TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Page

LIST OF ABREVIATIONS AND ACRONYMS

API	Application Program Interface
BOM	Bill Of Material
BP	Business Process
BPEL	Business Process Execution Language
CODP	Content Ontology Design Pattern
CRM	Customer Relationship Management
DL	Descriptive Logic
ERP	Enterprise Resource Planning
HQDM	High Quality Data Model
III-RM	Information Integration Infrastructure - Reference Model
MDM	Master Data Management
MES	Merchandizing Enterprise System
NAF	NATO Architecture Framework
NAICS	North American Industry Classification System
NAPCS	North American Product Classification System
NATO	North Atlantic Treaty Organization
NLP	Natural Language Processing
ODP	Ontology Design Pattern
OWL	Ontology Web Language
PO	Purchase Order
RA-EKI	Reference Architecture - Enterprise Knowledge Infrastructure
SCM	Supply Chain Management

SLR	Systematic Literature Review
SQL	Structured Query Language
TOGAF	The Open Group Architecture Framework
TOVE	Toronto Ontology Virtual Enterprise

INTRODUCTION

The role of the critical researcher is always to go beyond mere studying and theorizing, to actively affect change in the phenomena investigated.

W. Orlikowski and J. Baroudi (1991) citing (Benson, 1983)

Preamble

The *raison d'être* of this research project is to solve the semantic heterogeneity problem. The semantic heterogeneity problem detrimentally affects the capacity of an enterprise to maintain system interoperability, i.e. the capacity of the organization to have its systems exchange data in a seamless manner. The present delivery concludes this doctoral research project, hereafter referred to as "the project". This delivery additionally and partially fulfills the requirements of a Ph.D. program. The project first delivered and presented in conferences the Reference Architecture – Enterprise Knowledge Infrastructure (RA-EKI). RA-EKI is described in greater detail in (Fitzpatrick, Coallier, & Ratté, 2013; Fitzpatrick, Ratté, & Coallier, 2013). It was initially presented as a reference architecture for a semantic enterprise data warehouse (Fitzpatrick, 2012; Daniel Fitzpatrick, François Coallier, & Sylvie Ratté, 2012), then reformulated in the more generic RA-EKI. An earlier research plan can be found in annex I (Fitzpatrick, 2012). RA-EKI, illustrated in figure 0.1, represents one of the first published frameworks that encompass knowledge, know-how and intelligence in addition to traditional data and information. RA-EKI covers the full range of the epistemological building blocks as represented in figure 0.2 i.e. processing from data i.e. factual symbols (unstructured, semi-structured, structured) to information i.e. data with context; to knowledge i.e. actionable information; to know-how i.e. functional knowledge; and finally to intelligence i.e. cognitive know-how. RA-EKI can also be considered as a reference model of a cognitive architecture as defined and described by Lieto and co-authors (Lieto, Lebiere, & Oltramari, 2018). The epistemological foundation of this project is further explained in (Fitzpatrick, 2012).

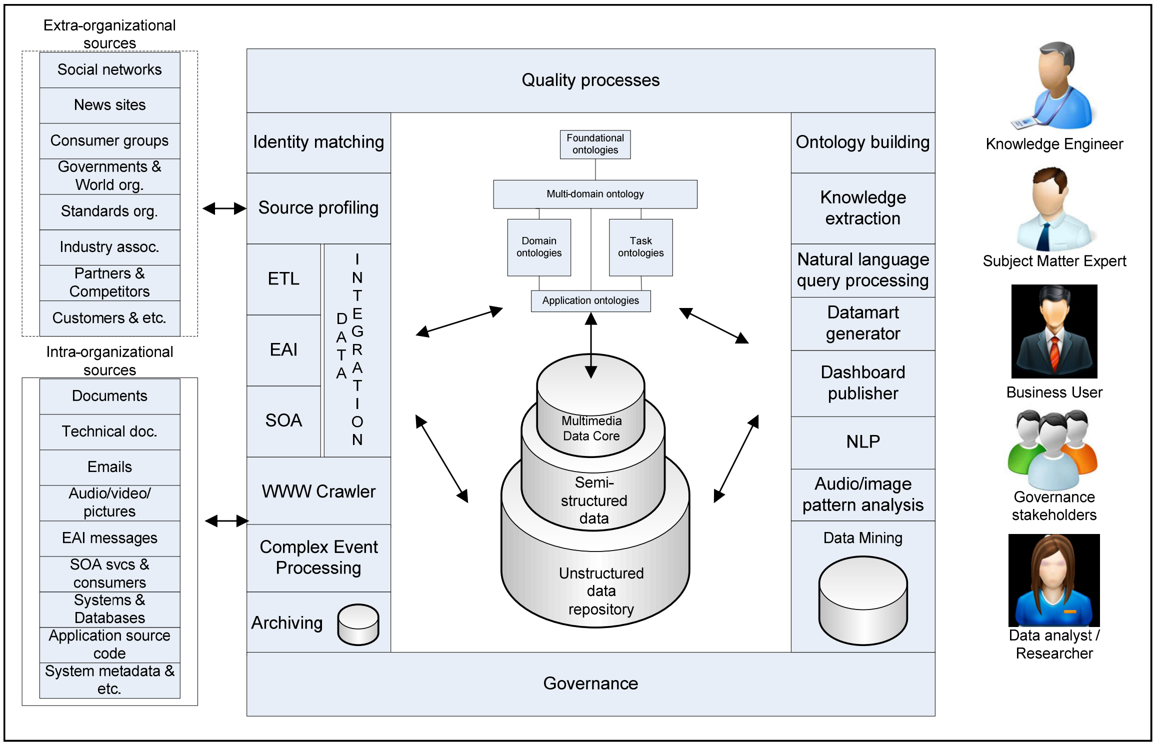


Figure 0.1 Reference Architecture – Enterprise Knowledge Infrastructure
(Daniel Fitzpatrick et al., 2013)

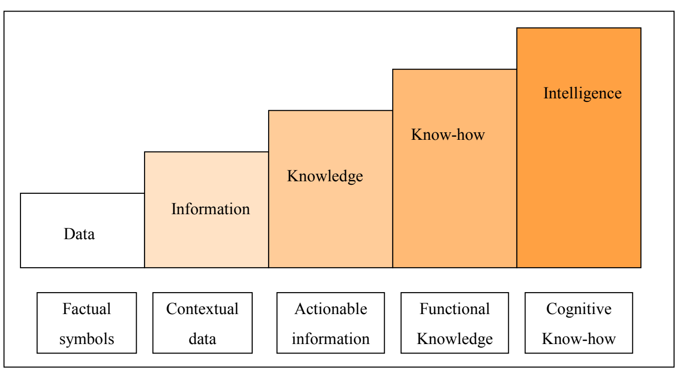


Figure 0.2 Epistemological foundation of this project
(Fitzpatrick, 2012)

RA-EKI also contributes a new type of mid-level (formal) ontology called multi-domain ontology. The multi-domain ontology serves as part of the terminological component ("T-Box") of a cognitive (inferential) application purposed for data integration and other functions such as Natural Language Processing (NLP). As illustrated in figure 0.3, the focus of the project has changed in a zoom in fashion from RA-EKI as a whole to the design of the internals of the multi-domain ontology's modules, RA-EKI's cornerstone. The (formal) multi-domain ontology comprises ontology modules the equivalent of subject areas for (semi-formal) data models.

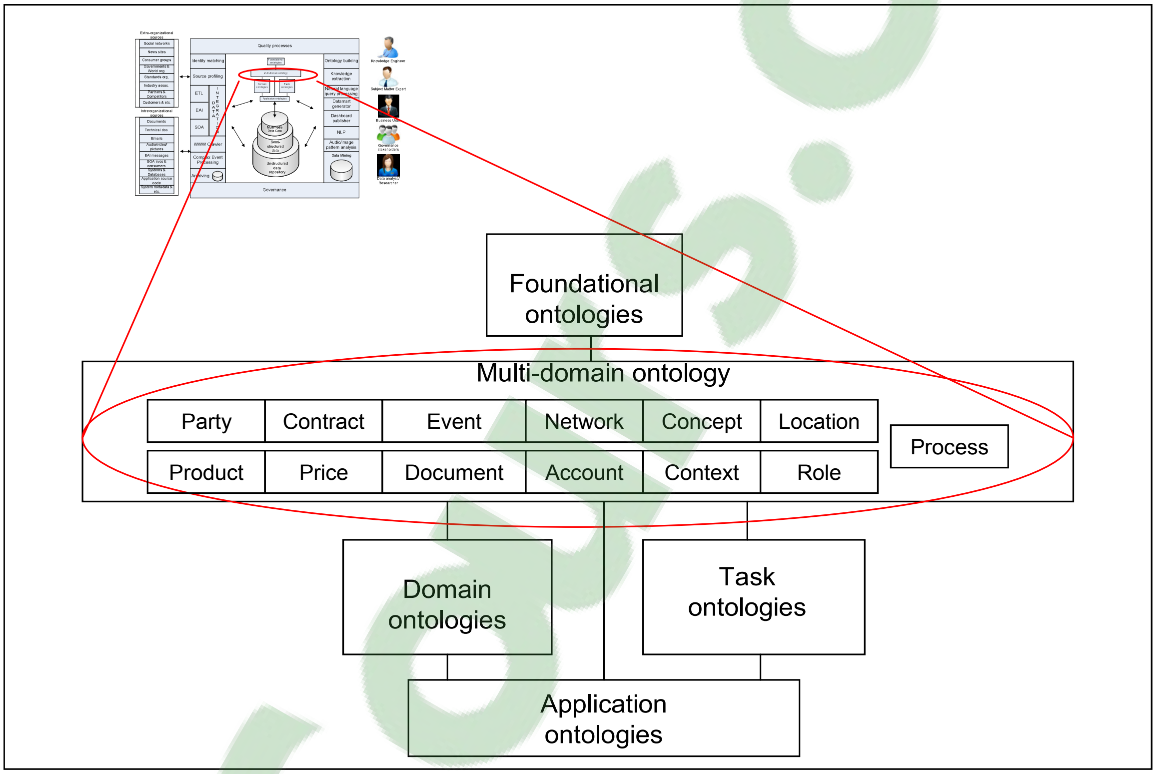


Figure 0.3 Focus on the design of the multi-domain ontology

This dissertation describes the research approach to specifically elicit agnostic data model patterns to eventually incorporate these patterns as axiomatized terminological rules in the multi-domain ontology's modules. This project intends to pursue the research effort with the ultimate goal to definitely resolve the semantic heterogeneity problem. The remainder of the introduction section comprises the following subsections:

1. Definition of important terms.

   In the case of the concept of ontology, there are several definitions. (Asunción Gómez-Pérez, Fernández-López, & Corcho, 2006) surveyed over a dozen different definitions of an ontology. This project intends to provide the most significant and consistent definitions in the context of this research, while attempting to avoid controversy;

2. Problem statement.

   The problem statement motivates the execution of this project. Although the project ends with the thesis defense, other research projects will need to be started and executed to achieve the desired theoretical saturation and ultimate resolution of the problem. Furthermore, shortcomings of the greater Information Technology (IT) and software engineering domain related especially to the selection and application of scientific methodology justify a greater diligence in the choice and design of a research approach;

3. Context.

   This subsection provides the holistic socioeconomic backdrop and factors related directly to the enterprises' requirements for system interoperability notably the creation of virtual enterprises and coalitions;

4. Research Objective.

   The project's intent related to the resolution of the stated problem. The objective can vary over time as research progresses;

5. Research questions.

   This subsection covers two research questions. The primary question addresses directly the project's objective and problem. The secondary question deals with the need to properly select the right approach to effectively solve the problem without the influence of scientific domain social factors;

6. Statement of the thesis argued in this project.

   The thesis statement constitutes the primary assertion that is defended in this dissertation's argumentation;

7. The research project's starting postulates.

   The starting postulates inspire the research question formulation;

8. Fundamental research approach,

   The project's research approach is summarily described using business process modeling;

9. Scope of the research project.
   This subsection outlines the expected findings of the project. The primary findings, the common thread of all research processes, represent the focus of the project. Secondary findings are also collected from the phenomenological research protocol mainly to provide context and preliminary data to be useful in subsequent projects;
10. Limits to the research project.
    The limits of the project represent what is to be excluded from the project but are likely to be included in future phases or projects related to the current problem;
11. Recapitulative overview of the project.
    This subsection provides the project's main themes in a data flow like representation. This concept map also illustrates post-project main activities leading to the resolution of the problem;
12. Structure of the dissertation.
    This subsection briefly describes the chapters.

1. Definition of important terms

The following terms constitute key notions for the project. Their definition intends to facilitate the reading of the introduction and the remaining chapters although some of the articles comprise a definition section as well.

Cognitive application

The project considers a cognitive application as a set of functions as represented in RA-EKI, figure 0.1. A cognitive application consists in a set of functions that transforms data or any of the other epistemological elements represented in figure 0.2 into a more advance stage e.g. data into information, information into knowledge, etc. These functions include NLP, data integration, knowledge extraction, ontology building and others. Also as prescribed in RA-EKI some or all of the functions may be ontology driven with the use of an inference engine (Lieto et al., 2018) (Daniel Fitzpatrick et al., 2013). Finally, an ontology driven cognitive application may stochastically infer its axioms using probabilistic reasoning (Kelly, 2015). In the case of being processed by probabilistic reasoning engine, the ontology is referred to as a fuzzy ontology (Carlsson, 2018).

Specification

This project defines a specification as a detail and shareable i.e. explicit description of a thing or a collection of things using a language, such as a detail design represented in a Unified Modeling Language (UML) class diagram. A specification may be deemed expressive in its capacity to represent a conceptualization (see next definition) in a machine-readable form to be processed by an automated application, including a cognitive application that processes axiomatic terminological rules (Guarino, Oberle, & Staab, 2009). A specification is also considered as the language dependent aspect of an ontology (Nicola Guarino, 1998).

Conceptualization

This project considers a conceptualization as a set of semantic elements, e.g. concepts, relationships, properties and human readable definitions (Lacy, 2005). Guarino and co-authors consider conceptualization as what is «*private to the mind of the individual*» (Guarino et al., 2009). Guarino considers conceptualization as the language independent aspect of an ontology as illustrated in figure 0.4 (Nicola Guarino, 1998).

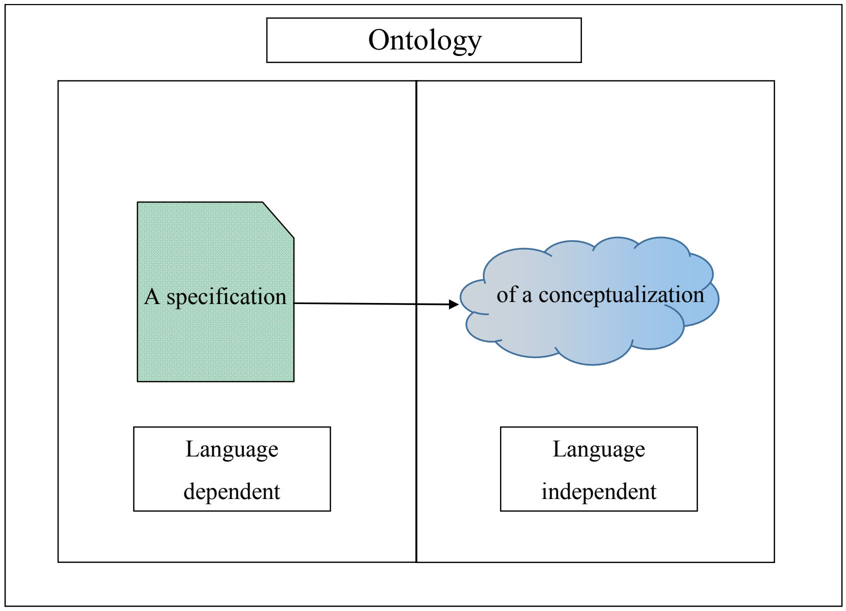


Figure 0.4 Language dependent and independent aspects of an ontology

Ontology

This project defines an ontology as a specification of a conceptualization. Gruber defines an ontology as an «*explicit specification of a conceptualization*» (Thomas R. Gruber, 1993). The project's definition removes the unnecessary *explicit* qualifying term since a specification is explicit by definition. Figure 0.5 outlines the language dependent aspect of an ontology. The specification aspect of an ontology comprises four levels: informal, semi-informal, semi-formal and formal. The informal level incorporates the natural language. Concept maps compose the semi-informal level. The semi-formal level encompasses the Entity Relationship Diagram (ERD) techniques and UML. The formal ontology level contains languages that define axioms forming a partial account of reality that can be processed by a semantic reasoning system or semantic reasoner (Guarino et al., 2009) (Bae, 2014). A semantic reasoner also known as an inference engine infers new axioms by deducting them from a base ontology. An ontology engineer provides a base ontology and validates consistency and correctness of the resulting superset (Lee, Matentzoglu, Sattler, & Parsia, 2015) (Bouten et al., 2016). Four exemplary formal languages are illustrated in figure 0.5. The Foundation for Intelligent Physical Agents' (FIPA) Agent Communication Language (ACL) supports the representation of ontological reasoned messages (Hsu & Cheng, 2015). The Semantic Web rule Language (SWRL) allows specifying axioms and knowledge rules (de Farias, Roxin, & Nicolle, 2016). The Resource Description Framework Schema (RDFS) language represents knowledge in the form of triple stores (subject, verb, object predicates) that can be used for semantic queries (Su et al., 2018). The Web Ontology Language (OWL) allows the representation of knowledge in an eXtensible Markup Language (XML) document encoding format (Rattanasawad, Buranarach, Saikaew, & Supnithi, 2018).

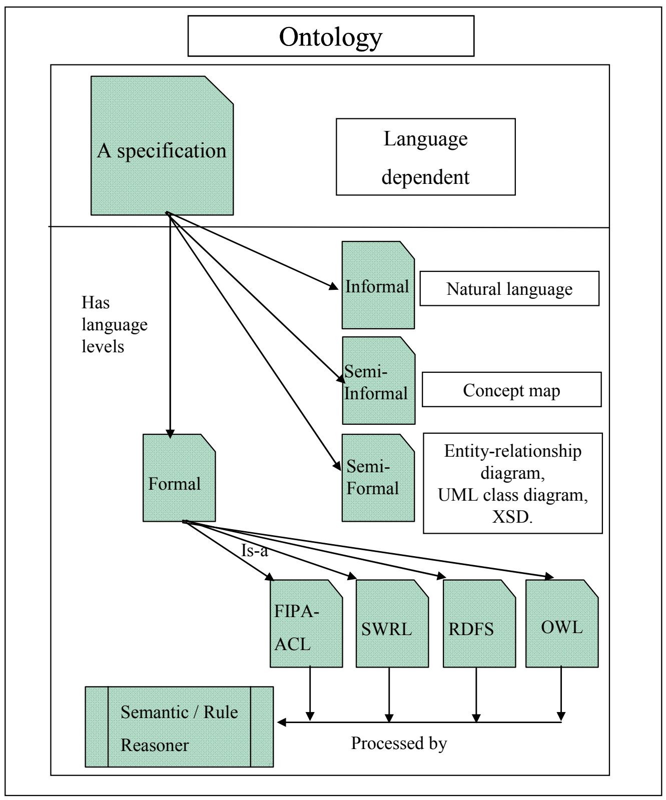


Figure 0.5 The language dependent aspect of an ontology

Figure 0.6 illustrates a conceptualization as the language independent aspect of an ontology (Nicola Guarino, 1998). A concept definition represents a human readable narrative that in supplies meaning to the concepts (Gruber, Liu, & Ozsu, 2009) (Noy & McGuinness, 2001). Lowering an ontology's abstraction may affect the robustness and flexibility of the conceptualization (Spyns, Meersman, & Jarrar, 2002). Semantic relationships are categorized as synonymy, antonymy, hyponymy, meronymy and holonymy relations. Synonymy relationships relate concepts with the same meaning. An antonymy relation associates opposing or disjoint concepts. The Hyponymy relationship subsumes a specific concept to a generic one. The meronymy and holonymy relationships support the equivalent of the UML composition relationship, the former indicates that a concept composes another one, while the latter indicates that one concept includes another one (Nicola Guarino, 1998) (Lacy, 2005).

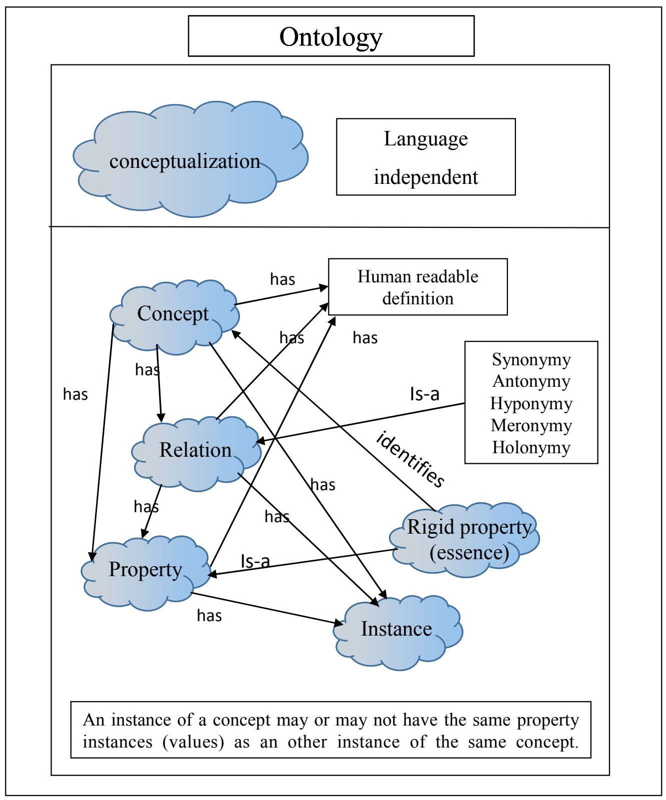


Figure 0.6 The language independent aspect of an ontology

Ontology Design Pattern (ODP)

An ODP represents «*a set of ontological elements, structures or construction principles that solve a clearly defined particular modeling problem*». ODPs for formal ontologies are translated into axioms in a specialized language such as OWL during ontology development. Ontology architecture patterns only cover the ontology as a whole or modules as the ones within the multi-domain ontology illustrated in figure 0.3. ODPs pertain to specific concepts or relations (Blomqvist, 2009b). (Blomqvist, 2010).

Content Ontology Design Pattern (CODP)

According to (Gangemi & Presutti, 2009) (Blomqvist, 2009a), a content ODP, or a CODP, is a design pattern that addresses business concepts found in a domain ontology. This research project specifically investigates CODPs representing business concepts that are meant to be applicable to all industry sectors.

Agnostic CODP

This project defines an agnostic CODP as an abstract ODP that possesses a distinct definition among other concepts and that can apply to any industry sector. This definition is inspired by Thomas Erl's definition of the term Agnostic in the context of Service Oriented Architecture as an application service that is business process independent and reusable across all contexts and domains in the enterprise (Erl, Merson, & Stoffers, 2017). Furthermore, an agnostic CODP is defined in such a way that it cannot be confused with other agnostic concepts.

Multi-domain ontology

A mid-level formal ontology that comprises a collection of interrelated agnostic CODPs that allow a cross-industry conceptualization (Daniel Fitzpatrick et al., 2012). Concepts related to any industry may be represented using the multi-domain ontology. The multi-domain ontology comprises modules that would possibly assist the ontology engineer in optimizing the agnostic axioms' interactions .(Hitzler & Shimizu, 2018).

2. Problem Statement

Starting in the 1990s, system interoperability, the ability of application systems to exchange information and conduct coordinated processes, has become an important intra and extra organizational requirement. This organizational requirement stems from the increasing need for the organizations to cooperate within and between organizations (Lu, Panetto, Ni, & Gu, 2013) (Estublier, Cunin, Belkhatir, Amiour, & Dami, 1998).

This research project targets the semantic heterogeneity problem, dubbed the «*old problem*» by De Giacomo and co-authors in (De Giacomo, Lembo, Lenzerini, Poggi, & Rosati, 2018). Some work pertaining to its solution, data integration, dates back over 30 years ago (Deen, Amin, & Taylor, 1987). Semantic heterogeneity originates from having application systems designed with different vocabularies, data models or ontologies. It affects the capacity of enterprises to have their systems interoperate within and between organizations. Systems interoperability represents a crucial capability to industry and government sectors alike. Semantic heterogeneity plagues the industry sectors by costing valuable funds (Lenz, Peleg, & Reichert, 2012) (M. Dietrich, Lemcke, & Stuhec, 2013) (Lemcke, 2009) (Brodie, 2010) (Jhingran, Mattos, & Pirahesh, 2002). It also hinders medical and pharmaceutical sectors in depriving them from some research funds needed to preserve and save lives (Williams et al., 2012) (Mirhaji et al., 2009). The scientific community has yet to propose a final solution for this problem (Doan, Halevy, & Ives, 2012) (Olivé, 2017) (Olivé, 2018).

The IT scientific community has conducted research notably in the development of formal ontologies for reasoning applications to resolve the semantic heterogeneity problem. Cognitive applications would perform the data integration function with the use of formal ontologies containing knowledge assertions (Bergamaschi et al., 2018) (Haziti, Qadi, Bazzi, & Elhassouni, 2018). Ontology science and engineering lack the maturity to provide a coherent theoretical framework to allow truly cross-enterprise semantic interoperability solutions (Pinkel et al., 2015). To illustrate the lack of maturity, in (Bennett & Bayrak, 2011), the authors define a data integration system as a «*general-purpose (application) used to* *provide interoperability among autonomous heterogeneous database systems*». Later in the same article, the authors refer to data integration as a «*problem*». In (Lenzerini, 2002), the authors define data integration as «*the problem of combining data residing at different sources, and providing the user with a unified view of these data*». Confusing the problem, i.e. semantic heterogeneity, with the solution, i.e. data integration, sheds doubts in the theory-building research process.

Dietrich and co-authors reported, citing an Aberdeen report (Kastner & Saia, 2006), that semantic heterogeneity may cost 40% of IT budget in deploying data integration platforms (M. Dietrich et al., 2013). The cited Aberdeen report does not explain the research method used to determine the cost of a significant problem such as semantic heterogeneity. If applied hypothetically to the United States of America's 2016 global output (Anonymous, 2016) of over $31.9 trillion and considering that IT costs in average 3.3% of corporate revenues in all industry sectors (Hall, 2016), the problem of semantic heterogeneity would cost the US economy each year in excess of $400 billion. Simply quoting an unsubstantiated number such as the cost of semantic heterogeneity in terms of the expenditures in developing data integration may not constitute effective scientific research, let alone sound theory building.

This research project intends to perform more disciplined theory building based on a dual method qualitative research approach. This project's approach is based on a similar dual method research described in (Bano, Zowghi, & da Rimini, 2017) to alleviate the issues raised in this section. The project's approach aims to demonstrate trustworthiness and hopefully stimulate a more definitive progress to resolve the semantic heterogeneity problem.

3. Context

The impact of economic woes, in the aftermath of the great recession of 2007 (Elsby, Hobijn, & Sahin, 2010), and the increase of compliance regulations render the enterprises more dependent on internal and external collaborations to cut costs and to achieve their strategic objectives and fulfill their mission more efficiently (Duygan-Bump, Levkov, & Montoriol-

Garriga, 2015) (De Toni, 2016). The significant pressure to reduce waste, in addition to costs, motivates the organizations of all industries to internally operate more efficiently with their existing customer base. Globalization, removal of trade constraints and the evolving regulatory landscape impose further pressure notably on the service industry (Bagheri & Jahromi, 2016). Direct relationship marketing monopolizes excessively financial and other resources to maintain good relations with existing customers. Again, as in the case of partnerships, the organizations' information systems must also interoperate to allow individual enterprises to strive in retaining their customers and expand their business.

Defense government agencies are affected as well by semantic heterogeneity in their attempt to implement system interoperability. Semantic heterogeneity constitutes an important challenge for large enterprises and notably for organization such as the US Department of National Defence (Morosoff, Rudnicki, Bryant, Farrell, & Smith, 2015). In manufacturing, new approaches to design products are proposed to allow product manufacturers to be more competitive: Set-Based Design (SBD) (Kerga, Schmid, Rebentisch, & Terzi, 2016), a new product development process proposed in (Belay, Welo, & Helo, 2014) and the modular approach, popular notably in aerospace manufacturing (Buergin et al., 2018). The SBD approach, for example, can contribute reducing in average by 25% the project duration and by 40% the total project costs as demonstrated in laboratory simulations (Kerga et al., 2016). These new product design approaches require that the Product Lifecycle Management (PLM) systems interoperate. Semantic heterogeneity adversely affects system interoperability thus hindering efforts to execute the new product design methodologies (Daniel Fitzpatrick et al., 2013).

4. Research Objective

This research project aims to elicit data model patterns from experienced practitioners and from rigorously selected publications. The data model patterns are to be re-engineered as agnostic axioms and to compose the multi-domain (formal) ontology. Although data model patterns are only used in semi-formal ontologies, e.g. database and software design, they can contribute for building formal ontologies, such as the multi-domain ontology (Blomqvist, 2010). The use of formal ontologies within data integration cognitive platforms constitutes an efficient approach to solve semantic heterogeneity (Jirkovský, Obitko, & Mařík, 2017).

5. Research questions

In this project, the research questions allow to transition, during the project, from the research objective to the actual research protocols. This project considers a research protocol an instance of a method with a specific research question or set of questions. The first question pertains directly to the objective:

Research question #1

What are the conceptualization patterns found in semi-formal ontologies, e.g. data model patterns, software engineering patterns, etc., that can be agnostic to any domain or industry sector in the context of enterprise semantic interoperability and can be used as the basis of agnostic CODPs to resolve semantic heterogeneity in enterprise systems?

Research question #1 is to be translated into more detail forms of investigation in the design and execution of the research protocols. The second question raises the contentious issue about choosing a research approach, specifically in selecting between theory testing or theory building approaches. As indicated in (P. Leedy & Ormrod, 2012), two fundamental approaches can be used: theory building and theory testing. Theory testing or quantitative methods typically use known variables to statistically measure and validate the extent to which a theory can explain a phenomenon. Theory building or qualitative methods, on the other hand, attempt to explain a phenomenon and explore its various facets. While quantitative methods are relatively standard, qualitative research methods do not benefit from standardization and are still evolving (P. Leedy & Ormrod, 2012). The use of qualitative research methods in information systems may constitute a highly contentious matter

(Marshall, Cardon, Poddar, & Fontenot, 2013). The project formulates the second question as in the following:

Research question #2

What research method or methods can be used in the attempt to effectively answer the first research question while providing sufficient evidence to instill confidence in the methodology employed and in the findings?

The second question requires reviewing the literature pertaining to research methods in information systems, information technology and software engineering. The literature review performed in this project to address research question #2 included a text book well cited by researchers: (P. Leedy & Ormrod, 2012), which provides guidance on selecting between theory testing and theory building. A contribution from Orlikowsky and Baroudi raised the issue in 1991 about the detrimental effect of the exclusive use of quantitative research for information systems (Orlikowski & Baroudi, 1991). Chapter 1 provides a more complete perspective on the literature review performed for addressing research question #2 and for the design decision made and indicated in the upcoming introduction's fundamental research approach subsection.

6. Statement of the thesis argued in this project

As indicated in the problem statement subsection, this project in effect addresses the problem pertaining to semantic heterogeneity and secondly the need to perform more disciplined theory building. This research project argues the following thesis as the position defended by the dissertation (Anonymous, 2018):

There is a set of data model patterns that are applicable to any private industry or government sector that can be used as agnostic CODPs and collectively constitute, after being translated into axioms, a (formal) multi-domain ontology that can be used by a cognitive data integration application to resolve the semantic heterogeneity problem.

7. The research project's starting postulates

The project's starting postulates describe the researcher's sources of inspiration for specifying the primary research question. The researcher draws from professional experience to formulate a first research question construed as potentially beneficial for an optimized research roadmap. These postulates only apply for the beginning of the project and may become irrelevant as new phases or projects pursue the exploration, theory building and theory testing efforts leading to the ultimate resolution of the problem. The starting postulates are:

- Agnostic CODPs that ensure ontology reusability are needed for a multi-domain ontology to be used in a cognitive data integration platform. This postulate conceptually originates from (Erl, 2008) and (Erl et al., 2017);
- Data model patterns can be used to kick start the development of formal ontologies (Blomqvist, 2010);
- The conceptualization aspect of an ontology is key to the richness of an ontology's axioms (Guarino et al., 2009);
- Best practice for formulating CODPs consist in the use of ODPs that can be used across several domains (Blomqvist, 2010);
- Data model patterns, such as those proposed in (West, 2011) and (Blaha, 2010b), may contribute to a more efficient multi-domain ontology for a cognitive data integration platform.

8. Fundamental research approach

In the problem statement subsection, the semantic heterogeneity problem represents the focus of this project. While performing research to contribute in solving this problem, this project also proposes a research design to demonstrate trustworthiness. The research design needs to ensure the elicitation of agnostic data model patterns or agnostic CODPs used interchangeably, fulfilling the first thesis while establishing the credibility, dependability, confirmability and transferability of the proposed dual method approach, supporting the second thesis. A purely qualitative research approach is proposed in this project to start the theory-building process.

This decision stems in part from a position taken in (Orlikowski & Baroudi, 1991) who were the first to argue that the exclusive use of positivist (hypothetico-deductive) methods may detrimentally affect the effort of effectively engaging all scientific challenges in information systems. Shirley Gregor posits in (S. Gregor, 2006) and (Shirley Gregor, 2017) that the science of design, to which the project subsumes, requires a theory-building approach. A qualitative research method is prescribed (P. Leedy & Ormrod, 2012) to build theory when needed. In (Alemu, Stevens, & Ross, 2011), the authors clearly argue that semantic interoperability research requires a qualitative research approach. This decision about the selection of the research methodology is also problematic since some IT postgraduate faculties with a positivist stance, and under pressure to produce studies, react in a hostile manner against qualitative (constructivist) studies (Marshall et al., 2013). Marshall et al. also argue the scarcity of methodological standards in qualitative research, notably to establish the trustworthiness of the research process. This controversial situation motivates a careful and diligent approach for designing the research methodology for this project.

This project's research design is based on a concurrent dual qualitative research approach that represents one of the first actual utilizations of such research methodology. The consequences of the decision to only perform qualitative research entail that this project is attempting to establish trustworthiness and not validity (Guba & Lincoln, 2001) (Cypress, 2017). Furthermore, this qualitative research process being essentially exploratory is driven only by a research question and not by hypotheses such as in the case of hypothetico-deductive or mix methods research (Wohlin & Aurum, 2015). Future phases of this project may involve a research design using a mixed-method phenomenological approach as proposed by (Flynn & Korcuska, 2018) where strengths of both qualitative and quantitative approaches may be used to solidify this emerging theory's foundation.

This project's research approach is inspired from another dual qualitative research method approach designed by M Bano, D Zowghi and F Da Rimini. Bano and her team's approach uses qualitative SLR and case studies to investigate requirements engineering, specifically in the relationship of user involvement and system success in software development (Bano et al., 2017).

Figure 0.7 holistically illustrates this project's dual method research approach. Process P1 comprises the high-level design activities to define the two theory elicitation protocols, i.e. the qualitative SLR and the phenomenological research method, two use cases and the concluding trustworthiness establishment activity. Process P1 also prescribes a strategy to establish trustworthiness for the research process. Process P2 pertains to the detail design and execution of the SLR protocol, based on (Okoli, 2015; Okoli & Schabram, 2010). P2 formulates a practical screen that retains or rejects publications in two stages. The practical screen's first stage filters papers based on their metadata. The second stage requires reading the publications. Then, the SLR's analysis and synthesis stages are based on this project's phenomenological research method (C. Moustakas, 1994) and from (Thomas & Harden, 2008). Process P3 covers the detail design and the execution of the phenomenological protocol. P3 establishes a purposeful sampling approach in selecting participants. The participants, called co-researchers, provide an insight to their experience in the phenomenon defined as data integration. Also P3 elaborates the semi-structured interview questionnaire (Bevan, 2014), the analysis and synthesis activities (C. Moustakas, 1994). Finally, P3 defines the computation method for determining data saturation that is also used in P2 (Marshall et al., 2013). Processes P4 and P5 execute use cases for collaborative product design for manufacturing and collaborative logistics planning for military coalition deployments. The use cases intend to show transferability from the SLR's findings in both contexts respectively. Finally, process P6 establishes the trustworthiness of this project's dual qualitative research method approach using the criteria proposed in (Guba & Lincoln, 2001) (Anney, 2014) (Forero et al., 2018) (Suri, 2011).

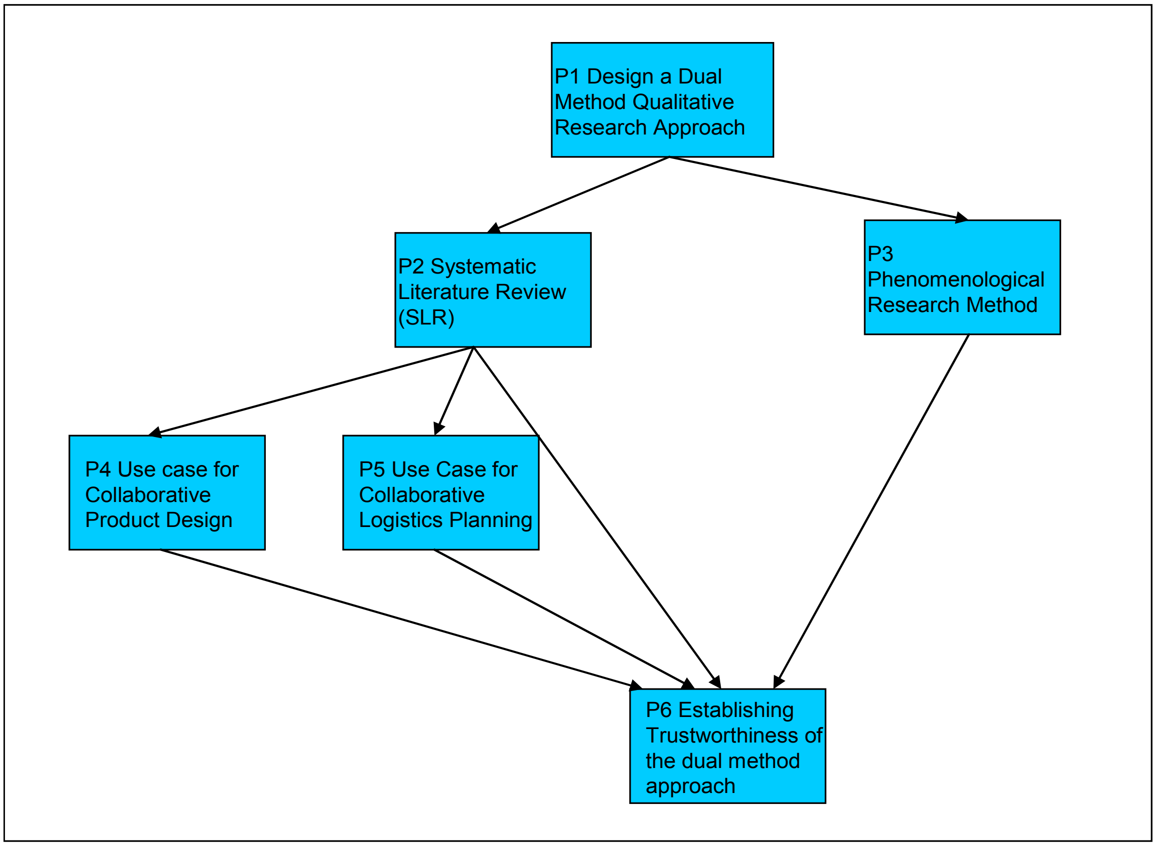


Figure 0.7 Overview of the dual method qualitative research approach

9. Scope of the research project

The focal point of the project is to collect agnostic CODPs, i.e. ontology design patterns that represents business concepts of various domains and that can apply to any industry sectors. The completion of this project consists in the development of a run-time multi-domain ontology functioning as the terminological component or T-box of a data integration cognitive platform. Concretely, the multi-domain ontology will comprise agnostic axioms, produced from the translation of the agnostic CODPs elicited in this delivery. Although less expressive than formal ontology languages such as OWL, UML can still show hyponymy, meronymy and holonymy relationships that constitutes valid ontology design pattern material based on the definition of an ontology design pattern formulated in (Blomqvist, 2010). Since the project for the time being does not translate patterns into axioms, agnostic data model patterns are considered agnostic CODPs. Blomqvist also considered the benefits of accelerating the development of axioms intended for cognitive applications using data model patterns; a critical consideration for this project. The early of the project for the current delivery only elicits data model patterns. The project considers collecting from formal ontologies only at a later phase. As indicated earlier in this subsection, agnostic CODPs constitute the common thread of this project's research processes.

10. Limits to the research project

By virtue of the research question, this project and its dual-method design concentrate exclusively on data models patterns, or semi-formal ontology patterns, elicited from selected publications through the SLR's practical screen and by interviewing experienced practitioners using the phenomenological research method. The recommendations set forth by Blomqvist in (Blomqvist, 2009a), to elicit data model patterns to be used as CODPs, defines this project's fundamental purpose and inspiration. This project's limitations include the following:

- Only data model patterns are considered for this research. No formal ontologies are studied for concept elicitation;
- Only data model patterns related to a business context are handled;
- Only publications written in English or French can be retained;
- Only participants speaking English or French can be retained;
- The SLR only covers papers published between 2009 and 2017 inclusively;
- Only domain level concepts are considered. No foundational concepts such as "Instance" are considered;
- The conversion of agnostic CODPs to axiomatic representation in the multi-domain ontology and further design and development of the ontology are not part of this project;

- The logical representation in a Description Logic language is not covered in this project. Agnostic CODPs are represented in light UML;
- This project's phenomenological research method limits the number of co-researchers to 15. Although this number may increase within the five to 25 range proposed by (P. Leedy & Ormrod, 2012), this project does not intend satisfying data or theoretical saturation. This project considers data saturation as the point where no new knowledge is created with the current research question. Further work may be required to achieve data saturation. Theoretical saturation represents here the point where no new knowledge is created after all possible research methods and protocols, qualitative, quantitative or mixed, have been used;
- The methodology to assemble and integrate the agnostic CODPs into the multi-domain ontology and consistency checks are excluded as well;
- A formal audit has not been performed although the data is available for such review on demand along with an audit protocol;
- A multi-researcher triangulation process for establishing methodological and data trustworthiness has not being performed during this project but is planned for upcoming phases.
- The project may remove certain of the aforementioned limitations in future phases.

11. Recapitulative overview of the project

Figures 0.8 and 0.9 summarize the project in its current form using concept maps. Figure 0.8 illustrates the triangle between the semantic heterogeneity problem, the enterprise's capacity this problem affects, i.e. system interoperability and the data integration cognitive platform, the solution to the semantic heterogeneity problem.

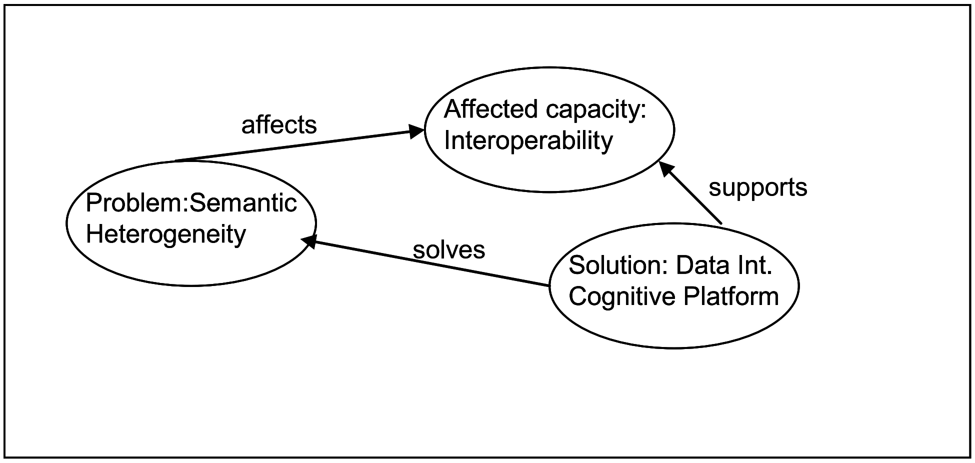


Figure 0.8 The problem, the affected capacity and the solution triangle

Figure 0.9 provides a recapitulative and holistic perspective of this project in its current state starting with the formulation of the problem. This holistic perspective finishes with the establishment of trustworthiness and the building of a proposed theory containing agnostic CODPs, the common thread to this entire research, also referred as phenomenon knowledge. Additionally, the phenomenological protocol provides on its own other elements of knowledge, or peripheral knowledge, such as quality and efficiency, elicited material that may assist the project to clearly define metrics in a future phase. The semantic heterogeneity problem requires a research objective that orients the research efforts toward what is believed to be key pieces of the solution, the agnostic CODPs. RA-EKI defines as its centerpiece the multi-domain ontology that is to be composed of agnostic axioms. The multi-domain agnostic axioms, the logical semantic rules represented in formal language and executable in cognitive applications, will originate from the agnostic CODPs elicited during this project.

Also illustrated in figure 0.9, the research objective generates two research questions as previously indicated, the primary question relative to the existence of agnostic CODPs and the secondary question regarding the research methodology that should be used during the initial stage of this project. The research methodology to be used in the initial stage of this project consists in a dual qualitative research method approach. The dual qualitative research approach comprises the SLR and phenomenological research methods, which design is described in chapter 1. Each method is instantiated into a protocol that specifically addresses the research question. The SLR protocol comprises a practical screen that filters queried publications and extracts agnostic data model patterns for analysis and synthesis. The phenomenological protocol comprises a questionnaire that is used to extract agnostic data model patterns from participants also for analysis and synthesis. The phenomenological protocol is also designed to elicit peripheral knowledge related to quality and efficiency of agnostic data model patterns, or agnostic CODPs. Both protocols build proposed theory as documented in chapters 2 and 5. They also contribute to the establishment of trustworthiness that provides the means to assess, in the context of qualitative research, to what extent the methods and the findings may be trusted, as covered in chapter 6. The protocols also provide the means to determine a certain level of theoretical saturation. Theoretical saturation partly guides the formulation of the research objective and questions and the evolution of the project also documented and discussed in chapter 6. In future phases, the proposed theory built during the present stage will be experimentally deployed. The agnostic CODPs will be translated into agnostic axioms and integrated in the multi-domain ontology using a formal ontology modeling tool. The multi-domain ontology will be incorporated in a data integration cognitive platform, in addition to data integration, cognitive and other task ontologies. The data integration cognitive platform will be developed to solve the semantic heterogeneity problem within the RA-EKI framework.

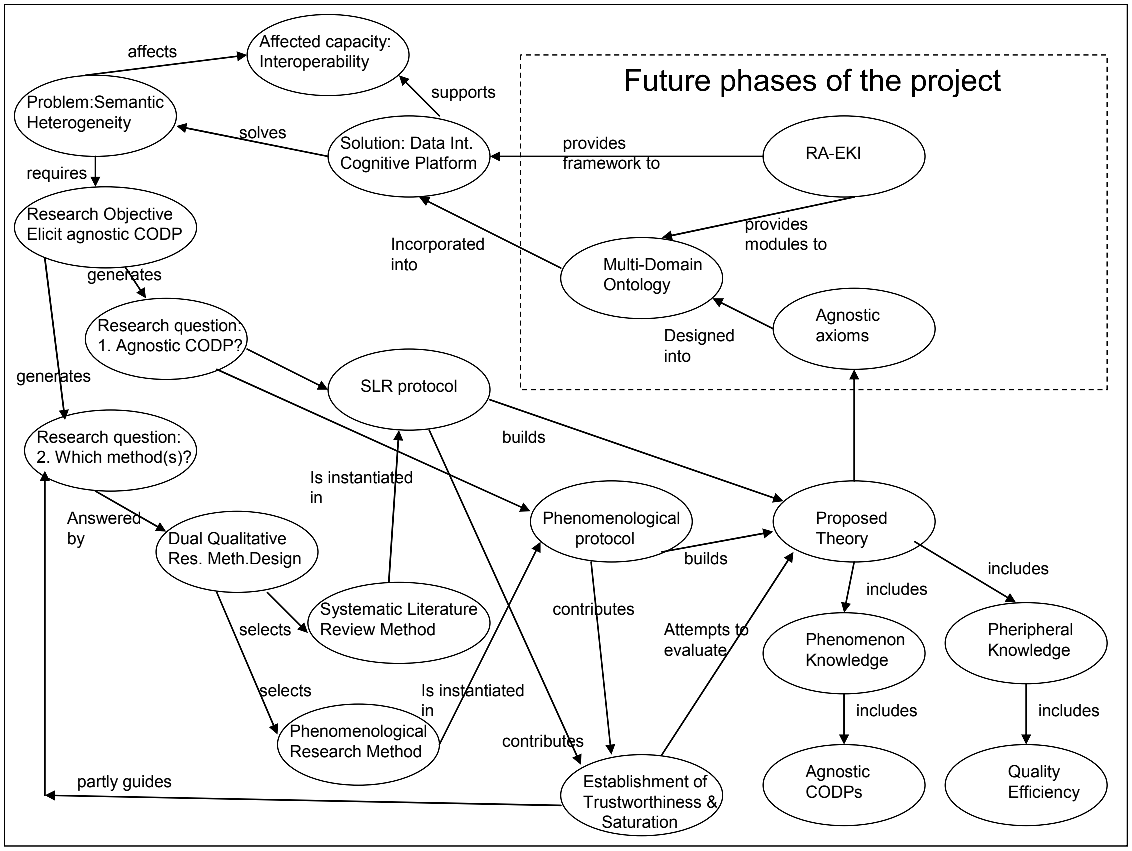


Figure 0.9 Recapitulative overview of the project

12. Structure of the dissertation

This manuscript-based dissertation comprises six chapters, each corresponding to an article that covers the work, and in some cases findings, performed in the context of the processes illustrated in figure 0.1. These chapters intend to argue both theses described earlier in this section. Chapter 1 covers the high-level design activities for the dual method qualitative research approach associated with the two main protocols themselves, the two supporting use cases and the concluding trustworthiness establishment process. Chapters 2 and 5 constitute the main processes in which two distinct and autonomous research methods, the SLR and the phenomenological research methods, have their protocols and findings richly documented. Chapter 2 outlines specifically the in-depth description of the protocol, the practical screen, the search query and the findings, including a set of UML diagrams representing the results of the analysis and synthesis of the retained publications, which consists in the agnostic data model patterns. Chapter 5 covers its protocol that is centered on a semi-structured interview questionnaire. Chapter 5 similarly to chapter 2 outlines the outcome of the analysis and synthesis steps in the form of UML diagrams representing the sought data model patterns. Additionally, chapter 5 provides context knowledge relative to the co-researchers average years of experience and industry sectors they were involved; peripheral knowledge also provided insight in the co-researchers' belief regarding, for example, the notions of efficiency and quality measurements, to be used in future stages of the project. Chapters 3 and 4 pertain on specific industry applications of the agnostic CODPs elicited in the SLR study. These two specific industry applications were randomly selected from several other industry domains and sectors that are subject to research on data integration. Chapter 3 examines the potential application of the SLR's elicited agnostic data model patterns in the context of collaborative logistics planning for military coalition deployment. Chapter 4 covers the SLR's data model patterns application in the context of collaborative product design in manufacturing. Chapter 6 establishes the trustworthiness of the dual method qualitative research approach by applying the four trustworthiness criteria: credibility, dependability, confirmability and transferability.

CHAPTER 1

A DUAL METHOD QUALITATIVE RESEARCH DESIGN FOR ELICITING AGNOSTIC CONTENT ONTOLOGY DESIGN PATTERNS FOR A MULTI-DOMAIN ONTOLOGY

Daniel Fitzpatrick[1], François Coallier[1], Sylvie Ratté[1]

[1]Department of Software Engineering & Information Technology, École de technologie supérieure,
1100 Notre-Dame West, Montréal, Quebec, Canada H3C 1K3

Paper submitted for publication to Empirical Software Engineering in September 2018

Abstract

In all private and government sectors, the semantic heterogeneity problem constitutes an important roadblock to organizations' efforts to implement systems interoperability. Semantic heterogeneity, an unnecessary ill, originates from application systems designed with different vocabularies or data models within an enterprise. Systems interoperability represents a crucial capability to the industry and government sectors. This paper proposes a dual method approach to establish the trustworthiness of a qualitative research project to elicit agnostic Content Ontology Design Patterns (CODPs). These two methods are covered in separate publications. First, the (qualitative) Systematic Literature Review (SLR) approach studies relevant publications using a rigorous approach to elicit the sought agnostic CODPs. Secondly, the phenomenological research method investigates through semi-structured interviews primarily the agnostic CODPs and other secondary topics. The SLR approach intends to elicit data to construct theory around a specific type of mid-level ontology called a multi-domain ontology. The concept of multi-domain ontology has been proposed previously in (Daniel Fitzpatrick et al., 2012, 2013; D. Fitzpatrick et al., 2013). The SLR approach uses a practical screen that comprises a set of criteria to select publications to have their content examined, analyzed and synthesized. The findings are represented in the form of CODP template. This paper's research approach draws from Clark Moustakas' phenomenological research methods. Clark Moustakas' phenomenological research methods, applied in clinical

psychology, elicit theoretical material through the experience of participants Moustakas referred to as co-researchers. The concept of abstract, or agnostic, concepts used for data integration represents the studied phenomenon. As in the case of the SLR, the content elicited from the interview is examined, analyzed and synthesized.

Keywords: Content ODP, Ontology Design Patterns, Ontology, inference application, multi-domain ontology, Systematic Literature Review, phenomenological research method, trustworthiness, constructivism, dual method, qualitative research.

1.1 Introduction

In all private and government sectors, the semantic heterogeneity problem constitutes an important roadblock to organizations' efforts to implement systems interoperability. Semantic heterogeneity originates from application systems designed with different vocabularies or data models within an enterprise. Systems interoperability represents a crucial capability to the industry and government sectors. The scientific community has yet to propose a solution for this problem (Doan et al., 2012) (Olivé, 2017). This problem has a financial impact in respect to IT expenses that can be used for more productive functionality, (M. Dietrich et al., 2013), (Lemcke, 2009) as well as (Brodie, 2010) and (Jhingran et al., 2002). Furthermore, there may be consequences in terms of human life since there is logically a cost stemming from valuable medical and pharmaceutical research funds wasted in addressing semantic heterogeneity (Lenz et al., 2012). In (Williams et al., 2012) and (Mirhaji et al., 2009) the authors stress that efforts in deploying data integration pose significant challenges in biomedical research and hinders knowledge discovery critically needed to develop new drugs.

One solution to the semantic heterogeneity problem is data integration using semantic web-capable technologies (De Giacomo et al., 2018). Data integration is a capability that allows harmonizing the meaning of data originating from various sources in a seamless manner, as if the data came from one single source (Jirkovský et al., 2017). (Daniel Fitzpatrick et al., 2013) propose a knowledge management model: the Reference Architecture – Enterprise

Knowledge Architecture (RA-EKI), which comprises high-level specifications for several ontology-driven applications such as Natural Language Processing (NLP), knowledge extraction and data integration. RA-EKI comprises a mid-level type ontology, a form of ontology more specific than a foundational ontology but more generic than a domain ontology (Obrst, Chase, & Markeloff, 2012) (Zuanelli, 2017) called the multi-domain ontology. The multi-domain ontology is designed to fulfill the requirements of various semantic web-based applications, such as inferential or cognitive applications.

The multi-domain ontology intends to apply to all industry and government sectors. Its conceptualization is to be agnostic, a characteristic based on T. Erl's Service Oriented Architecture (SOA) principle of service reusability through agnostic design (Erl, 2008). Erl defines agnostic design as a design that can apply across the enterprise. (Fitzpatrick, Ratté, & Coallier, 2018a) extend this definition to a design which semantics can apply across all industry sectors. The objective of this paper is to propose a research design to elicit agnostic Ontology Design Patterns (ODPs) for the design of the multi-domain ontology. ODPs are defined by as: «*a set of ontological elements, structures or construction principles that intend to solve a specific engineering problem and that recurs, either exactly replicated or in an adapted form, within some set of ontologies, or is envisioned to recur within some future set of ontologies*» (Blomqvist, 2010).

As recommended by (Blomqvist, 2010), data model (semi-formal ontology) patterns can be used to «*kick-start the usage of [formal ontology] patterns*». Based on the latter recommendation and by (Fitzpatrick, Ratté, et al., 2018a) classifying data models as semi-formal ontologies, this project states the following as its research objective: «*To elicit data model patterns. The data model patterns are to be re-engineered as agnostic CODPs and to compose the multi-domain ontology*» (Fitzpatrick, Ratté, et al., 2018a). The research question strictly focuses on eliciting the agnostic (design) data model patterns considered, by definition, as agnostic CODPs. The research question is formulated as:
« *What are the conceptualization patterns found in semi-formal ontologies, e.g. data model patterns, software engineering patterns, etc., that can be agnostic to any domain or industry*

sector in the context of enterprise semantic interoperability and can be used as the basis of agnostic CODPs to resolve semantic heterogeneity in enterprise systems?» (Fitzpatrick, Ratté, et al., 2018a).

This problematic situation consequently influences the decision regarding the selection of the fundamental scientific approach to use. As argued in (Daniel Fitzpatrick et al., 2012):

The current IT theoretical frameworks do not adequately support the industry in terms of knowledge and know-how in respect to ontology-based data integration. No existing methodology would allow, without research, to elaborate an ontology-based [design] approach of a cross enterprise data integration capability... A qualitative research project to achieve the research objective is therefore warranted. For this purpose, a theory-building qualitative research approach is considered here.

This decision stems from a position taken by Orlikowski and Baroudi in (Orlikowski & Baroudi, 1991) who were the first to argue that the exclusive use of positivist (hypothetico-deductive) methods may be detrimental to the effort of engaging all scientific challenges in information systems. Shirley Gregor posits in (S. Gregor, 2006) and (Shirley Gregor, 2017) that in the science of design, to which the present project subsumes, a theory-building approach is warranted. A qualitative research method is prescribed (P. Leedy & Ormrod, 2012) to build theory when needed. This decision about the selection of the research methodology is also problematic since some IT postgraduate faculties with a positivist stance, and under pressure to produce studies, react in a hostile manner against qualitative (constructivist) studies (Marshall et al., 2013). This controversial situation motivates a careful and diligent approach for designing the research methodology for this project.

The mixed methods approach, using both quantitative and qualitative research designs as recommended in (John W Creswell & Creswell, 2017), is not used by this project on the basis that this a first exploratory effort and all project resources are concentrated in eliciting knowledge. The utilization of a mixed methods approach design may be considered for future research efforts. This project's research design is based on a concurrent dual qualitative research approach that represents one of the first actual utilizations of such research methodology. The consequences of the decision to only perform qualitative research entail that this project is attempting to establish trustworthiness and not validity (Guba & Lincoln, 2001) (Cypress, 2017). Furthermore, this qualitative research process is driven by a research question and not by hypotheses such as in the case of hypothetico-deductive research.

This project's research approach and strategies consider the trustworthiness criteria as defined in (Guba & Lincoln, 2001) and (Anney, 2014). Added to the trustworthiness criteria, the concept of theoretical and data saturation, first introduced in the grounded theory method, that allows to determine at a point during the qualitative research process when no new data or theory are created (Saunders et al., 2017). This is an emerging and elusive concept that is difficult to apply since theoretical sufficiency can only be determined post-mortem (Sim, Saunders, Waterfield, & Kingstone, 2018). Since this project intends to serve as a starting point in a series of other research initiatives, the project does not set saturation goals. The project is set to only measure theoretical (data) saturation for the purpose of planning future work.

Table 1.1 describes the trustworthiness criteria prescribed by (Guba & Lincoln, 2001) and (Anney, 2014) to conduct qualitative research and the key design decisions made to ensure that the research process design satisfies these criteria. First of the trustworthiness criteria is the credibility criterion, which entails that the findings are considered believable by various stakeholders such as publication's editorial boards and the participants (co-researchers) to the research. This is done through thick description and by triangulation, i.e. the relative similarity of the findings using methods with different data sources such as a Systematic Literature Review (SLR) eliciting data from rigorously selected publications and a Phenomenological research method extracting data through semi-structured interviews. Secondly, the transferability criterion allows examining how the findings can be used in a specific context through use case scenarios for example. Thirdly, the dependability criterion involves an audit trail. Finally, the confirmability is established by the capacity of the research design to allow very similar findings to be produced by other researchers.

Table 1.1 Trustworthiness criteria for a dual method qualitative research (Guba & Lincoln, 2001) (Anney, 2014) and key design decisions

Criteria	**Description of the criteria**	**Key design decisions**
Credibility	The findings are construed as believable by readers.	Publications for each of the two research methods will be written using the thick description approach; The two publications will be using two different sources of research data, which are to be compared for relative similarity. Anney in (Anney, 2014) recommends at least one triangulation, ideally two. An SLR method, a standalone publication-oriented method as defined in (Fitzpatrick, Ratté, et al., 2018a) citing (Okoli, 2015). The second method used is the phenomenological research method as detailed in (Fitzpatrick, Ratté, & Coallier, 2018c) citing (C. Moustakas, 1994). The phenomenological approach uses semi-structured interviews.
Transferability	The findings are used in a specific context, e.g. use cases.	To examine how the findings can be used, in the execution of a competency question in a specific context. This project has adopted two scenarios: • Collaborative logistics planning in a (military) coalition force deployment; • Collaborative product design in Product Lifecycle Management (PLM).

Table 1.1 Trustworthiness criteria for a dual method qualitative research (Guba & Lincoln, 2001) (Anney, 2014) and key design decisions (continued)

Criteria	**Description of the criteria**	**Key design decisions**
Dependability	Criterion involves an audit trail.	Artifacts are produced to allow an auditor to verify the veracity and accuracy of the findings. Artifacts include interview recording, interview live notes, transcripts, content analysis and synthesis spreadsheets (Forero et al., 2018).
Confirmability	Capacity of the research design to allow very similar findings to be produced by other researchers.	Detailed description steps for each research method to allow another researcher to reconstitute the findings to a high degree of confidence (Forero et al., 2018).

In section 1.2, we start with the state of the art related to both research methods, i.e. the SLR and the phenomenological research. Section 1.3 provides a holistic perspective of the overall research process. Section 1.4 concludes the paper with a discussion on the research project's next steps.

1.2 State of the art

As stated in the previous section this project aims in eliciting agnostic CODPs from data model patterns. After this project, these agnostic CODPs are to be eventually axiomatized and developed as a multi-domain ontology for performing data integration. A dual method qualitative research process is proposed to perform the required elicitation of agnostic CODPs. Although no similar dual method qualitative research with the purpose of eliciting agnostic CODPs were found, related publications were extracted and examined as indicated in this section.

(Simsion, Milton, & Shanks, 2012) and (Anglim, Milton, Rajapakse, & Weber, 2009) used qualitative research approaches using interviews or surveys to acquire insight from data modeling professionals. (Anglim et al., 2009) studied the current and expected practice in data modeling. Anglim and co-authors elicited from experienced data modelers insight in respect to high-level data modeling. Their approach, with a documented method, involved semi-structured interviews. The latter research reached out to the practitioners by contacting professional associations. (Simsion et al., 2012) directly addressed the issue of the purpose of data modeling, i.e. descriptive versus design, which this project intends to explore in a future phase as a variable that may be associated with the semantic heterogeneity problem. Simsion and his co-authors also diligently documented the research method that used surveys intended for practitioners and semi-structured interviews intended to data modeling «*thought leaders*» identified by name in the publication. The research design does not explain the method to determine how the «*thought leaders*» were selected. This research attempted to identify the purpose of data modeling, either descriptive, i.e. to foster communication of requirements, to design semantic structures such as databases. Following the synthesis of the survey and interview data (Simsion et al., 2012) concluded that data modeling was better characterized as design.

In (Olivé, 2017), the author covers a new variation of the notion of ontological agnosticism, a similar concept to the multi-domain ontology. This research proposes the concept a universal ontology. This paper elicits positive and negative reactions from the scientific community in regards to an ontology that is intended to solve semantic integration, which we interpreted as semantic heterogeneity.

In respect to SLRs, only seven papers used the SLR approach on the broad subject of ontologies and were identified using the following search query in the scholar google publication database:

«allintitle: ontology "systematic literature survey" OR "systematic survey" OR "systematic literature review" OR "systematic review»

Table 1.2 summarizes the SLR's content, concentrating on the relevant material for this project. It is important to mention that none of the papers used a thick descriptive approach that would allow progressively improving the method for future usage. The project considers thick description as a crucial characteristic for qualitative research that may help future researchers to use qualitative researchers.

Table 1.2 Related SLR publications

SLR publication title	**Relevant summary**
(Blanco, Lasheras, Fernández-Medina, Valencia-García, & Toval, 2011)	Thick description of the research method, including the practical screen as recommended by (Okoli, 2015). The papers indicated that reusability was important and the abstraction quality of the elicited concepts. A light UML is used to represent the concepts.
(Hammar & Sandkuhl, 2010)	Although the central subject is Ontology Design Pattern (ODP), the purpose of the study is not to elicit ODPs but to study the motives of the primary studies.
(Subbaraj & Venkatraman, 2015)	This research described an SLR approach provided a superficial perspective on ontology based content management systems.
(Aranda-Corral, Borrego-Díaz, & Jiménez-Mavillard, 2010)	This SLR provides a framework for future research, in a similar fashion in respect to our project. The study pertains to interoperability of healthcare systems.
(Gharib, Giorgini, & Mylopoulos, 2017)	This SLR elicits privacy requirements. The SLR provides a descriptive account of the research design. Also, as performed in this project's SLR, a semantic model is provided as the result of the synthesis step.
(Setiawan, Budiardjo, Basaruddin, & Aminah, 2017)	This SLR attempts to elicit combine ontology functionality with Bayesian network to obtain a combination of logical and stochastic reasoning capabilities.
(Verdonck, Gailly, de Cesare, & Poels, 2015)	This SLR is by far the most richly described of all such reviews. While describing the research protocol in great detail, the paper also indicates validation challenges, albeit the qualitative nature of the study.

While the SLR publications, in a very small number, richly describe the practical screen step, which is used to select and extract the sought theoretical material, the actual analysis and synthesis activities were scarcely covered. In the next section, the overall design, the architecture, is examined.

As for phenomenological research methods, in respect to Information Systems (IS) (Bharadwaj, 2000) and in Information Technologies (IT) (Introna, 2005) provide insight to the use of the method. A phenomenological research method involves the individual interviews of 'first-persons', individuals that have actually participated in a phenomenon (Patton, 2002) (Tesch, 1990). The phenomenon here for this project is a multi-domain data integration capability, as perceived and lived by experienced practitioners.

1.3 Overview of the research process design

To answer the research question that pertains to eliciting agnostic CODPs to solve the semantic heterogeneity, the project is using a dual method qualitative research process. This dual method research process, while attempting to solve the problem, also intends to satisfy the trustworthiness criteria.

Figure 1.1 describes the processes performed for this project. This overview diagram illustrates using the Archimate notation (Lankhorst, Proper, & Jonkers, 2009) the business processes for the dual method qualitative research method project. Table 1.3 describes the business processes involved in the overall dual method qualitative research process.

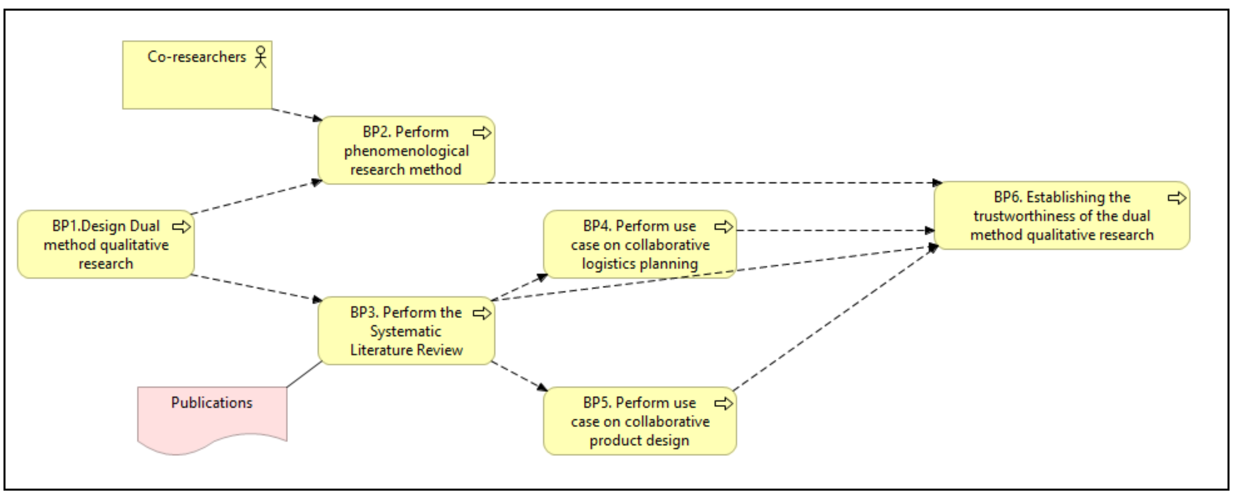


Figure 1.1 Overall business processes for the dual method qualitative research process

Table 1.3 Description of the dual method qualitative research processes

Business process name	**Business process description**
BP1.Design Dual method qualitative research	The current paper outlines the design for the dual method qualitative research process.
BP2. Perform phenomenological research method	This process elicits theoretical material through the experience of participants referred to as co-researchers. The concept of abstract, or agnostic, concepts used for data integration represents the studied phenomenon. This process and the outcome are documented in (Fitzpatrick, Ratté, et al., 2018c).
BP3. Perform the Systematic Literature Review	This process elicits theoretical material from publications selected by a query search and meeting criteria defined in a practical screen. This process and the outcome are documented in (Fitzpatrick, Ratté, et al., 2018a).
BP4. Perform use case on collaborative logistics planning	This process involves: • A literature review about military collaborative planning and collaborative logistics planning for coalition force deployment; • A literature review about interoperability ontologies for coalition force deployment; • The execution of a competency question for collaborative logistics deployment (Fitzpatrick, Coallier, & Ratté, 2018).
BP5. Perform use case on collaborative product design	This process involves: • A literature review about collaborative product design including notably Set-Based Design (SBD) and modular product design; • A literature review about interoperability ontologies for collaborative product design; • The execution of a competency question for collaborative product design in the context of PLM (Fitzpatrick, Ratté, & Coallier, 2018d).
BP6. Establishing the trustworthiness of the dual method qualitative research	This process completes the dual method qualitative research approach by providing a holistic perspective on the findings of all of the previous processes.

As described in (Fitzpatrick, Ratté, et al., 2018a) and (Fitzpatrick, Ratté, et al., 2018c). The research protocol used for both the SLR and phenomenological methods, follow the same techniques for the analysis and synthesis stages. The exceptions, i.e. the differences between the SLR and phenomenological methods, are:

- The techniques used to select the knowledge sources. In the case of the SLR, a practical screen is designed to systematically and rigorously select the publications to be studied to answer the research question. In the case of the phenomenological method, the selection criterion, for example, targeted practitioners with a minimum of eight years' experience in conceptualizing that speaks either French or English;
- The elicitation of the knowledge from the knowledge sources. In the case of the SLR, a note-taking approach allows to extract the sought concepts from publications. In the case of the phenomenological method, notes are taken and the conversations are recorded.

1.4 Conclusion and future work

The research question motivated the inquiry into the elicitation of agnostic concepts that can be used as agnostic CODPs in a multi-domain ontology. Although positivist or hypothetico-deductive criteria of validation cannot apply here in a qualitative research (Guba & Lincoln, 2001), evidences are emerging to indicate that the findings of this paper's phenomenological research method is significantly consistent, in the similarity of the findings, with two other sources: this paper's companion publication (Fitzpatrick, Ratté, et al., 2018a) and the best practice research on CODPs in (Blomqvist, 2010). This significant similarity in the outcome of qualitative research, as in the case of this project's two companion papers along with Blomqvist research on CODP best practices, is referred to as triangulation. Anney in (Anney, 2014) recommends that one or two such triangulations be demonstrated as a criterion to establish the research's trustworthiness. The authors posit that, although this is an initial phase of a multi-phase project, the outcome of this phenomenological study demonstrated a credible inductive process in eliciting data model patterns from experienced practitioners that may be considered as experts in twenty out of twenty-two individuals based on criteria established in (S. Ahmed, Hacker, & Wallace, 2005). Furthermore, the companion SLR is also followed by two use case papers: (Fitzpatrick, Coallier, et al., 2018) and (Fitzpatrick, Ratté, et al., 2018d). These use cases allow determining the transferability of the SLR. (Anney, 2014) indicates that transferability is the equivalent of positivism's generalizability criterion for qualitative research. Anney also posit that «thick description» and purposeful sampling facilitates transferability. Along with the involvement of several co-researchers in the execution of the phenomenological protocol (use of peer debriefing) (C. Moustakas, 1994) (Anney, 2014), an audit trail, thick documentation and the application of Okoli's best practice approach for conducting qualitative, this research has shown evidence of trustworthiness following the guidelines established in (Guba & Lincoln, 2001).

The authors consider that the phenomenological research method has supported quite adequately their needs for eliciting agnostic CODPs and other insights, such as prescriptive directions to eventually study design methods for multi-domain ontology based applications to resolve semantic heterogeneity. While it is expected that qualitative research protocol will predominate in this research project for some time in the future, it is conceivable that, on occasions, when sample size and other conditions are met to perform hypothetico-deductive methods that theory testing protocols may complement the current approach.

Following this phase of the project, where an SLR approach and a phenomenological research method were used, a new group of about twenty-five participants will be solicited to become co-researchers. The phenomenological research method will be executed identically to the present study without the imaginative variation technique to attempt to establish theoretical saturation. Additional semi-structured interview questionnaire, surveys and focus group sessions will be designed to further investigate some questions studied in this paper such as additional agnostic CODPs, additional domain-specific concepts, the influence of lines of business and others. This project intends to increase the size of the co-researcher group from twenty-two to approximately 100.

CHAPTER 2

AGNOSTIC CONTENT ONTOLOGY DESIGN PATTERNS FOR ENTERPRISE SEMANTIC INTEROPERABILITY: A SYSTEMATIC LITERATURE REVIEW

Daniel Fitzpatrick[1], François Coallier[1], Sylvie Ratté[1]

[1]Department of Software Engineering & Information Technology, École de technologie supérieure,
1100 Notre-Dame West, Montréal, Quebec, Canada H3C 1K3

Paper submitted for publication to the International Journal of Metadata, Semantics and Ontology in April 2018

Abstract

All organizations from the private and government sectors attempt to implement system interoperability in their information ecosystem to allow the exchange of data to solve business problems and engage in commercial opportunities. Semantic heterogeneity is the problem that affects system interoperability. Enterprises spend significant efforts and money to implement palliative measures to address this problem. No definitive solution has been and is likely to be developed in the foreseeable future. This Systematic Literature Review (SLR) intends to elicit generic conceptualization structures, language-independent semantic constructs, to solve the enterprise semantic heterogeneity problem.

This SLR intends to elicit data to construct theory around a specific type of mid-level ontology called a multi-domain ontology. The concept of multi-domain ontology has been proposed previously in (Daniel Fitzpatrick et al., 2012, 2013; D. Fitzpatrick et al., 2013). Multi-domain ontology comprises the concept of the agnostic Content Ontology Design Pattern (CODP). The agnostic CODPs form a conceptualization that intends to establish the semantics of real world concepts applicable to all industries. In this paper, such agnostic concepts are intended to be represented in a formal ontology to provide data integration functionality to all private and government sectors. This paper uses the SLR method, as a standalone research method, to elicit agnostic patterns from data models, domain models and

other types of schemas (semi-formal ontologies) usually applied in (non-cognitive) contemporary information technologies, such as relational databases. The axiomatic form of these patterns would constitute collectively the multi-domain ontology.

Keywords: Content ODP, Ontology Design Patterns, data model patterns, ontology, inference application, multi-domain ontology, systematic literature review, trustworthiness, constructivism.

2.1 Introduction

2.1.1 General Context

Semantic heterogeneity challenges affect organizations, or enterprises, in private and government sectors. This problem adversely affects the enterprise in its attempt to allow interoperability between systems required to support intra and extra organizational business processes. The scientific community has yet to propose a solution for this problem (Doan et al., 2012) (De Giacomo et al., 2018).

The information technology scientific community has conducted research notably in the development of formal ontologies for reasoning applications to resolve the semantic heterogeneity problem. Cognitive applications would perform the data integration function with the use of formal ontologies containing knowledge assertions (Bergamaschi et al., 2018) (Haziti et al., 2018). Ontology science and engineering lack the maturity to provide a coherent theoretical framework to allow truly cross-enterprise semantic interoperability solutions (Pinkel et al., 2015).

The impact of economic woes and the increase of compliance regulations render the enterprises more dependent on internal and external collaborations to cut costs, achieve their strategic objectives and fulfill their mission (Duygan-Bump et al., 2015) (De Toni, 2016). The scientific community prescribes ontology-driven integration, for the most part, to provide the required semantic interoperability. Commercial and government organizations have gained the interest to create partnerships in various domains such as medical research, fight against terrorism, law enforcement and to bail out economies on the verge of collapse. In the wake of what is now called the great recession of 2007, organizations worldwide had the need to create partnership networks notably to cut expenses and become more efficient. Free exchange of quality information and business process harmonization contribute significantly to the survival and sustainability of partnerships. These essential capabilities require that the partnerships' information systems can interoperate.

The significant pressure to reduce cost and waste motivates the organizations of all industries to internally operate more efficiently with their existing customer base. Globalization, removal of trade constraints and the evolving regulatory landscape impose further pressure notably on the service industry (Bagheri & Jahromi, 2016). Direct relationship marketing monopolizes excessively financial and other resources to maintain good relations with existing customers. Again, as in the case of partnerships, the organizations' information systems must also interoperate to allow individual enterprises to strive in retaining their customers and expand their business.

The problem of semantic heterogeneity plagues the efforts of the organizations to establish interoperability between their information systems. Semantic heterogeneity consists in having information systems each narrowly designed with semantics specific to a business domain. The problem of semantic heterogeneity also impacts multiple organizational partnerships.

2.1.2 Research Context

Despite advances made by academia in ontology engineering tool development, ontology integrative capabilities rarely contribute to knowledge discovery or any other applications in the industry. For over 25 years, the research community has conducted projects to develop machine learning capabilities based on formal ontologies, to perform data integration and thus solving the semantic heterogeneity problem. Alon Halevy, lead researcher at Google's and renowned authority on data integration, with his colleague AnHai Doan and Zachary Ives in (Doan et al., 2012) indicated that the semantic heterogeneity problem may possibly never be solved.

This paper covers a systematic literature review that is part of a project with the primary objective to elicit agnostic ontology design patterns. This project proposed an approach to perform data integration with the use of a multi-domain ontology (Daniel Fitzpatrick et al., 2012). It has first introduced the concept of multi-domain ontology in 2012 as a formal ontology that can perform data integration thus supporting interoperability between an enterprise and a group of enterprise's information systems. In the context of a group of enterprises in partnership, the multi-domain ontology's data integration capability services all of the group's information systems that are involved in the partnership agreement to interoperate.

A set of agnostic CODP composes the multi-domain ontology. A CODP pertains to the conceptualization of a specific domain (Gangemi & Presutti, 2009). Agnostic CODPs relate to concepts that apply pervasively to an entire enterprise of any industry or government sector. In the context of the data integration capability, such agnostic conceptualization constitutes a "domain", even if it encompasses all business domains in respect to the Gangemi-Presutti ODP classification, that is further detailed in section 2.2 Definition of terms.

This paper thus intends to elicit data models patterns that may eventually be re-engineered as (formal) agnostic CODPs as proposed by (Blomqvist, 2010). Furthermore (Hammar & Sandkuhl, 2010) encourages the discovery of «best practices» of patterns in data models that aim to facilitate interoperability between information systems. He also considers this field of research as immature and in need of formalization.

The SLR approach used here in this research is defined by Okoli and Schabram (Okoli, 2015)for information systems research. The SLR approach is a rigorous scientific method introduced originally in the life sciences and other mature research domains. Using quantitative research techniques, life sciences' SLRs notably apply hypothetico-deductive theory testing processes on data already collected and analyzed by a number of original studies. Inspired notably by the SLR approach for software engineering in (Kitchenham, 2004), Okoli and Schabram proposed a qualitative SLR approach suitable for this project inductive theory-building processes to unearth the sought agnostic CODPs (Okoli & Schabram, 2010). The paper analyzes selected publications published between 2009 and 2016 using a qualitative approach inspired from the Okuli and Schabram approach.

In section 2.2, we start with Definition of terms that defines the fundamental concepts relevant to this project. Section 2.3, Problem Statement, enunciates the project's primary uncertainty for which it was designed to resolve. Section 2.4 formulates the objective of this research. Section 2.5, Research Method, comprises subsection 2.5.1, Research Protocol, that describes the SLR methodology used in this paper. Section 2.6, Research Question, describes the intended inquiry at the heart of this paper. Section 2.7 describes in detail the practical screen and its two sets of criteria, the metadata level and content level criteria. Section 2.8 provides the logical query formulation. Section 2.9 provides statistics on the actual search after query execution. Section 2.10, Content Analysis, describes the findings from the systematic examination of the selected publications. Section 2.11, Content Synthesis, presents light UML (Archimate notation) diagrams with accompanying descriptions for each derived agnostic CODP. Section 2.12 concludes the paper with a discussion on the SLR's outcome and the research project’s next steps.

2.2 Definition of terms

The following definitions have been formulated based on research performed before this SLR, notably through a traditional literature review. This SLR’s authors consider these definitions necessary to establish a solider conceptual basis to this research effort.

2.2.1 Conceptualization

Conceptualization is defined here as a process that implicitly creates semantic structures. Semantic structures establish the meaning of things. Semantic structures are set of concepts, properties and their relationships. Pierdaniele Giaretta and Nicola Guarino define conceptualization as «*an intensional semantic structure which encodes the implicit rules constraining the structure of a piece of reality*» (Giaretta & Guarino, 1995). Guarino also refers to a conceptualization as an «*intended meaning of a formal vocabulary*» (Nicola Guarino, 1998).

2.2.2 Representation

It is an externalized depiction, or specification, of concepts that can be shared amongst people or machines. Representing concepts involves converting implicit concepts lodged in a person's brain into explicit concepts using a language. For example, domain ontologies that are created to share a vocabulary amongst a community are represented using one or several of the following languages: natural, concept map, SQL, XSD, OWL, etc. The represented domain ontology is submitted to the members of its community through a consensus-building process to be officially recognized and used accordingly. Nicola Guarino defines a representation or a specification of an ontology as «*a logical theory accounting*» (Nicola Guarino, 1998).

2.2.3 Ontology

The formulation of a universally accepted definition of an ontology represents in itself a problem, caused by the confusion in attempting to elicit one (Welty, 2003). Gruber defines an ontology as an «*explicit specification of a conceptualization*» (Thomas R. Gruber, 1993). Gruber's definition constitutes the most cited definition of an ontology, amid several other definitions (Guarino et al., 2009). (Guarino et al., 2009) cites (Borst, 1997) who defines an ontology as «*formal specification of a shared conceptualization*». Borst's definition entails that an ontology is formal, i.e. that it can be executed as a set of axioms in an inference engine and that it is shared, i.e. adopted consensually by a group of at least two persons, thus using a common vocabulary to communicate (Basu, 2018). Consequently, if an ontology is designed for an actual semantic application, the inherent obligation to gain consensus on the ontology's structure, instead of limiting the number of designers to an individual or a very small group of specialists, would likely caused a delay in delivering a solution (Maier & Rechtin, 2009).

This project defines an ontology, since a specification is explicit by nature, simply as a specification of a conceptualization. A data model and a domain model constitute ontologies as well (West, 2011), albeit of lower ontology level, i.e. semi-formal, as described later in this sub-section. It aims in providing a shareable and reusable knowledge to be used by people and computer systems. Ontologies would favor the trend toward a greater universal interoperability across all industries.

Conceptualization is independent of language. However, an ontology's representation is dependent on a language. An ontology is a logical theory that describes the intended meaning to its defined vocabulary, in other words, using the committed concepts to a particular conceptualization of the real world. Guarino stresses that ontologies only approximate a conceptualization. He also indicates that the only way to enhance the representation is to develop a richer set of axioms (N. Guarino, 1998). The search for a richer set of axioms explains this project's interest for data model patterns, here used interchangeably with CODPs, for multi-domain data integration developed in the industry and academia for acquiring the sought semantic richness.

All ontologies may be classified in five types:

- Top level or foundational ontologies, such as Cyc, SUMO and Proton describe some of the basic objects of reality such as time, matter, action etc. These concepts are independent of a particular problem or domain. This type of ontology supplies the fundamental concepts serving as the basis to define the other type of ontologies (Ruy, Reginato, Santos, Falbo, & Guizzardi, 2015);

- Mid-level ontologies such as the multi-domain ontology as proposed by (Daniel Fitzpatrick et al., 2012), are described by (Obrst et al., 2012) as being «*less abstract (than foundational ontologies) and span multiple domain ontologies. Mid-level ontologies also encompass core ontologies that represent commonly used concepts, such as Time and Location*». Core ontologies may be voluminous and can be more difficult to develop (Gangemi & Presutti, 2009);
- Domain ontologies represent the vocabulary of a domain, e.g. civil engineering domain;
- Task ontologies describe a generic process structure that can be used to solve a certain type of problem, such as for semantic integration described in (Calhau & de Almeida Falbo, 2012);
- Application ontologies, which describe semantic entities that stem from a domain and task ontology or ontologies, both providing a specific function context (N. Guarino, 1998).

There are essentially three types of ontology applications:

- To support the mediation between people and ontology representing a vocabulary for the exchanges between people and organizations;
- Domain interoperability: support to develop (development time application) or to operate (run time application) systems of the same or different domains;
- Knowledge reuse: requires the highest level of rigor, in addition to axioms, other concepts and their properties; ontologies for knowledge reuse will rely heavily on constraints and other types of restrictions. Problem solving methods or PSM have the capacity to support shared knowledge. They often include generic algorithms to perform various functions within the domain. Figure 2.1 illustrates a summarized definition of an ontology. One type of application that is growing in popularity in the research domain is ontology-based information extraction through natural language

processing (NLP). (Navigli & Velardi, 2008; Völker, Haase, & Hitzler, 2008; Wimalasuriya & Dou, 2010) In (Ratté, Njomgue, & Ménard, 2007), NLP processes are proposed to extract information from the organization's internal documents. These aspects constitute key elements behind the proposed reference architecture in this research project (Daniel Fitzpatrick et al., 2013). Figure 2.1 illustrates the two basic facets of the ontology concept: language dependent and language independent characteristics.

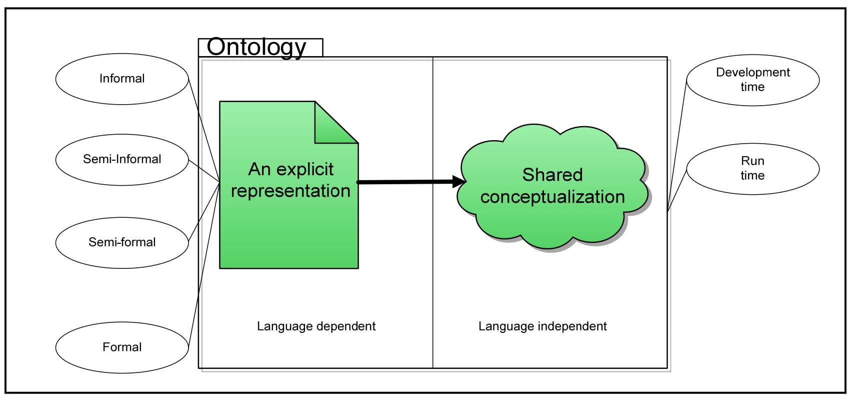


Figure 2.1 Summarized definition of an ontology

An ontology does not impose the application of properties to a given instance of a class or concept. The finality here should be to build libraries of reusable knowledge and knowledge services available on networks. Ontological commitments or agreements pertaining to classes and relationships of an ontology are discussed among software agents and knowledge bases. (T. R. Gruber, 1993). A concept definition is a human readable text that in itself provides significance, meaning therefore semantically whole (Gruber et al., 2009) (Noy & McGuinness, 2001).

An effective equilibrium must be achieved in defining ontology constrains rules in order to avoid affecting the concept abstraction level in the ontology even if it supports interoperability in a more effective manner. Affecting the ontology's abstraction level may lower the robustness and flexibility of the vocabulary (Spyns et al., 2002).

Semantic relationships are categorized as synonymy, antonymy, hyponymy, meronymy and holonymy relations. Synonymy relationships relate two similar concepts. An antonymy relation indicates opposing or disjoint concepts. The Hyponymy category pertains to a generic to specific relationship between concepts. The meronymy and holonymy relationships support the build of material structure between concepts, the former indicates that a concept is included in another one, while the latter indicates that a concept includes the object of the relationship. Figure 2.2 illustrates the conceptualization aspect of an ontology that is language independent (Lacy, 2005) (Nicola Guarino, 1998).

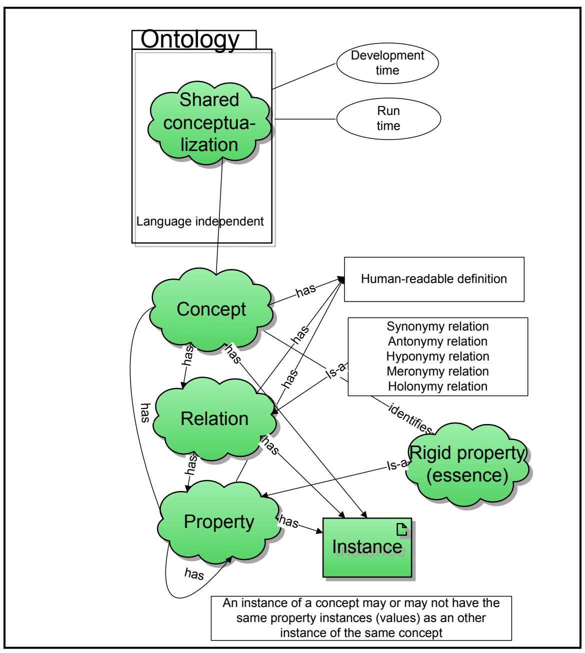


Figure 2.2 Language independent
aspects of ontologies

Ontologies can be used to solve syntactic and semantic problems, and to automate data integration. However, some of the ontologies written in specialized languages such as OWL, RDF, RDFS, PLIB and SWRL have grown to be voluminous and are becoming difficult to execute in main memory. A hybrid solution has been proposed by both academic and industrial organizations to address the in-memory loading of voluminous ontologies (Khouri & Bellatreche, 2010).

Figure 2.3 illustrates the language dependent aspects of ontologies. In terms of their level of formalism, there are: highly informal, semi-informal, semi-formal and formal ontologies.

The first level of formalism is the informal level. It refers to a natural language text. In the case of semi-informal ontology is represented as a restricted and structured form of natural language, such as a concept map. In a case of a semi-formal ontology, the vocabulary would be expressed in an artificial language such as pseudo-code or an entity relationship diagram. Finally, at the formal level, ontologies possess:

Meticulously defined terms with formal semantics, theorems and proofs of such properties as soundness and completeness, i.e. classes including property information, value restrictions, more expressivity, arbitrary logical statements, first order logic constraints between terms and more detailed relationships such as disjoint classes, disjoint coverings, inverse relationships, part and whole relationships, etc. (Xie & Shen, 2006).

Formal ontologies can be based on first-order logic, frame-based constructs or both. (A. Gómez-Pérez, Fernández-López, & Corcho, 2004; Lacy, 2005) The concept of multi-domain ontologies has been researched to facilitate the exchange of data, information and knowledge between domains (Jinxin et al., 2002).

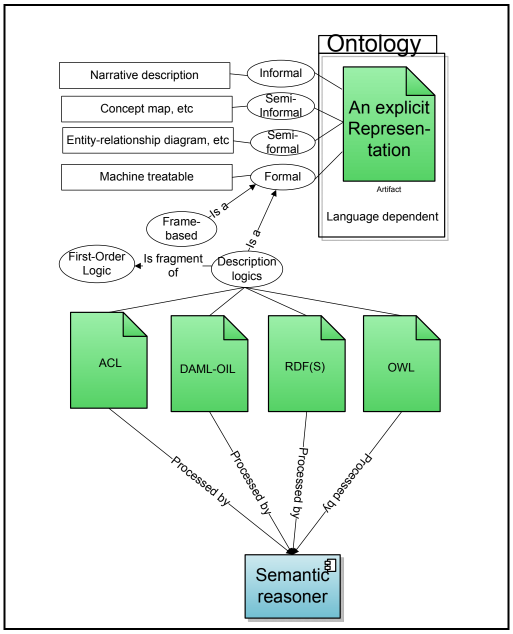


Figure 2.3 The language dependent aspects of ontologies

2.2.4 Pattern

Alexander introduces the notion of pattern in defining it as a generic solution to a recurring problem from the building architecture domain (Alexander, 1977) (Alexander, 1979). Later in 1993, the software engineering scientific community adapted the pattern concept to object-oriented design (Gamma, Helm, Johnson, & Vlissides, 1993). (Poveda, Suárez-Figueroa, & Gómez-Pérez, 2009) indicates that its fundamental meaning of a pattern pertains to something that can be imitated, that can serve as a starting point.

2.2.5 Ontology Pattern

Blomqvist defines an ontology pattern as «a set of ontological elements, structures or construction principles that intend to solve a specific engineering problem and that recurs, either exactly replicated or in an adapted form, within some set of ontologies, or is envisioned to recur within some future set of ontologies» (Blomqvist, 2010).

This project excludes ontology structure patterns since foundational concepts are excluded. Also, ontology architecture patterns are excluded since the project considers concepts and relationships other than what is found in a taxonomy (Blomqvist, 2009b). (Blomqvist, 2010) considers that ontology architecture patterns only covers the ontology as a whole or modules, but not specific concepts or relations. This SLR only covers ontology design patterns that are related to business concepts and that agnostic, i.e. applicable to any industry or domain.

2.2.6 Ontology Design Pattern (ODP)

An Ontology Design Pattern is a «*an ontology design pattern is a set of ontological elements, structures or construction principles that solve a clearly defined particular modeling problem*» (Blomqvist, 2010). It is a pattern used for the formulation of an ontology to be processed by a reasoning application. ODPs are represented as axioms in a specialized language such as OWL, a derivative of the XML language, for the purpose of logical processing. However, for the purpose of publication, an ODP can be represented in a natural language, concept map, UML, etc. This article uses the Archimate architecture modeling formalism, a simplified derivative of the Unified Modeling Language (UML), to represent the CODPs for the proposed multi-domain ontology.

2.2.7 Content ODP

According to (Gangemi & Presutti, 2009) (Blomqvist, 2009a), a content ODP, or a CODP, is a design pattern that addresses business concepts found in a domain ontology. This article represents CODPs that correspond to business concepts that are meant to be applicable to all domains.

2.2.8 Enterprise

According to The Open Group Architecture Framework (Anonymous, 2009), an enterprise can be a commercial profit driven entity, a no-profit organization or a government agency. An enterprise can also be a group of organizations such as a coalition or a partnership. A subdivision of another enterprise such as an affiliate company or department of a government can be considered as an enterprise.

2.2.9 Domain

A domain represents a community or collection of knowledge and know-how shared by a group of individuals within an enterprise, across an industry or universally (Tennis, 2003).

2.2.10 Abstract concept

An abstract concept is defined as the quality of a general concept that can be instantiated in several forms depending on a given context. In the context of this article, the sought abstract (agnostic) concepts from the elicited data model patterns can apply to any domain.

2.2.11 Agnostic concept

An agnostic concept is defined here as an abstract concept that possesses a distinct definition amongst other concepts. Thomas Erl defines the term Agnostic in the context of Service Oriented Architecture software component logic as logic that is reusable across all contexts and domains in the enterprise (Erl et al., 2017). Furthermore, it is implied here that an agnostic concept is defined in such a way that it cannot be confused with another agnostic concept.

2.2.12 Multi-domain ontology

A mid-level formal ontology composed that comprises a collection of interrelated agnostic CODPs that allows a cross-industry conceptualization (Daniel Fitzpatrick et al., 2012). Concepts related to any industry may be represented using the multi-domain ontology.

2.3 Problem statement

Semantic heterogeneity hampers enterprise application systems' interoperability. Semi-formal and formal ontology-based data integration solutions have yet to be successful and commoditized (Doan et al., 2012). Furthermore, the ontology engineering research community, albeit significant advancements that were made, still cannot consensually formulate a single unifying definition of an ontology, the prime element of a theory (Welty, 2003).

As indicated earlier in this SLR's introduction, the financial impact of this problem on the US economy (output) in 2016 was in the order of magnitude of $400 billion. This constitutes the cost of palliative measures that do not provide added business value to any aspect. Since the life sciences' research, including the medical domain, is equally affected by this problem, it is reasonable to assert that quality of life and even the capacity to preserve and save lives may also be affected by this problem. In (Laínez, Schaefer, & Reklaitis, 2012), the authors raise the issue that the pharmaceutical research domain is data rich but knowledge poor. We stipulate that semantic heterogeneity may affect the pharmaceutical research domain, notably, in its capacity to convert raw data into insight.

2.4 Research Objective

This SLR aims to elicit data model patterns from selected publications published between 2009 and 2017. The data model patterns will be re-engineered as agnostic CODPs and will compose the multi-domain ontology. Although data model patterns are only used in semi-formal ontologies, e.g. database and software design, they can contribute for building formal ontologies, such as the multi-domain ontology (Blomqvist, 2010).

This paper specifically deals with ontology patterns that can be found in the conceptualization of semi-formal ontologies, for example in an object-relational database schema or a canonical model. The sought semi-formal ontology constructs enact semantic interoperability allowing the enterprise’s application systems to work jointly intra and extra organizationally, and, will be re-engineered as agnostic CODP.

This SLR seeks to elicit existing conceptualization patterns that transcend any representation form (semi-formal vs. formal) that are domain agnostic and perhaps industry agnostic.

2.5 Research method

This SLR is based on methodologies documented in (Kitchenham, 2004), (Okoli, 2015) and (Okoli & Schabram, 2010). These guides propose an approach to plan, prepare, document and conduct a SLR for software engineering (Kitchenham) and information systems research (Okoli and Schabram). Pioneered mainly by the life sciences research domain, the SLR approach constitutes a method to produce rigorous stand-alone secondary reviews that are meant to be, as much as possible, reproducible.

SLR can be done for both quantitative and qualitative research methods types. This paper outlines a qualitative SLR based on the need to create theory about agnostic CODPs for a multi-domain ontology for performing data integration (Fitzpatrick, 2012).

2.5.1 Research protocol

Mainly inspired by (Okoli, 2015), the research protocol includes the following activities.

Previous exploratory literature survey

A previous exploratory literature survey conducted in this research project as identified conceptualization patterns in semiformal ontologies. Prior to the undertaking of this SLR, a lengthy multiyear conventional literature review was performed. Over 200 articles were found and assessed. This conventional literature review supported a qualitative research project conducted using a phenomenology method in an exploratory fashion.

As indicated in (Okoli, 2015), some steps in the research protocol, as this one, are not reproducible. The previous literature survey was performed on a broader subject, the Reference Architecture – Enterprise Knowledge Infrastructure, for which the multi-domain ontology was one of several components. This regular literature survey elicited key aspects that were used in the present SLR such as the more focused research question, the search and quality criteria and the query formulation. This particular SLR constitutes the first secondary study to elicit semi-formal data model patterns to build a formal multi-domain ontology. Since no previously published SLR with such a research objective has been found, most steps in this SLR's protocol are not reproducible, as indicated in (Okoli, 2015), except the research objective formulation, the research protocol drafting, the literature analysis and the synthesis activities. Although the guides used in this SLR do not prescribe to start with an SLR research with an exploratory literature survey, this project includes it as a necessary primer step.

Formulation of the research objective

This activity indicates the purpose of the research and is reproducible. In the context of a qualitative SLR, as it is the case here, the objective is broad (P. Leedy & Ormrod, 2012).

Formulation of a research question

As indicated by (P. Leedy & Ormrod, 2012) and (John W Creswell, 2003), a research question, not hypotheses, guides the remaining activities for a qualitative research.

Drafting the protocol

The design of the protocol for this SLR draws from (Okoli, 2015; Okoli & Schabram, 2010) for all steps of the protocol except for the Analysis and Synthesis steps. The Analysis and the Synthesis steps originate from the adapted phenomenology research method outlined in (Fitzpatrick, 2012).

Formulating the practical screen

The practical screen establishes the criteria that will allow the researcher of this SLR to select the publications that will be analyzed and synthesized. The criteria ensure the feasibility of completing the SLR by allowing a number of publications that can be read and treated by the authors. The practical screen comprises two subdivisions: metadata level and content level. The metadata level comprises any information available without actually reading the publication. The metadata level part of the practical screen allows only to either entirely reject the publication or allowing it to be further examined at the content level part of the practical screen. The content level provides the criteria that will allow the researcher of this SLR to retain and further process part or all of the content.

A key consideration that supports the necessity of a previous exploratory literature survey consists in providing this SLR's researcher with a list of publications that contained sought data model patterns. A search query too rigidly inspired from the research question would have missed too many valuable papers. However, the search query allowed too many publications that required being read and that were rejected.

Search results

The logical query defined in the previous step is executed in each of the publication databases earmarked in the practical screen. The metadata level criteria allow the retention or the rejection of publications without actually reading the content in first elimination. Once the metadata level part of the screening is completed, the retained publications' content is examined, but not analyzed, to determine if there is any material that can be used in the content of this SLR. Some publications may be rejected if no material of interest is found. All remaining publications not rejected on the metadata and content levels are then registered in the EndNote reference management software.

Content analysis

Each publication is then read for analysis. This SLR authors' previous publications are the first to be analyzed. The note-taking technique employed here consists in using Nuance Communications's Dragon Naturally Speaking dictation software where speech is converted into text and inserted in a Microsoft Word document. The extracted components are: the main agnostic concept, the subsumed subordinate concepts, the definitions and relationships. The properties, rigid properties and instances are not covered by this SLR. The documentation is segmented by publication and then by main agnostic CODP.

Content Synthesis

Agnostic CODPs found in all retained publications are then merged with same concepts that were elicited in the previous step. The documentation for the content synthesis step is segmented by agnostic CODP and represented in a simplified domain diagram where the patterns are represented as classes and not in an axiomatic form. The axes for the synthesis activity are for each CODP: the universal Thing concept, the main agnostic concept, the subsumed subordinate concepts, the definitions and relationships. Table 2.1 describes the rules used to synthesize the selected agnostic data model patterns into agnostic CODPs. These rules are based on the same rules used in this paper's companion publication that uses a phenomenological research method to also elicit agnostic CODPs for a multi-domain ontology. The ontology elements and structures are considered as meaning units as in the phenomenological approach. And as in the phenomenological research method, the semantic material extracted in this SLR is coalesced using the described rules.

Table 2.1 Rules to synthesize data model patterns into agnostic CODPs, based on (Fitzpatrick, Ratté, et al., 2018c)

Meaning unit number	**Meaning unit type description**	**Meaning unit coalescence rule description**
1	The agnostic concepts.	Concepts defined in the same manner are retained if it was identified by at least two publications; • In the case of synonyms, only the term with the greatest selection by publications is retained. In case of equal number of selections, the researcher makes the final decision; • In the case of concepts that have been defined in more than one way, the same rule as in the case of synonyms applies.
2	The subsumption and other relationships between the agnostic concepts.	• The relationships need to be selected only once to be retained; • In case of conflicting relationships, only the one with the greatest number of selections is retained.

Table 2.1 Rules to synthesize data model patterns into agnostic CODPs, based on (Fitzpatrick, Ratté, et al., 2018c) (continued)

Meaning unit number	Meaning unit type description	Meaning unit coalescence rule description
3	The definition or description of the agnostic concepts.	The texts are integrated by the researcher.
4	The de facto agnostic CODPs derived for the above-mentioned meaning units.	The aforementioned meaning units are then integrated in distinct modules using the SLR's module structure as a starting point. The SLR's architecture module's structure may be modified during the synthesis step to adapt to the emerging agnostic CODPs.

2.6 Research question

(N. Guarino, 1998) stresses that ontologies only approximate a conceptualization. He also indicates that the only way to enhance the representation is to develop a richer set of axioms, which are derived from concepts. Guarino stipulated that conceptualization is language-independent. This project posits that the elicitation of richer concepts, albeit being light ontological structures, as ontology design patterns, and their conversion into axiomatic rules or axioms as proposed by (Blomqvist, 2009b), would enhance the use of inference engine technologies described notably by (McGuinness & Da Silva, 2004). Data integration, also referred to as semantic data integration by (De Giacomo et al., 2018), represents a potentially effective application for ontology-based inference technologies. As proposed by (Daniel Fitzpatrick et al., 2013), a multi-domain ontology would leverage agnostic design patterns, based on semi-formal ontologies, to perform data integration and resolve the semantic heterogeneity problem. The research question's formulation intends to be accurate in relation to the desired results, i.e. the set of publications that will be filtered and examined (Okoli, 2015).

For this SLR, the research question is formulated by the following:

What are the conceptualization patterns found in semi-formal ontologies, e.g. data model patterns, software engineering patterns, etc and that can be agnostic to any domain or industry sector in the context of enterprise semantic interoperability and can be used as the basis of agnostic CODPs to resolve semantic heterogeneity in enterprise systems?

This research question guides the design of this project's qualitative research approach. The research question serves as the foundation of the search query for this SLR. During the pilot phase for this Project, a query statement written closely as the research question is executed first. The query formulation is progressively phrased in a manner that it identifies a minimal set of publications previously reviewed during the conventional literature survey performed in the early phase of the Project.

2.7 Practical screen

The criteria for the practical screen are grouped in two categories: metadata level and content level. The metadata level criteria, in Table 2.2, are used when the researcher of this SLR examines the general information made available by the publisher of the publication without actually reading the content, such as the title, abstract, keywords, etc. The content level criteria, in table 2.3, require the researcher of this SLR to visually scan part of or the entire publication. Some criteria may be used in both practical screen levels. The practical screen constitutes a subjective topic in the SLR and is not reproducible (Okoli, 2015).

Table 2.2 Metadata level criteria

Name of criterion	Description of the criterion
Ontology level	Only semi-formal domain ontologies are sought for this SLR. Research publications that pertain to formal ontologies will be discarded.
Ontology type	Data models that do not pertain to business concept domains are not retained.
Publication language	Only publications written in English and French will be retained.
Publication year	Publications are retained only if they were published after 2009 inclusively and before 2018.
Publication types	All scientific and industry peer reviewed publications are eligible to be selected. PhD theses are also to be considered. Masters theses are not to be retained.
Authors	The publications written by this SLR's authors will be retained regardless of that the practical screen identifies them or not. Such self-reference criterion is noted in Okoli's guide (Okoli, 2015).
Research source libraries	scholar.google.com, IEEEXplore, ACM Digital library, Springer Link, Web of Science, Scopus, Science Direct, Compendex & Inspec.
Study type	Only primary studies are to be considered in this SLR. Other SLRs are to be excluded.

Table 2.3 Content level criteria

Name of criterion	Description of the criterion
Ontology level	Only semi-formal domain ontologies are sought for this SLR. Content that pertain to formal ontologies will be discarded.
Agnostic business concepts	Only business concepts that can be used in any industry domain, e.g. Financial, Retail, Government and others can be considered.
Industry specific (low abstract) concepts	Low abstract business concepts are to be retained only if they are associated to agnostic business concepts.

2.8 Logical query formulation

This search query is specifically designed to answer the research question to extract conceptualization patterns from semi-formal ontology primary studies. The following search query is adapted only for the time period between 2009 and 2017 inclusively. During the Project’s the pilot phase, a query statement closely formulated as an abridged version of the research question is first tried without any returns from the source libraries. The query formulation is then diluted by trial and error until several publication previously identified in the standard literature survey and considered by the researcher as essential were returned.

A preliminary research of the selected publication sources indicates that there are fewer than a dozen systematic literature reviews that contain the term ontology in the title for all year at the writing of this SLR. The query to find existing systematic reviews was submitted in Scholar Google as:

> allintitle: ontology "systematic literature survey" OR "systematic survey" OR "systematic literature review" OR "systematic review"

A total of five publications are identified by the search, excluding irrelevant papers: (Hammar & Sandkuhl, 2010), (Subbaraj & Venkatraman, 2015), (Diaz, Antonelli, & Sanchez, 2017; Setiawan et al., 2017; Verdonck et al., 2015). None of the latter publications intended to elicit ontology patterns or any other form of patterns. The novelty of systematic literature reviews, let alone for qualitative research, for ontology research explains the small number of such publications. The logical query is formulated in a form that can be adapted in the selected research source libraries as listed in the practical screen found in section 2.7:

> enterprise "patterns of data modeling" OR "pattern of data modeling" OR "data model pattern" OR "semantic pattern" OR "class model pattern" OR "data vault model" "data model"

As (Okoli, 2015) indicates, this step is not reproducible. In the case of this project, it was developed over time through a traditional literature survey with very few examples, i.e. systematic literature reviews on ontology development, to inspire from as indicated in this section.

2.9 Search results

The statistics in Figure 2.4 show the total number of publications displayed after executing the search query in all research source libraries from 2009 through 2017 inclusively. The search query listed a total of 860 publications from the source libraries prescribed in the practical screen over nine years.

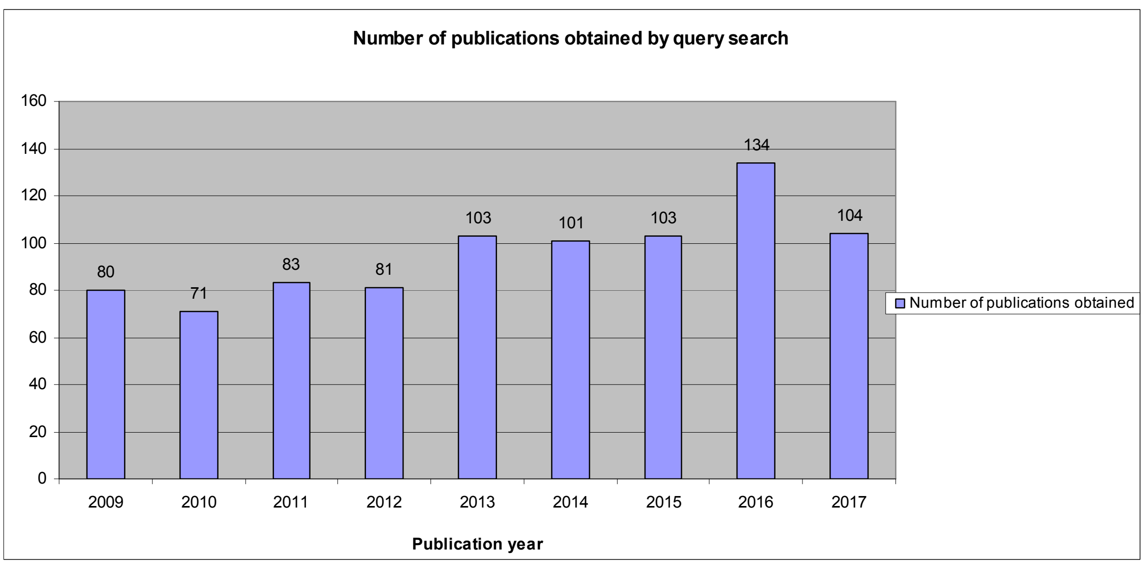


Figure 2.4 Number of publications per year returned and scrutinized

Figure 2.5 shows 69 papers, or eight percent of the 860 returned publications from the query, retained publications for analysis and synthesis once the filtering criteria are applied. As established in the metadata level criteria of the practical screen, this SLR's authors' publication, (Daniel Fitzpatrick et al., 2012) are included in the statistics although being elicited in the query. The small number of publications that were finally retained can be explained mainly by publications that treated the matter regarding data model patterns without actually showing any.

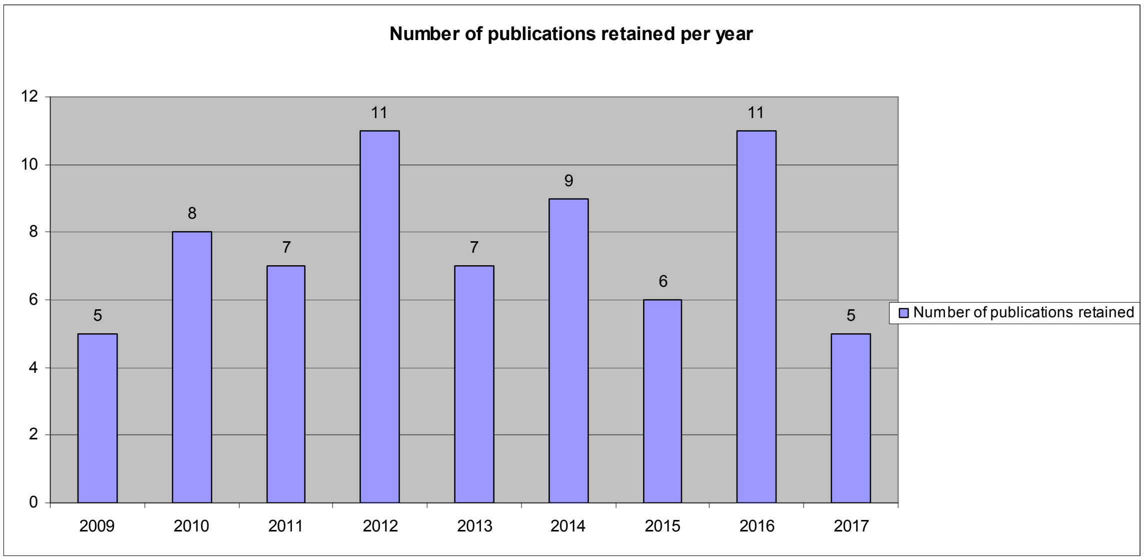


Figure 2.5 Number of publications per year screened and retained

Following the search results, the publications are a studied more in-depth for agnostic data model patterns, i.e. that can be used in any private industry or government sector. The analysis step elicits agnostic concepts, their relationships and definitions, isolating these elements from the rest of the text. While not in an axiom format, these semantic elements, or meaning units, are ultimately integrated in the synthesis step.

2.10 Content analysis

The analysis of the elicited publications breaks down the sought material in the following components: the main agnostic concept, the subsumed subordinate agnostic concepts (if any) and the definitions. The publications that are considered of greater interest, which contain a complete data model or that contain a greater number of agnostic concepts, are covered in this section in more depth at the beginning and summarized in tables 2.3, 2.4 and 2.5. Then Table 2.6 shows the other remaining analyzed publications with main and subordinate agnostic concepts.

The first papers analyzed are some of this SLR's authors' previous publications, i.e. (Fitzpatrick, 2012; Daniel Fitzpatrick et al., 2012, 2013; D. Fitzpatrick et al., 2013). These publications cover research performed on the concept of Reference Architecture – Enterprise Knowledge Infrastructure (RA-EKI). RA-EKI defines processes, data structures and ontologies to produce knowledge, actionable information, and know-how, functional knowledge. It proposes an assembly line like epistemological approach to convert data into information, then information into knowledge and know-how. Knowledge and know-how are stored and executed from an ontological structure composed notably of the multi-domain ontology, a contribution of this project. These publications, while describing RA-EKI, also provided the following descriptions of agnostic concepts in Table 2.3. Only concept names and descriptions are provided. This set of agnostic concepts and the multi-domain ontology architecture modules serve as the foundation, the starting point, for the content synthesis process. The RA-EKI multi-domain architecture modules used for grouping the elicited agnostic CODPs are shown in figure 2.6.

Table 2.4 Elicited agnostic concepts from this SLR's author previous papers

Name of concepts	Description of concepts
Party	A person or an organization. Also covers the notion of a taxonomy of persons and organizations and groupings to represent the composition of a group of people into organizations.
Product	A good or service resulting from a process. The UN PCS and NAPCS classification schemes can notably be used as taxonomies for products. The concept of a bill of material allows to package products.
Contract	A tacit agreement between parties.
Tariff	Covers the notion of price, rates, etc.
Event	A spatiotemporal object in the form of a change of state affecting a thing.
Document	A physical or electronic text material containing information.
Identity	A mechanism to distinguish two instances of the same class. This includes means of identifying persons such as the licence number, employee number, etc.
Infrastructure	A human made work such as buildings, roads, railroad, etc.
Financial	Includes the notions of transaction, account, instruments, etc.
Technology	A subclass of products consisting of man-made electronic and mechanical devices.
Strategy	A subclass of process specially designed to achieve a goal.
Network	A Petri-like structure composed of two non-segments for the purpose of transport of: energy, cargo, people, voice, data, etc.
Context	A set of things such as location, parties, products and events that may influence the use of vocabularies, chain of future events, etc.
Concept	An imaginary man-made construct that corresponds to real life imaginary or physical things.
Process	A unit of work in which resources are used resulting in the fabrication of goods or the rendering of services. A process can be performed by humans, by nature or a mix of both.
Location	A concept related to a coordinate system such as Earth location systems. This includes the notion of areas, segment and grid location. Geography also includes the notion of street addresses and electronic locations such as email and IP addresses.
Inventory	A set of specified goods or services, stored or offered at a given location.

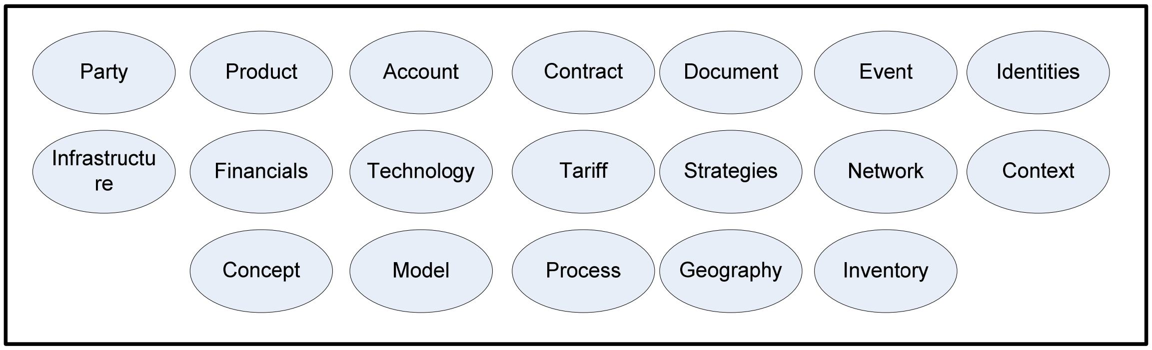


Figure 2.6 The RA-EKI ontology architecture modules
(D. Fitzpatrick et al., 2013)

The book of Matthew West, (West, 2011), covered a data model that the researcher of this SLR considers the most conceptually similar to the multi-domain ontology in terms of agnostic concepts and completeness. (West, 2011) proposed the High Quality Data Model (HQDM) approach. The author describes the notion of data model quality as accurate and reusable semantics.

This model is inspired by the ISO 10303 standard, informally known as Standard for the Exchange of Product model data or STEP (Pratt, 2005), a process industry standard. This model is also inspired by the ISO 15926 standard about lifecycle integration of process plant data including oil and gas production facilities (Leal, 2005).

West defines an integration data model as: «*a data model that integrates a number of separate applications*». It allows semantic interoperability between an enterprise's systems and between enterprises, for example, to support supply chain processes. The HQDM model can be used for any operational or transactional application system. West considers that data model concept definition should be expressed as real world concepts. Data model definitions should not be associated with adapted database artefacts. Table 2.4 illustrates elicited the concepts, definitions and relationships that meet the content level criteria from the practical screen.

Table 2.5 Elicited concepts from (West, 2011)

Name of concepts	**Description of concepts**	**Relations between concepts**
Activity	A thing that involves participating things. It causes at least one event, usually to a starting and ending event.	An activity causes an event.
Thing	A piece of reality or of the imagination.	A role is-a thing.
Event	A spatiotemporal thing of zero duration.	A event has a location.
Transaction	Represents an execution of an activity. A transaction records a party's business.	A party has transactions.
Role	A class of participating things in which each instance is involved in the same way in an activity or an association.	A thing (except Role) that participates in an Activity.
System	A physical object that is composed of physical objects. A system may be a functional system, a biological system, etc.	A collection of things is a system.
Person	A biological system of the human species.	A Person is a Party.
Employee	A person that is employed by an organization.	An employee is-a role.
Organization	A party that is a body of people.	An organization is-a party.
Party, Party Type	A system that is either a person or an organization.	A Party is a system.
State of Party	A temporal part of a party.	A party has a state of party.
Period of Time	A state that is a temporal part of some possible world.	Event is in a period of time.
Employee	A person that works for an organization.	An employee is a role played by a Person.
Position	A component of an organization occupied by a person usually at a given time.	A Position is an Organization.
Asset	A participant that plays the role of being owned in an ownership relationship.	An Asset is a Role.

Table 2.5 Elicited concepts from (West, 2011) (continued)

Name of concepts	**Description of concepts**	**Relations between concepts**
Product	A tangible good such as oil. A product can also be a generic class, such as a car model, and not its instance class, such as a specific car identified by a vehicle identification number. A product can have a role of being offered or sold. A brand is a type of product as well. An instance of a product can also be defined as a product offering. A product offering can be sold at a price and at a given location, through a sales channel, and for a period of time.	A Good is a Product. An Offered Product is a Role. A Sold Product is a Role. A Brand is a Product. A Product can have a Price.
Brand	A named instance of a product.	A brand is-a product.
Offer	A socially devised activity that leads to an exchange of a thing.	An Offer is a Thing.
Plan	A possible world that a party intends to make happen.	A plan involves a party.
Requirement	A spatiotemporal object that is a part of a plan has at least one intended state.	A Requirement is part of a Plan.
Price	An amount of money used to sell a product.	A Product can have a Price.
Currency	A class of money that is issued by an authority party.	A transaction has a currency.
Sale	A process agreed by parties where goods and money are exchanged.	A Sale is a Process.
Agreement	A course of action, or process, determined by two or more people.	An agreement involves roles.
Contract	A type of agreement that involves obligations typically in an exchange of goods or services for assets, usually money.	A Contract is an Agreement.

A book written by Michael Blaha (Blaha, 2010b), also covered a data model composed of a set of archetypes, i.e. patterns that the authors of this SLR considers similar to the multi-domain ontology in terms of agnostic concepts. (Blaha, 2010b) proposed a set of archetypes that «*are abstractions that often occur and transcend individual applications*». These agnostic concepts are listed and defined in Table 2.6.

Some of the archetypes contained in Michael Blaha's do not meet the agnostic requirement from the practical screen. For example, the concept of Course contained in Blaha's set of archetypes can be abstracted as a Service and is considered as a low abstract concept. The same applies the archetype Flight that can be abstracted as a (airline) Network segment, operated by an airline company through a service.

Table 2.6 Elicited concepts from (Blaha, 2010b)

Name of concepts	**Description of concepts**	**Relations between concepts**
Account	A Thing that «is a label for recording, reporting, and managing a quantity of something. The following are types of accounts: accounting account, service accounts, computing accounts, customer loyalty account».	Accounting Account is an Account. Service Account is an Account. Computing Account is an Account. Customer Loyalty Account is an Account.
Address	A mechanism to ensure communication between actors. May include postal address, email address, phone number, URL, etc.	A Postal Address is an Address. An Email Address is an Address. A Phone Address is an Address.
Asset	A thing that represents something having a value for an actor.	An Actor has an Asset.
Contract	An agreement to ensure the provisioning of products.	A Contract is an Agreement.
Customer	A role that can be played by a person or an organization.	A Customer is a Role.
Document	A physical or electronic representation of a body of data in a context.	A party play a role in a document.
Event	A (Thing) «that is an occurrence at some point in time».	A party plays a role in an event.
Item	A part of a Product.	An Item is part of a Product.

Table 2.6 Elicited concepts from (Blaha, 2010b) (continued)

Name of concepts	**Description of concepts**	**Relations between concepts**
Location	A Thing that represents a spatial object, i.e. a place on the globe or elsewhere.	A location is-a thing.
Opportunity	«*An inquiry that can result in business. Opportunities often arise in the context of sales*».	Party plays a role in an opportunity.
Part	An individual good that can be counted and described.	A Part is a Good.
Payment	A transfer of money done against the supply of goods or services.	A payment is-a transaction.
Position	A job occupied by a person in an organization.	A Position is occupied by a Person. An Organization has a Position.
Product	A package that contains items for a particular marketplace.	A Product contains Item.
Role	A function performed by a thing.	A Role is performed by a Thing.
Transaction	An exchange that must be done completely, mostly in finance and computing.	A transaction is-a thing.
Vendor:	A person or organization that provides a product to a customer.	A Vendor is a Role.
Identity	A means that allows to distinguish two instances of the same class.	An identity is-a role.
Name	A single word or sentence that attempts to distinctively identify a thing in the context.	A Name is an Identity.

The remaining extracted publications' analysis is summarized in Table 2.7 with the name of the main and subordinate concepts and the reference to the publications. The analyzed papers are associated for each elicited concept. The actual semantic material is broken down in a spreadsheet.

Table 2.7 Summary of the analysis of the remaining retained publications

Name of main concepts	**Name of subordinate Concepts**	**Publications**
Party	Person, Organization, Organization Unit, Company, Government, Government Agency, Society, Company	(Lubyansky, 2009; G. Piho, Tepandi, Parman, & Perkins, 2010), (Xi & Hongfeng, 2009), (Gunnar Piho, Roost, Perkins, & Tepandi, 2010), (Azizah, Bakema, Sitohang, & Santoso, 2009), (Luttighuis, Stap, & Quartel, 2011; Pfeiffer & Wąsowski, 2011), (Hofreiter, Huemer, Kappel, Mayrhofer, & vom Brocke, 2012), (Henderson-Sellers, Low, & Gonzalez-Perez, 2012), (Debruyne & De Leenheer, 2013), (Mamayev, 2014), (Collins, Hogan, Shibley, Williams, & Jovanovich, 2014), (Aibdaiwi, Noack, & Thalheim, 2014), (Frosch-Wilke & Scheffler, 2015), (Ptitsyn, Radko, & Lankin, 2016), (Ruan et al., 2016), (L. González, Echevarría, Morales, & Ruggia, 2016)
Product	Order, Product Item, Part, Service, Equipment, Vehicle, Order, Product Type, Order Line, vehicle, Product Type, Bill of Material (BOM), Brand, Electronic Equipment, Device	(G. Piho et al., 2010), (Sesera, 2011), (V Jovanovic & Pavlic, 2011), (Blaha, 2010a), (Van Grootel, Spyns, Christiaens, & Jörg, 2009), (Azizah et al., 2009), (Currim & Ram, 2010), (De Leenheer, Christiaens, & Meersman, 2010), (Pfeiffer & Wąsowski, 2011), (G. Piho, Tepandi, & Parman, 2012), (Blaha, 2013), (Vladan Jovanovic, Subotic, & Mrdalj, 2014), (Delfmann, Breuker, Matzner, & Becker, 2015), (Frosch-Wilke & Scheffler, 2015), (Puonti, Raitalaakso, Aho, & Mikkonen, 2016), (Zhao et al., 2017), (Kozmina, Syundyukov, & Kozmins, 2017)
Agreement	Contract, Service contract, Contract type	(Xi & Hongfeng, 2009), (West, 2011), (Sesera, 2011), (Knowles & Jovanovic, 2013), (Mamayev, 2014)
Price	Associated Fee, Rate Package, Book Rate	(Sesera, 2011), (Vladan Jovanovic et al., 2014)

Table 2.7 Summary of the analysis of the remaining retained publications (continued)

Name of main concepts	**Name of subordinate Concepts**	**Publications**
Event:		(Poels, Maes, Gailly, & Paemeleire, 2011), (Van Grootel et al., 2009), (De Bruyn, Van Nuffel, Verelst, & Mannaert, 2012), (Henderson-Sellers et al., 2012), (Laurier & Poels, 2012), (Camossi, Villa, & Mazzola, 2013), (Molnár & Benczúr, 2015)
Document		(Blaha, 2010a), (Mamayev, 2014), (Molnár & Benczúr, 2015)
Identity	Name, Identifier	(Silverston & Agnew, 2011), (West, 2009), (Blaha, 2010a), (Vladan Jovanovic & Bojicic, 2012)
Financial	Transaction, Transaction Type, Payment	(Sesera, 2011), (Poels et al., 2011), (Laurier & Poels, 2012), (Blaha, 2013), (Athenikos & Song, 2013), (Giraldo, España, Pineda, Giraldo, & Pastor, 2014), (Z. Ahmed, Arif, Ullah, Ahmed, & Jabbar, 2016)
Context	Contextual role	(De Leenheer et al., 2010), (Silverston & Agnew, 2011), (Luttighuis et al., 2011), (Stirna & Sandkuhl, 2014), (Tiwari & Thakur, 2015), (Serbanescu, Azadbakht, Boer, Nagarajagowda, & Nobakht, 2016), (Serbanescu et al., 2016)
Network	Node Type, Edge Type	(Blaha, 2010a)
Concept		(Z. Ahmed et al., 2016)
Process	Rules, Analysis Process, Quality Control, Testing, Task	(G. Piho et al., 2010), (G. Piho & Tepandi, 2013), (De Leenheer et al., 2010)
Location:	Point, Curve, Surface	(Wannous, 2014)
Inventory		(G. Piho et al., 2010), (Athenikos & Song, 2013)
Unit of Measure	Quantity, Measure	(G. Piho et al., 2010), (Gunnar Piho et al., 2010), (Frosch-Wilke & Scheffler, 2015)
Account	Account Type, Account contract	(Sesera, 2011)

Table 2.7 Summary of the analysis of the remaining retained publications (continued)

Name of main concepts	**Name of subordinate Concepts**	**Publications**
Role	Customer Product, Channel, Resource, Contextual Role, Contact Mechanism, Party Role, Name	(Sesera, 2011), (Poels et al., 2011), (G. Piho & Tepandi, 2013), (De Leenheer et al., 2010), (Silverston & Agnew, 2011), (West, 2009)
Asset		(Lubyansky, 2009)
Resource		(Bergholtz, Andersson, & Johannesson, 2010)
Requirement		(Khouri, Bellatreche, & Marcel, 2011)
Rule	Business Rule	(Silverston & Agnew, 2011)

The Content Analysis step involved the survey, in order, of this SLR authors' publications and of two publications, specifically elicited through the practical screen, that contain complete or near complete set of data model patterns, from Matthew West (West, 2011) and Michael Blaha (Blaha, 2010b). Tables 2.3, 2.4 and 2.5 list the sought agnostic concepts along with definitions. The remaining extracted publications covered a relatively small number of primary agnostic concepts. Table 2.6 identifies the primary and secondary agnostic concepts elicited along with the source publications. The content analysis of the 69 retained publications identified a total of 246 agnostic concepts. Table 2.8 lists the twenty agnostic concepts that were the most elicited in this SLR, the top twenty selections, and the number of papers that covered them as data model patterns.

Table 2.8 Top twenty agnostic concepts
elicited in this SLR

Name of the top twenty agnostic concepts	Number of the top twenty selections
Product	18
Customer	13
Person	13
Party	12
Role	12
Event	11
Location	11
Resource	9
Organization	8
Contract	7
Process	7
Service	7
Supplier	7
Time	7
Address	6
Context	6
Country	6
Employee	6
Order	6
Part	6

In the next step, all the agnostic concepts, their relations and definitions are consolidated across all retained publications. The synthesis step processes the agnostic concepts from the 69 retained publications first in order of publication years, in order of source libraries as listed in the practical screen in section 2.7 and in order for which the publications are analyzed. In the next step, all of the ontological elements and structures are merged as light-weighted CODPs. (Blomqvist, 2010) describes these light-weighted CODPs as «*not heavily axiomatized, but provide just a bit of formal semantics*». The agnostic concepts, relationships and definitions are consolidated across all retained publications. The synthesis step processes the agnostic concepts from the 69 retained publications first in order of publication year, in order of source libraries as listed in the practical screen in section 2.7 and in order for which the publications are analyzed.

2.11 Content Synthesis

Following the analysis performed in the previous section, the main agnostic concepts, the subsumed subordinate concepts, the definitions and relationships are synthesized starting with the material extracted from the SLR authors' previous publications: (Fitzpatrick, 2012; Daniel Fitzpatrick et al., 2012, 2013; D. Fitzpatrick et al., 2013). The synthesis process selects agnostic concepts that were elicited at least two times. It is important to note that, although (Okoli, 2015) considers this step as irreproducible, the synthesis of agnostic concepts and relationships reveals that part of the synthesis step may reproducible by involving a different researcher to perform this step.

The study of the saturation points is based on this Project's phenomenological research method. Originally created for the grounded theory method, the concept of theoretical saturation acquires popularity with other qualitative research methodologies in IT and social sciences (Marshall et al., 2013) (Saunders et al., 2017) (Sim et al., 2018). Figure 2.8 shows the number of saturation events, or saturation points, which are when an agnostic concept is selected a second time and becomes part of an agnostic CODP. The publications are ordered by publication years, by source libraries as listed in the practical screen and then by processed order. Since a minimum of two selections are needed for an agnostic concept to be retained, no saturation event is identified on the first publication. The publications are aggregated by a group of five papers for the purpose of the diagram in figure 2.7.

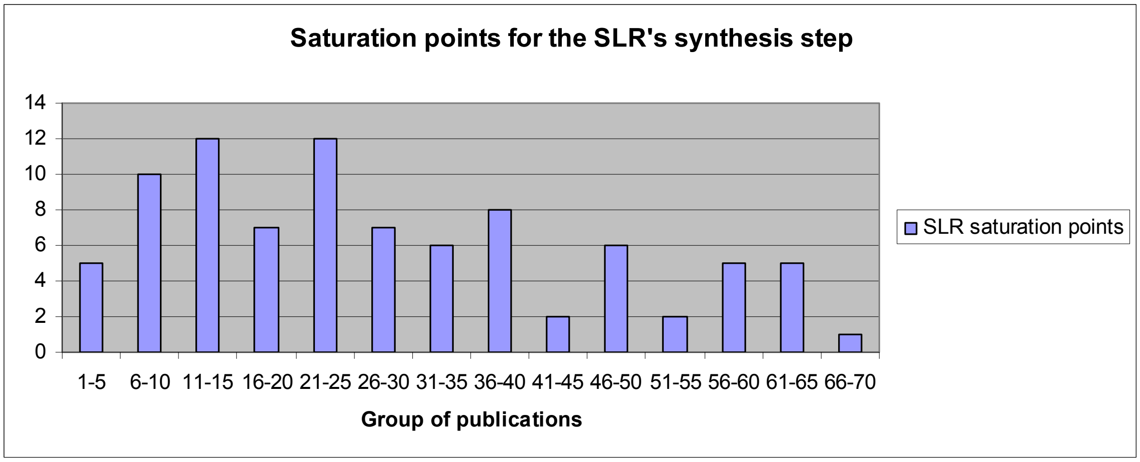


Figure 2.7 Saturation events in the SLR synthesis step

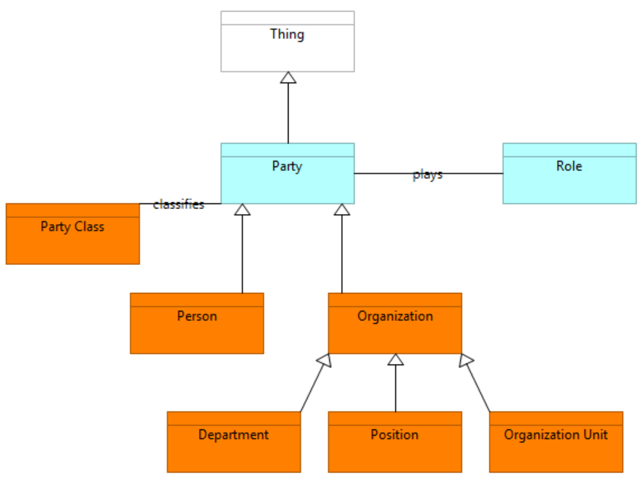
The saturation points diagram in figure 2.7 shows a downward trend that may demonstrate saturation. At this point, it would be illogical to expect a definitive state of saturation, especially in an exploratory research project. This saturation condition and the decision on how to treat with the very notion of theoretical saturation is to be re-examined in a future continuation of the SLR approach, as suggested by (Saunders et al., 2017). The elicited CODPs are represented in diagrams using the Archimate notation (Lankhorst et al., 2009). This notation standard meets the requirements to model CODPs at its appropriate level. Each agnostic CODP is described in a format inspired from a CODP template proposed in (Gangemi, Gómez-Pérez, Presutti, & Suárez-Figueroa, 2007). The agnostic concept Thing, present in all CODPs, is defined here as an element of reality or of the imaginary. This SLR also revealed that each of the 89 agnostic concepts are selected in average by approximately 6 publications each. Furthermore, a concept reaches a saturation point in average at around the 28th publication. Finally, 90% of the 89 agnostic concepts have reached their saturation point at the 59th publication, which may be indicative that the elicitation may reach a turning point.

Following the synthesis step, the resulting meaning units, the agnostic CODPs, are represented using the Archimate Open Group notation standard, a lighter form of UML (Lankhorst et al., 2009). As in the phenomenological research method performed in this project (Fitzpatrick, Ratté, et al., 2018c), each agnostic CODP is documented using a CODP template proposed in (Gangemi et al., 2007). The root entity is the main agnostic concept that bears the same name as the module.

2.11.1 The Party agnostic CODP

The Party CODP allows conceptualizing people and organizations as represented in table 2.9.

Table 2.9 SLR study Party CODP

Ontology Pattern Type	***Content Ontology Design Pattern***
Name	Party
General description	The Party CODP allows the conceptualization of the nature of a person and an organization.
Examples	• Any physical person regardless of what role or roles may be played, e.g. John Doe; • A private corporation, a job position, a government agency, a government as a whole, an informal group, a family.
Simplified UML diagram (Archimate)	
Definitions of the agnostic concepts	• Party: A thing that is either a person or an organization; • Party Class: A classification scheme for parties; • Person: A biological thing classified as a Homo Sapiens; • Organization: A group of persons; • Role: See the Role CODP.

2.11.2 The Product agnostic CODP

The Product CODP covers the goods and services that result from processes as illustrated in table 2.11. It includes the notions of classification and Bill of Material.

Table 2.10 SLR study Product CODP

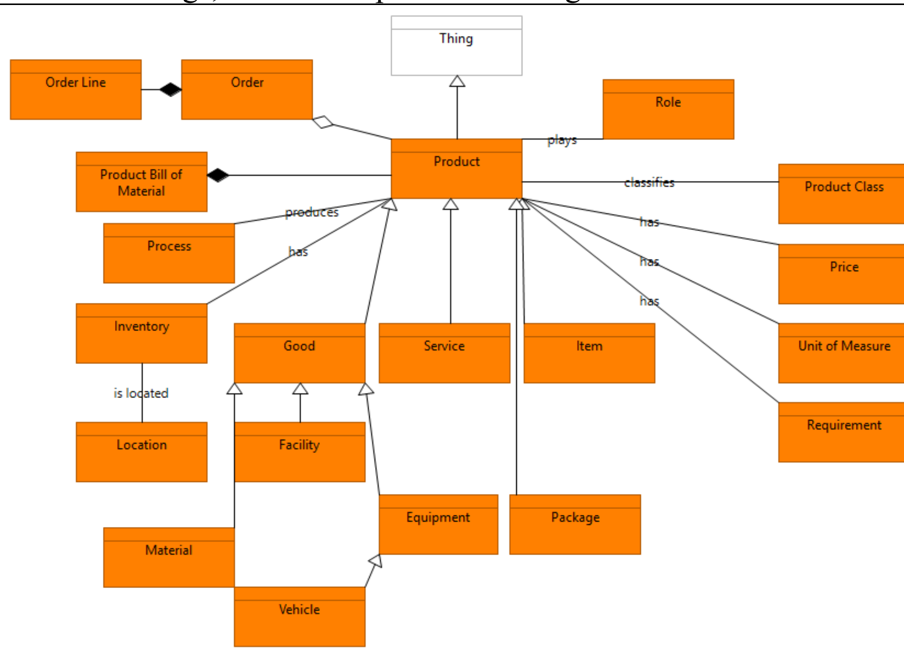
Ontology Pattern Type	***Content Ontology Design Pattern***
Name	Product
General description	A good or service resulting from a process. The UN PCS and NAPCS classification schemes can notably be used as taxonomies for products. The concept of bill of material allows to package products.
Examples	• Goods are tangible products such as automobile, an electronic equipment, salt, fuel; • Services are intangible services such as car rental, banking offerings, investment portfolio management.
Simplified UML diagram (Archimate)	

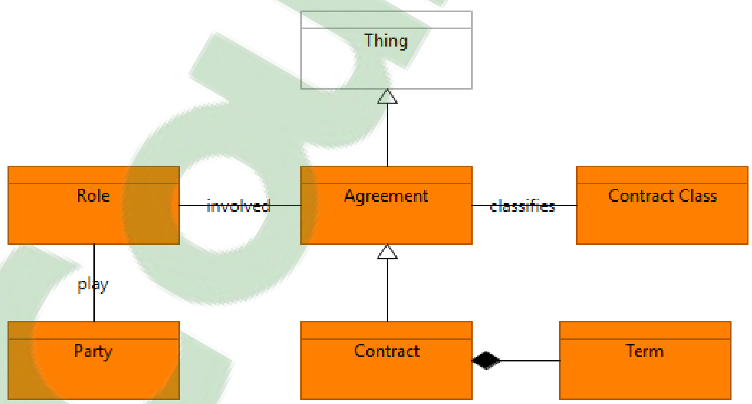
Table 2.10 SLR study Product CODP (continued)

Ontology Pattern Type	***Content Ontology Design Pattern***
Definitions of the agnostic concepts	• Product: A tangible good or an intangible service produced by a process. A product may be a grouping of other products or can be parts, which are also products; • Product Class: A classification scheme for products; • Order: Request for the fulfillment of a service or to supply goods; • Product bill of Material: A grouping, or packages, of products, that may be a product itself; • Inventory: A specification of goods or services stored or offered at a given location; • Good: A tangible product such as equipment, etc.; • Service: An intangible offering providing value to a consumer; • Brand: A factor of differentiation associated to a good or service for the benefit of a consumer; • Infrastructure: A human made thing such as buildings, roads, railroad, etc.; • Unit of Measure: A standard for establishing the quantity of a thing, e.g. Currency, weight, height, etc.; • Role: See the Role CODP; • Location: See the Location CODP; • Process: See the Process CODP; • Price: See the Price CODP.

2.11.3 The Contract agnostic CODP

The contract CODP covers any form of tacit agreement between parties.

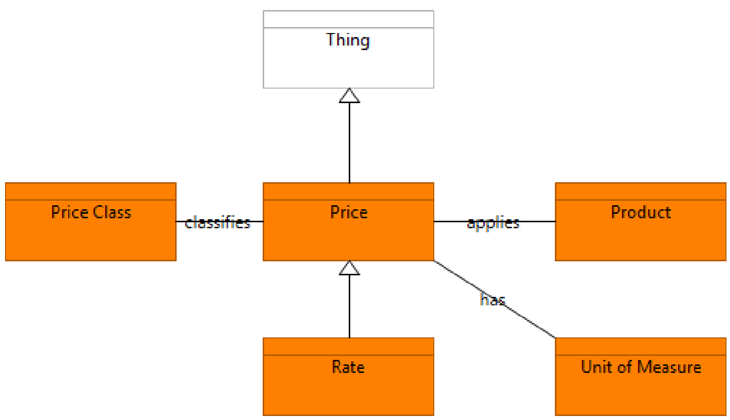
Table 2.11 SLR study Contract CODP

Ontology Pattern Type	***Content Ontology Design Pattern***
Name	Contract
General description	The Contract CODP allows the conceptualization of an agreement between parties playing roles.
Examples	• A legal binding contract for the sales of a house between two persons playing roles of buyer and seller; • A Service Legal Agreement for procuring an infrastructure cloud service to a user from a cloud provider; • The set of terms and conditions associated with a bank-checking service.
Simplified UML diagram (Archimate)	
Definitions of the agnostic concepts	• Contract: A tacit agreement between parties playing roles; • Contract Class: A classification scheme for contracts; • Role: See the Role CODP; • Party: See the Party CODP.

2.11.4 The Price agnostic CODP

The Price CODP optionally relates to products and allows the commercial operations to generate revenues.

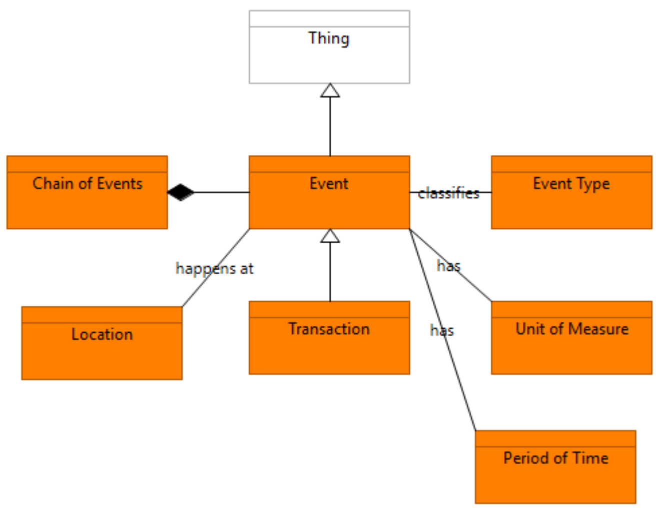
Table 2.12 SLR study Price CODP

Ontology Pattern Type	***Content Ontology Design Pattern***
Name	Price
General description	The Price CODP allows the conceptualization of the notion of rates, rate packages, fees, penalties, pricing curve (time varying cost structure) applicable to the consumption of products.
Examples	• A rack rate applicable for selling room nights in a hotel; • A driver's licence fee for the right to drive a motor vehicle as a service dispensed by a government agency.
Simplified UML diagram (Archimate)	
Definitions of the agnostic concepts	• Price: A financial quantity assigned to the procurement of products; • Price Class: A classification scheme for Price; • Product: See the Role CODP.

2.11.5 The Event agnostic CODP

The Event CODP relates to occurrences in space and time that affects the state of things.

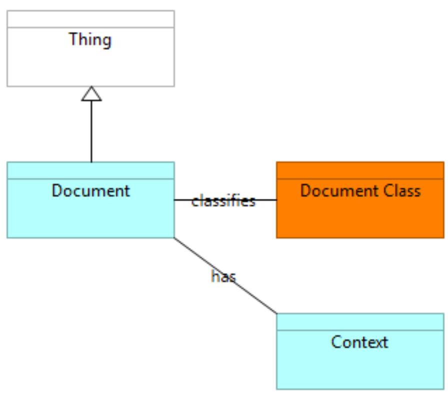
Table 2.13 SLR study Event CODP

Ontology Pattern Type	***Content Ontology Design Pattern***
Name	Event
General description	The Event CODP allows the conceptualization of the notion of a spatiotemporal occurrence that may affect a thing by changing its state.
Examples	• The start of a registration process for a student in a university; • A financial transaction reducing a cash accounting account after the disbursement of a pay cheque.
Simplified UML diagram (Archimate)	
Definitions of the agnostic concepts	• Event: An occurrence in time and space that may affect the state of a thing; • Event Class: A classification scheme for Event; • Chain of events: A grouping of events that is an event in itself; • Transaction: An event that has a quantity where an exchange between more than one thing occurred; • Unit of Measure: A standard for establishing the quantity of a thing, e.g. Currency, weight, height, etc.; • Location: See the Location CODP.

2.11.6 The Document agnostic CODP

The Document CODP is a media containing symbolic facts that a person may bring context and acquire as knowledge and know-how.

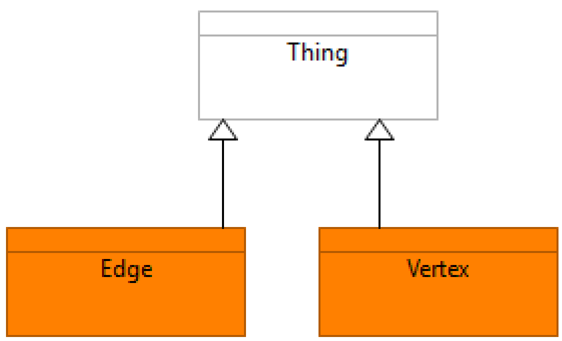
Table 2.14 SLR study Document CODP

Ontology Pattern Type	***Content Ontology Design Pattern***
Name	Document
General description	The Document CODP allows the conceptualization of physical or electronic representation of a body of concepts in a context;
Examples	• The Open Group Architecture Framework book purchased on the Open Group web site; • This SLR will be published as a journal article.
Simplified UML diagram (Archimate)	
Definitions of the agnostic concepts	• Document: A physical or electronic written account of concepts represented through symbols in accordance to a language; • Document Class: A classification scheme for documents; • Context: see the Context CODP.

2.11.7 The Network agnostic CODP

The Network CODP is the implementation of the Petri-network concept for conceptualization.

Table 2.15 SLR study Network CODP

Ontology Pattern Type	***Content Ontology Design Pattern***
Name	Network
General description	The Network CODP allows the conceptualization of a Petri-like structure composed of two nodes and a segment linking the nodes for the purpose of transport of: energy, cargo, people, voice, data, etc. A grouping of networks is also a network.
Examples	• A non-stop flight links Montreal, Canada to Chicago USA; • A telecommunication channel links switching node A to switching node B.
Simplified UML diagram (Archimate)	
Definitions of the agnostic concepts	• Network: A Petri-like structure composed of two nodes and a segment linking the nodes for the purpose of transport of: energy, cargo, people, voice, data, etc.; • Network Class: A classification scheme for Network; • Network Grouping: A group of networks that is also a network.

2.11.8 The Account agnostic CODP

The Account CODP is the only agnostic concept that possesses a dual nature, the Product Account, a mechanism to allow access to a product, and an Accounting Account that is used in financial recording and reporting.

Table 2.16 SLR study Account CODP

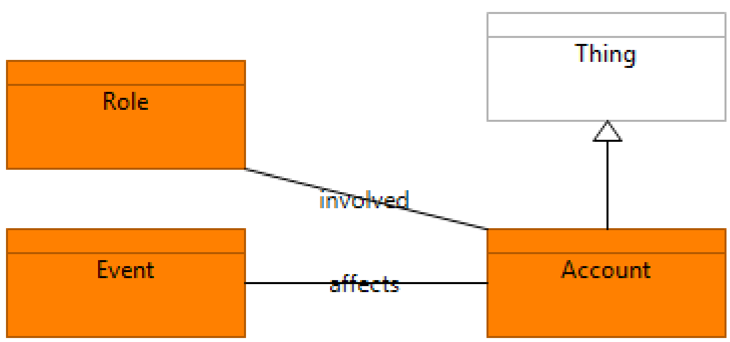
Ontology Pattern Type	***Content Ontology Design Pattern***
Name	Account
General description	The Account CODP allows the conceptualization of a thing used for recording transactions for the purpose of procuring products or tallying quantities for financial statements.
Examples	• A checking account allows the customer to write cheques without fees when the balance is more than $1000 for the whole month; • The Building – Asset account has been adjusted in the Consolidated Grand Ledger by a post-mortem transaction.
Simplified UML diagram (Archimate)	

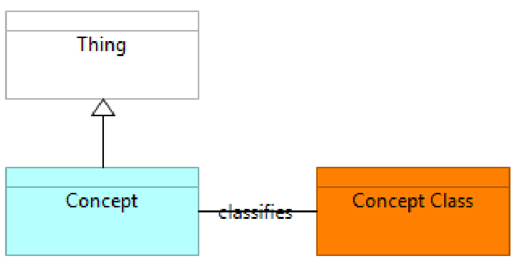
Table 2.16 SLR study Account CODP (continued)

Ontology Pattern Type	***Content Ontology Design Pattern***
Definitions of the agnostic concepts	• Account: A thing used for recording transactions for the purpose of procuring products or tallying quantities for financial statements; • Account Class: A classification scheme for Account; • Network Grouping: A group of accounts that is also an account; • Product Account: A mechanism that allows a customer access to a product under the terms and conditions of a contract; • Accounting Account: A recording structure to tally transaction in accordance to a financial system; • Contract: See the Contract CODP; • Role: See the Role CODP; • Event: See the Event CODP.

2.11.9 The Concept agnostic CODP

The concept CODP would allow the conceptualization of ontological elements and serves as the equivalent of metadata in semi-formal ontologies.

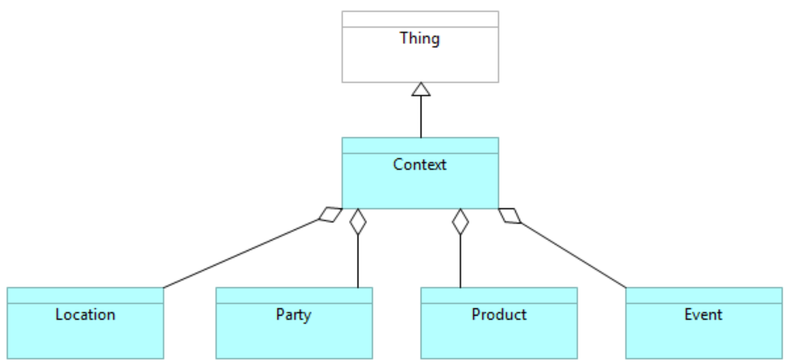
Table 2.17 SLR study Concept CODP

Ontology Pattern Type	***Content Ontology Design Pattern***
Name	Concept
General description	The Concept CODP allows the conceptualization of a man-made imaginary construct that corresponds to real life imaginary or physical things.
Examples	• The CODPs contained in this SLR are agnostic concepts; • The Context CODP is an imaginary concept.
Simplified UML diagram (Archimate)	
Definitions of the agnostic concepts	• Concept: A man-made imaginary things that correspond to real life imaginary or physical things; • Concept Class: A classification scheme for Concept.

2.11.10 The Context agnostic CODP

The Context CODP is scarcely covered in publications. This pattern may be quite useful for several applications including NLP as described in (Akman & Surav, 1997).

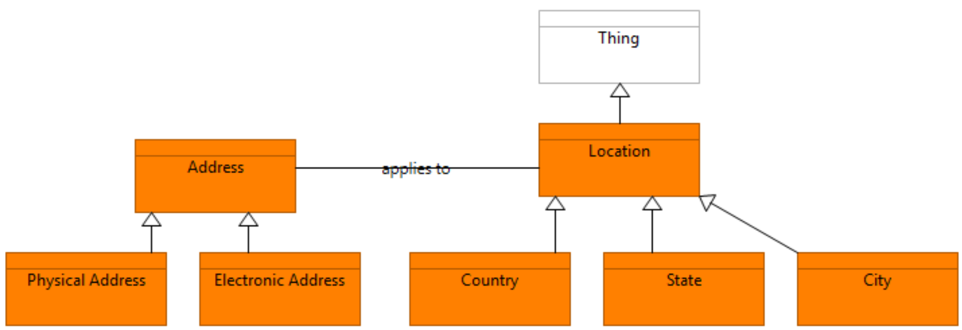
Table 2.18 SLR study Context CODP

Ontology Pattern Type	***Content Ontology Design Pattern***
Name	Context
General description	The Context CODP allows the conceptualization of a set of things such as location, parties, products and events that grouped together may influence the use of vocabularies, chain of future events.
Examples	• In the metaphor-rich American culture, an expression such as «passing the buck» may mean something quite different than when taken literately; • In the context of ACME Corporation, deploying Service-Oriented Architecture (SOA) services just means implementing plain web services.
Simplified UML diagram (Archimate)	
Definitions of the agnostic concepts	• Context: A set of concepts that grouped together may influence the use of vocabularies, chain of future events, etc.; • Context Class: A classification scheme for Context; • Location: see the Location CODP; • Party: see the Party CODP; • Product: see the Product CODP; • Event: see the Event CODP.

2.11.11 The Location agnostic CODP

The Location CODP covers geographical and other forms of coordinated systems.

Table 2.19 SLR study Location CODP

Ontology Pattern Type	***Content Ontology Design Pattern***
Name	Location
General description	The Location CODP allows the conceptualization of a thing related to a coordinate system such as Earth location systems. This includes the notion of area, segment and grid locations. Geography also includes the notion of street addresses and electronic locations such as email and IP addresses.
Examples	• The City of New York is a Location Area included in the State of New York; • The address of this house is 123 Main Streer, Littletown USA and has a centroid determined by a longitude and latitude.
Simplified UML diagram (Archimate)	
Definitions of the agnostic concepts	• Location: A thing related to a coordinate system such as Earth location systems and the concept of address; • Location Class: A classification scheme for Location; • Location Grid: A zero-dimensioned point on a coordinate system; • Location Area: A polygon on a coordinate system; • Location Segment: A curved line of zero width joining two points; • Address: A label affixed on various locations for communication and other purposes; • Physical Address: An address for geographical locations; • Electronic Location: An address used in a media environment such an email address, IP address, etc.

2.11.12 The Role agnostic CODP

The Role CODP constitutes a key concept that allows distinguishing between the nature of things and their behavior.

Table 2.20 SLR study Role CODP

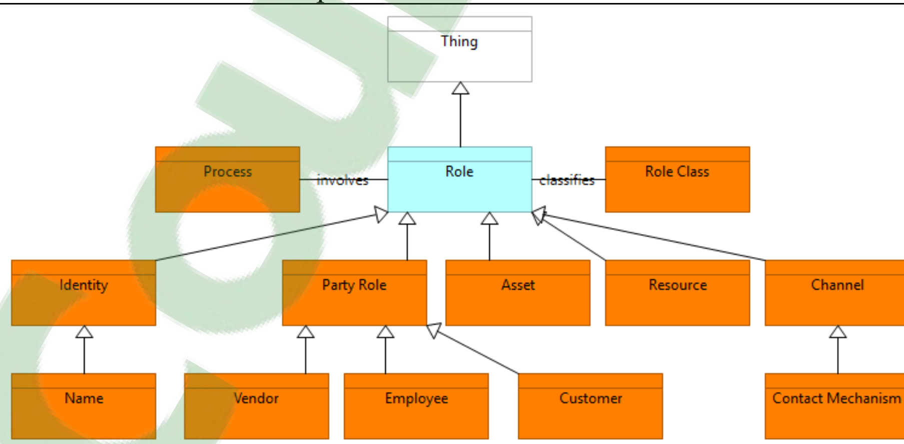
Ontology Pattern Type	***Content Ontology Design Pattern***
Name	Role
General description	The Role CODP allows the conceptualization of a form of involvement in a process or into anything other than a role. A thing playing a role would exhibit a behaviour that may not be related to its nature.
Examples	• A person plays the role of an employee in ACME Corporation; • This horse is an asset for this farmer and is a resource that is involved in farm processes.
Simplified UML diagram (Archimate)	

Table 2.20 SLR study Role CODP (continued)

Ontology Pattern Type	***Content Ontology Design Pattern***
Definitions of the agnostic concepts	• Role: A form of involvement in a Process or into any Thing other than a Role; • Role Class: A classification scheme for Role; • Identity: A Role being played by a Thing to uniquely designate a Thing; • Name: A form of Identity composed of one or more words; • Party Role: A form of Role played by a Party; • Vendor: A Party Role that involved supplying a Product; • Employee: A Party Role that involves being a full-time worker for an organization; • Customer: A Party Role that involves consuming a Product from a vendor; • Asset: A Role being played by a Thing that involves having a value for another Thing; • Resource: A Role being played by a Thing that involves participating in a Process; • Channel: A Role being played by a Thing for allowing access to another Thing; • Contact Mechanism: A Channel used for establishing a community of interest between two or more Things; • Process: see the Process CODP.

2.11.13 The Process agnostic CODP

The Process CODP covers all forms of human or natural activities.

Table 2.21 SLR study Process CODP

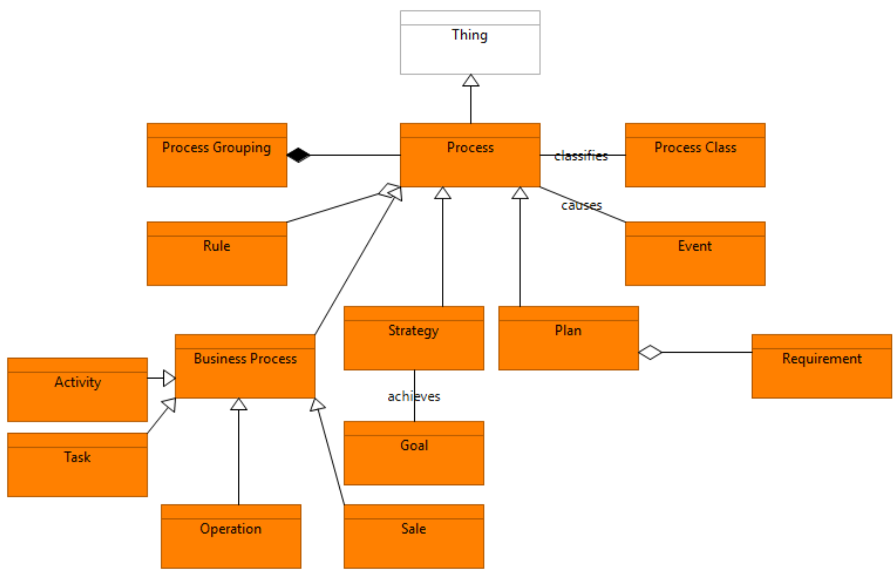
Ontology Pattern Type	***Content Ontology Design Pattern***
Name	Process
General description	The Process CODP allows the conceptualization of a form of a unit of work in which resources are used in the fabrication of goods or in the rendering of services. A process can be performed by humans, by nature or a mix of both.
Examples	• A set of activities in the manufacturing of a consumer electronic product is a Process; • The growth of an animal's fetus in an In Vitro facility is a Process.
Simplified UML diagram (Archimate)	

Table 2.21 SLR study Process CODP (continued)

Ontology Pattern Type	***Content Ontology Design Pattern***
Name	Process
Definitions of the agnostic concepts	• Process: A form of a unit of work in which resources are used in the fabrication of goods or in the rendering of services; • Process Class: A classification scheme for Process; • Process Grouping: A collection of Processes forming another Process; • Rule: A formulated logical constraint that would be used to control the execution of a Process; • Strategy: A Process specifically designed to achieve a goal and not a Product; • Goal: A desired state of a Thing; • Plan: A Process that proposes a sequence of processes and events with a predetermined outcome; • Requirement: An element of the predetermined outcome that is fulfilled by a Plan and relates to the state of a Thing; • Event: See the Event CODP; • Role: see the Role CODP.

The Content Synthesis step concludes the SLR research method by providing the consolidated set of agnostic CODPs. These agnostic CODPs are drawn from the literature using a qualitative form of the SLR approach proposed by (Okoli, 2015).

2.12 Conclusion and future work

The elicitation performed in this paper's SLR approach uncovered 89 light-weighted agnostic CODPs. Although it may be too premature to consider the notion of theoretical saturation as a decision-making technique for research planning, the downward trend may indicate possible opportunities for the use of other qualitative methods such as research action and focus groups. At this point in time, the SLR approach represents an efficient research methodology, especially when used in conjunction with an interview-based approach such the phenomenological research method.

It is important to note that the findings of the elicitation and synthesis of agnostic CODP performed in this SLR includes several CODPs that are also reported in a list of CODPs contained in (Blomqvist, 2010). (Blomqvist, 2010) describes twenty-one CODPs elicited during a research covering best practices in ontology design patterns that are common to several domains. Also, more than 80% of the twenty-one CODPs listed in (Blomqvist, 2010) are present in this SLR, e.g. Party and Person. The remaining CODPs are conceptualized in this SLR by more abstract CODPs, such as in the case of the CODP Analysis Modelling contained in (Blomqvist, 2010) and covered by Process, one of this SLR's key agnostic CODP.

Such close alignment of this SLR with the research findings found in (Blomqvist, 2010) constitutes a demonstration of triangulation as proposed by (Anney, 2014). Such triangulation represents an important means to establish the trustworthiness of the qualitative research method used in this SLR.

Following this SLR, use cases in the domains of Product Lifecycle Management and military logistics are to illustrate the role of the SLR's agnostic CODPs for solving competency questions. The competency questions are drawn from two conference papers that previously covered these domains at a more holistic architectural level (Daniel Fitzpatrick et al., 2013) and (D. Fitzpatrick et al., 2013). The new use cases will cover the competency questions at a more detail ontology design level, using this SLR's elicited CODPs.

Following the final formulation of the resulting conceptualization composed of the set of agnostic CODPs elicited in this research project, the multi-domain ontology is to be formulated as a formal ontology using the OWL language with an approach as proposed in (J. Dietrich & Elgar, 2005) and deployed in the form of an Application Programming Interface (API) as prescribed by (Horridge & Bechhofer, 2011).

Finally, in the wake of this SLR, this project intends to argue for a position in which single domain ontologies would be contraindicated for run-time operation of any cognitive applications. This contraindication would apply for cognitive application capable of knowledge reuse, as described in this SLR at section 2.2.3, for data integration or any other inferential applications. However, single domain ontologies would be used in development time as input to the design of the multi-domain ontology prior to its deployment in run time within a cognitive application.

CHAPTER 3

A USE CASE OF A MULTI-DOMAIN ONTOLOGY FOR COLLABORATIVE LOGISTICS PLANNING IN COALITION FORCE DEPLOYMENT

Daniel Fitzpatrick[1], François Coallier[1], Sylvie Ratté[1]

[1]Department of Software Engineering & Information Technology, École de technologie supérieure,
1100 Notre-Dame West, Montréal, Quebec, Canada H3C 1K3

Paper submitted for publication to the International Journal of Intelligent Defense Support Systems in April 2018

Abstract

The government defense agencies increasingly rely on coalitions to deploy military assets. The defense domain, and the coalition it creates, requires system interoperability. The coalitions need to ensure that their systems interoperate. Interoperability between coalition members involves exchanging data, information (contextualized data), knowledge (actionable information) and know-how (functional knowledge). Coalitions require full interoperability to accomplish their missions at maximum efficiency and efficacy. In this paper, a multi-domain ontology is applied to resolve a competency question about the collaborative logistics planning for force deployment. To plan the deployment and provisioning of military coalition, the logisticians and commanders need to access in a seamless manner, data, information, knowledge and know-how. This paper proposes the use of a formal multi-domain ontology to perform data integration that would allow the seamless exchange of data in a coalition's heterogeneous information technology ecosystems.

This use case utilizes elicited agnostic Content Ontology Design Patterns or CODPs grouped as a specific type of mid-level ontology called a multi-domain ontology (Fitzpatrick, Ratté, et al., 2018a). The concept of multi-domain ontology was proposed previously in (Daniel Fitzpatrick et al., 2012, 2013; D. Fitzpatrick et al., 2013). Agnostic CODPs constitute a conceptualization that covers real world concepts usable across all industries. In this paper, such agnostic concepts are intended to be represented in a formal ontology to provide data integration functionality to perform collaborative logistics planning for force deployment. This paper uses the resulting set of agnostic CODP elicited using a qualitative SLR method. These agnostic CODPs originate from data models, domain models and other semi-formal ontologies usually applied in contemporary non-cognitive information technologies, such as canonical models. Transformed as axioms, these patterns would constitute collectively the multi-domain ontology. This use case primarily serves to demonstrate the transferability (Anney, 2014) , or generalizability, of the agnostic CODPs elicited by the SLR in (Fitzpatrick, Ratté, et al., 2018a).

Keywords: Content ODP, Ontology Design Patterns, Ontology, inference application, multi-domain ontology, military ontology, collaborative logistics planning, trustworthiness, constructivism.

3.1 Introduction

Defence government agencies are affected by semantic heterogeneity in their attempt to implement system interoperability. The scientific community is still attempting to commoditize data integration (Doan et al., 2012) (Olivé, 2017). Semantic heterogeneity constitutes an important challenge for large enterprises and notably for organization such as the US Department of National Defence (Morosoff et al., 2015).

Military coalitions usually include at least one major country and a few local governments to mitigate the risks associated with counterinsurgencies. There is a distinct possibility that coalitions may allow potentially unreliable parties in their midst. (Roberts, Lock, & Verma, 2007).

Coalition members unite for a very specific time with limited goals and do rarely engage in long-term commitments. The International Security Assistance Force (ISAF) in Afghanistan, under the direction of NATO and created in 2001, constitutes a notable exception as a long-lived coalition. Around forty countries joined this partnership for providing military civilian and military capabilities to rebuild Afghanistan. The exchange of reliable information diminishes the chances of discords within the coalition. Access to information is provided according to the members' role and in accordance to agreements (Grant & van den Heuvel, 2010).

Coalitions depend on network-centric warfare capability. A network-centric warfare capability enables battlefield dominance. Ontology based cognitive applications, such as data integration and Natural Language Processing (NLP) represent essential tools for a network-centric warfare capability. These tools allow the coalition to acquire situational awareness of the terrain (Pai, Yang, & Chung, 2017).

The military logistics planning processes are still today primarily manual once the operations have started using office automation software. The logistics processes involved in deploying coalitions' assets and workforce are highly complex. This complexity is explained by the high multitude of variables and the volatility of the situation in the theatre of operations (J. Patel, M. C. Dorneich, D. Mott, A. Bahrami, & C. Giammanco, 2010).

Military planning involves a great variety of business domains and specialties and requires constant and extensive orchestration. The military logisticians face the constant challenges of sharing and broadcasting accurate information and knowledge in a timely fashion to the entire coalition (Jitu Patel et al., 2010). Semantic heterogeneity constitutes a significant hurdle in the exchange of information in the coalition.

This use case attempts to answer a competency question dealt with in a previous use case (D. Fitzpatrick et al., 2013). The competency question was formulated as: «*what is the required logistics load and movement plan for a given coalition force deployment and what are the factors associated with this plan?*».

In the previous use case, an architectural model, the Reference Architecture of an Enterprise Knowledge Infrastructure (RA-EKI), addressed the competency question.
RA-EKI conceptually originates from TOGAF's information integration infrastructure reference model (III-RM). NATO's Architecture Framework (NAF) extensively covers reference architectures, or architectural patterns, that can be applied to business, data, application and technology architectures. The concepts of reference architecture, reference model and architectural patterns constitute synonyms for the purpose of the SLR and used interchangeably. The architectural pattern for the multi-domain ontology is described in detail, as a set of agnostic CODPs, in (Fitzpatrick, Ratté, et al., 2018a) hereafter referred to as the SLR. RA-EKI proposes in figure 3.1 an application reference architecture (Daniel Fitzpatrick et al., 2013), (D. Fitzpatrick et al., 2013) and in figure 2 the architectural pattern for the multi-domain ontology (Fitzpatrick, Ratté, et al., 2018a).

RA-EKI proposes, as illustrated in figure 3.1, «a set of generic applications that transforms unstructured, semi-structured and structured data into information mostly in execution time and information into knowledge in design time. RA-EKI also comprises a unique ontology structure» (D. Fitzpatrick et al., 2013). RA-EKI's ontological structure comprises foundational, mid-level (multi-domain), domains, task and application ontologies.

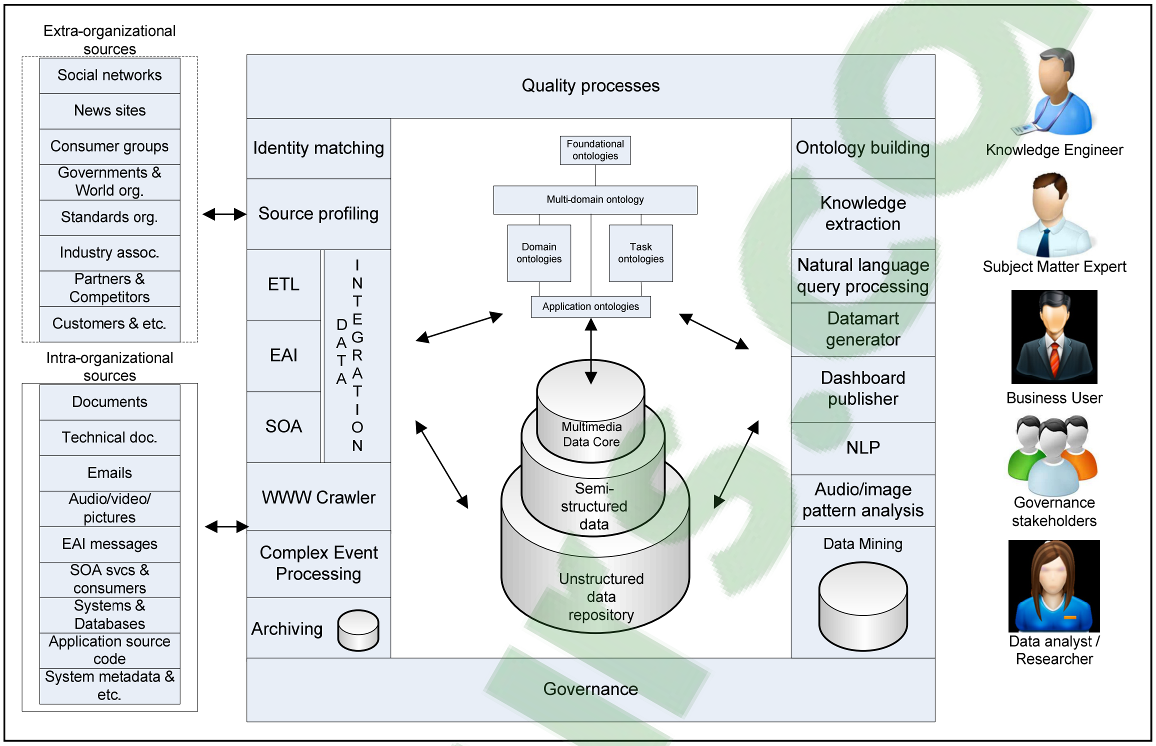


Figure 3.1 Reference Architecture of an Enterprise Knowledge Infrastructure (Daniel Fitzpatrick et al., 2013)

The corner stone of RA-EKI is a multi-domain ontology first introduced in (Daniel Fitzpatrick et al., 2012). This multi-domain ontology proposes agnostic concepts that are applicable across all industries. (Obrst et al., 2012) introduced a new type of ontology, the mid-level ontologies, which are more grounded than foundational ontologies but more abstract than domain ontologies. RA-EKI's multi-domain ontology, a type of mid-level ontology, intends to conceptualize all business concepts that are found across all industries. The multi-domain ontology and its modules aim in providing a cross-domain semantic capability appropriate for a military coalition's requirements for system interoperability.

The SLR revisited the architectural pattern for the multi-domain ontology, found in (D. Fitzpatrick et al., 2013), by eliciting agnostic CODPs using a qualitative systematic literature review approach richly documented in the SLR. This use case applies the agnostic CODPs by attempting to answer the aforementioned competency question. This use case serves as a means to establish the trustworthiness of the SLR by examining the transferability, as prescribed by (Anney, 2014), of the agnostic CODPs in the context of collaborative logistics planning. Anney posits that transferability, a criterion for establishing the trustworthiness of a qualitative research, consists in applying the elicited concepts to a different context that has an actual real life purpose and with different respondents, e.g. reviewers for a scientific journal.

This SLR is primarily based on a methodology described in (Okoli, 2015). Okoli proposed an approach to perform a qualitative systematic literature review. Initially from the life sciences research community, the SLR research method intends to rigorously search and select publications based on a research question. However, Okoli's methodology provided only partial guidance for the analysis and the synthesis of the elicited material. This SLR, (Fitzpatrick, Ratté, et al., 2018a), prescribes more accurately the analysis and synthesis steps inspired from the phenomenology research method proposed in (C. Moustakas, 1994).

The research question is formulated as follows: «w*hat are the conceptualization patterns found in semi-formal ontologies, e.g. data model patterns, software engineering patterns, etc, that can be agnostic to any domain or industry sector in the context of enterprise semantic interoperability and can be used as the basis of agnostic CODPs to resolve semantic heterogeneity in enterprise systems?*» (Fitzpatrick, Ratté, et al., 2018a).

This research question is then translated into a search query, executed in various publication databases and then selected based on a practical screen. The retained publications are then analyzed. The analysis step consists in breaking down the content in the following topics: the primary agnostic concept, the secondary agnostic concepts and the definitions. The synthesis step consolidates the entire elicited material first by primary agnostic concepts. This yields a set of modules under which are associated for each the main agnostic CODP, the subordinate agnostic CODPs, their relations and their definitions. At this point, the present use case attempts to show that the set of agnostic CODPs produced by the SLR's research protocol may apply in the context of collaborative logistics planning for coalition force deployment.

Section 3.2 provides a definition of important concepts for this research. Section 3.3, Related work, describe similar research initiatives. Section 3.4 outlines the multi-domain ontology modules used in this use case. Section 3.5 illustrates and defines the business processes for collaborative logistics planning. Section 3.6 describes the application of the agnostic CODPs to answer the competency question. Section 3.7 concludes the paper with a discussion on the present use case.

3.2 Definition of terms

The following definitions are extracted and summarized from the SLR (Fitzpatrick, Ratté, et al., 2018a)for the present use case. The SLR's provided these definitions to establish a conceptual foundation to this research.

3.2.1 Conceptualization

«*Conceptualization is defined here as a process that implicitly creates semantic structures. Semantic structures establish the meaning of things. Semantic structures are set of concepts, properties and their relationships*» (Giaretta & Guarino, 1995), (Nicola Guarino, 1998).

3.2.2 Representation

«*It is an externalized depiction, or specification, of concepts that can be shared amongst people or machines. Representing concepts involves converting implicit concepts lodged in a person's brain into explicit concepts using a language*» (Nicola Guarino, 1998).

3.2.3 Ontology

Gruber defines an ontology as an «explicit specification of a conceptualization» (Thomas R. Gruber, 1993). «Guarino stresses that ontologies only approximate a conceptualization. He also indicates that the only way to enhance the representation is to develop a richer set of axioms (N. Guarino, 1998).

There are two basic facets of the ontology concept: language dependent, the representation, and language independent, the conceptualization, characteristics (Nicola Guarino, 1998). Based on (Héon, 2010), the four ontology levels are:

- Informal: e.g. natural text;
- Semi-informal: e.g. concept maps;
- Semi-formal : e.g. data models, canonical models, XSDs;
- Formal: a set of logical rules that can be processed by an inference engine for cognitive applications.

3.2.4 Ontology Pattern

Blomqvist describes an ontology pattern as «a set of ontological elements, structures or construction principles that intend to solve a specific engineering problem and that recurs, either exactly replicated or in an adapted form, within some set of ontologies, or is envisioned to recur within some future set of ontologies» (Blomqvist, 2010).

3.2.5 Ontology Design Pattern (ODP)

An Ontology Design Pattern is a «an ontology design pattern is a set of ontological elements, structures or construction principles that solve a clearly defined particular modeling problem» (Blomqvist, 2010).

3.2.6 Content ODP

Based on (Gangemi & Presutti, 2009) (Blomqvist, 2009a), a content ODP, or a CODP, represents a design pattern that describes business concepts found in a domain ontology. This use case provides CODPs that represents business concepts that are relevant across all domains and industries. This project intends to elicit agnostic cross-industry CODP that form the multi-domain ontology.

3.2.7 Enterprise

The Open Group Architecture Framework (Anonymous, 2009), an enterprise is defined as commercial profit driven entity or a no-profit organization or a government agency. An enterprise is also defined a coalition or a partnership. A subdivision of another enterprise such as a subdivision of a company or of a government constitutes an enterprise.

3.2.8 Domain

A domain is defined as set of knowledge and know-how shared by a community, an enterprise or an industry sector (Tennis, 2003).

3.2.9 Agnostic concept

«An agnostic concept is defined here as an abstract concept that possesses a distinct definition amongst other concepts. Thomas Erl defines the term Agnostic in the context of Service Oriented Architecture software component logic as logic that is reusable across all contexts and domains in the enterprise. Furthermore, it is implied here that an agnostic concept is defined in such a way that it cannot be confused with another agnostic concept» (Erl et al., 2017).

3.2.10 Multi-domain ontology

«A mid-level formal ontology composed that comprises a collection of interrelated agnostic CODPs that allows a cross-industry conceptualization. Concepts related to any industry may be represented using the multi-domain ontology» (Daniel Fitzpatrick et al., 2012).

3.3 Related work

This section first surveys literature pertaining to general and logistics collaborative planning. This review mainly studies the military-related business processes, the organizational challenges of a coalition and the critical requirement for system interoperability in military coalition. Secondly, this section investigates ontology applications related to topics such as situation awareness, truck transportation navigation, cargo loading, battlefield dynamics, etc.

Kuster in (Egon Kuster, 2007) considers interoperability crucial for a coalition. Differences caused by semantic heterogeneity constitute important challenges to maintain interoperability. (J. Patel, M. Dorneich, D. Mott, A. Bahrami, & C. Giammanco, 2010) and (Dorneich, Mott, Bahrami, Patel, & Giammanco, 2011) prescribes the extension of the planning processes to encompass all military functions (logistics, operations, intelligence, etc). This approach requires the coalition members' systems to interoperate. Such interoperability supports critical knowledge extraction, essential to the success of the coalitions' missions. Interoperability is also critical for knowledge reusability in that previous plans can be used to accelerate the production of new plans to fulfill new operational requirements.

(J. Patel et al., 2010) and (J. González, de Castro, & Güemes, 2011) prescribe a Service-Oriented Architecture (SOA) approach for collaborative military planning activities. These authors are also proposing a business process layer that involves using Business Process Execution Language (BPEL) scripts. This SOA approach entails the invocation of application assets through data integration. As prescribed by T. Erl's agnostic design and reusability principle, such an approach would be highly dependent on agnostic CODPs, which the present use case proposes.

Reference models such as ICODES (Pohl & Morosoff, 2011), ONISTT (Ford, Martin, Elenius, & Johnson, 2011) and those proposed by Chmielewski (M. Chmielewski, 2009), Gonzalez et al (J. González et al., 2011) and Kuster (E. Kuster, 2007) outline notably NLP, data integration and knowledge extraction applications. Certain projects, such as in (M. Chmielewski, 2009), (Ford et al., 2011), (Pohl & Morosoff, 2011), (Glöckner & Ludwig, 2017), (Hofman & Rajagopal, 2015), (Katsumi & Fox, 2018), (Fokoue, Srivatsa, Rohatgi, Wrobel, & Yesberg, 2009) and (Pai et al., 2017) comprises low abstract specific domain ontologies. These ontologies apply to specific, focused domains such as situation awareness, truck transportation navigation, cargo loading, battlefield dynamics, etc. Another example, the Unified Battle Space Ontology (UBOM) covers a large set of very specific non-agnostic concepts related to the military operations domain, assets and battlefield decision-making. The SOA paradigm, prevalent in the aforementioned projects, is significantly prescribed due to more demanding performance requirements in terms of latency.

The analyzed reference models cover a wide array of ontology patterns. In all cases, the ontology patterns were non-agnostic and do not support reusability, an essential attribute for data integration. The emerging notion of a coalition comprises an extensive set of concepts that are not related to the traditional military doctrines. The richness of related domains such as intermodal logistics, supply chain provisioning and others have extended the military doctrines in a significant manner (D. Fitzpatrick et al., 2013). The SLR and the present use case proposed set of agnostic CODPs intend to ultimately solve the semantic heterogeneity problem and provide coalitions the required support for their systems' interoperability.

3.4 Multi-domain ontology modules

This section introduces the revised modules and definitions that compose the multi-domain ontology. These modules and their associated agnostic CODPs are used in section 3.6 for the intended resolution of the selected competency question. Table 3.1 provides modules' name and description, which are drawn from (Fitzpatrick, Ratté, et al., 2018a).

Table 3.1 Description of the revised agnostic multi-domain modules
(Fitzpatrick, Ratté, et al., 2018a)

Module name	**Module description**
Party	«The Party CODP allows the conceptualization of the nature of a person and an organization».
Product	«A good or service resulting from a process. The UN PCS and NAPCS classification schemes can notably be used as taxonomies for products. The concept of a bill of material allows to package products».
Contract	«The Contract CODP allows the conceptualization of an agreement between parties playing roles».
Price	«The Price CODP allows the conceptualization of the notion of rates, rate packages, fees, penalties, pricing curve (time varying cost structure) applicable to the consumption of products».
Event	«The Event CODP allows the conceptualization of the notion of a spatiotemporal occurrence that may affect a thing by changing its state».
Document	«The Document CODP allows the conceptualization of physical or electronic representation of a body of concepts in a context».
Network	«The Network CODP allows the conceptualization of a Petri-like structure composed of two nodes and a segment linking the nodes for the purpose of transport of: energy, cargo, people, voice, data, etc. A grouping of networks is also a network».
Account	«The Account CODP allows the conceptualization of a thing used for recording transactions for the purpose of procuring products or tallying quantities for financial statements».
Concept	«The Concept CODP allows the conceptualization of a man-made imaginary construct that corresponds to real life imaginary or physical things».
Context	«The Context CODP allows the conceptualization of a set of things such as location, parties, products and events that grouped together may influence the use of vocabularies, chain of future events».

Table 3.1 Description of the revised agnostic multi-domain modules (Fitzpatrick, Ratté, et al., 2018a) (continued)

Module name	**Module description**
Location	«The Location CODP allows the conceptualization of a thing related to a coordinate system such as Earth location systems. This includes the notion of area, segment and grid locations. Geography also includes the notion of street addresses and electronic locations such as email and IP addresses».
Role	«The Role CODP allows the conceptualization of a form of involvement in a process or into anything other than a role. A thing playing a role would exhibit a behaviour that may not be related to its nature».
Process	«The Process CODP allows the conceptualization of a form of a unit of work in which resources are used in the fabrication of goods or in the rendering of services. A process can be performed by humans, by nature or a mix of both».

Each of the described modules comprises primary and secondary agnostic CODPs used in the resolution of the competency question in section 6.

3.5 Business process definition for collaborative logistics planning

The business process definition provides the backdrop for the resolution of the competency question. Inspired from (D. Fitzpatrick et al., 2013) and related projects indicated in section 3.3, these business processes set the requirements for interoperability, thus for data integration. This is a simplification for the purpose of the use case since the actual business processes are far more numerous and complex.

Figure 3.4 illustrates the business processes in sequence, albeit some of the processes may be executed concurrently. The Archimate notation is used to represent the business processes. Table 3.2 provides the definitions for these business processes.

The following information elements constitute input mainly from J3 Operations (we assume here a joint headquarters for a significant force coalition) (D. Fitzpatrick et al., 2013), (Antkiewicz et al., 2012), (Mariusz Chmielewski, Gałka, Jarema, Krasowski, & Kosiński, 2009):

- Concept of Operations, fundamental document describing the core concepts of the mission;
- Asset and commodities inventory and requirements;
- Unit composition and human resource requirements;
- Coalition composition, including civilian organizations;
- Threat analysis issued by J2 Intelligence;
- Operations plan.

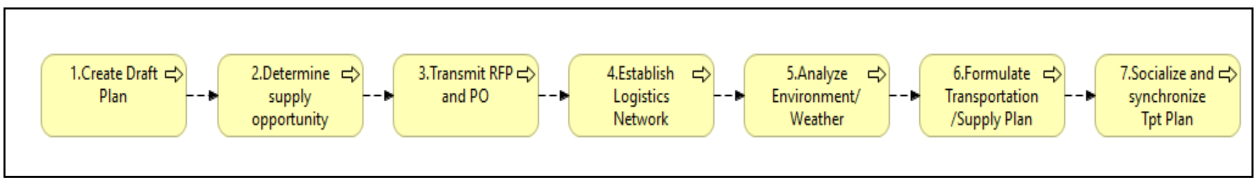


Figure 3.2 Business processes for collaborative logistics planning

Table 3.2 Business process descriptions

Business process name	**Business process description**
1.Create Draft Plan	Based on existing plans, a cognitive planning application, the application, would infer a draft plan to be reviewed by the J4 Logistics branch staff (J. Patel et al., 2010), (Dorneich et al., 2011).
2.Determine supply opportunity	The application would search for the coalition members' supply requirements and would infer a consolidated purchase strategy to minimize costs (Dorneich et al., 2011).
3.Transmit RFP and PO	The application transmits the Request for Proposals, selects the vendors and issues the purchase orders (Glöckner & Ludwig, 2017).
4.Establish Logistics Network	The application identifies the distribution centres, the transportation hubs, modes of transportation, logistics services vendors, network segments, etc (Glöckner & Ludwig, 2017), (Hofman & Rajagopal, 2015), (Katsumi & Fox, 2018).

Table 3.2 Business process descriptions (continued)

Business process name	Business process description
5.Analyze Environment/Weather	The application considers the threat analysis, the weather forecasts and others for establishing various environmental conditions that may affect the deployment of assets and workforce (Katsumi & Fox, 2018), (Antkiewicz et al., 2012), (Smart et al., 2008).
6.Formulate Transportation/Supply Plan	The application consolidates all the information, knowledge and know-how received and generated and produces transportation and supply plans for the deployment (J. Patel et al., 2010), (Dorneich et al., 2011), (Glöckner & Ludwig, 2017), (Hofman & Rajagopal, 2015).
7.Socialize and synchronize Tpt Plan	The application transmits the proposed transportation and supply plan with the coalition members and updates all individual plans upon approval. It also keeps up-to-date the plan when revisions are applied in reaction to events (J. Patel et al., 2010), (Dorneich et al., 2011), (Smart et al., 2008).

In the next section, competency question resolution associates each business process to the agnostic CODPs that used for the resolution of the competency question. It is important to note that only the selected agnostic CODPs used for the competency question resolution is shown. Either the ontology axioms or the assertions are included in the scope for the present use case.

3.6 Competency question resolution

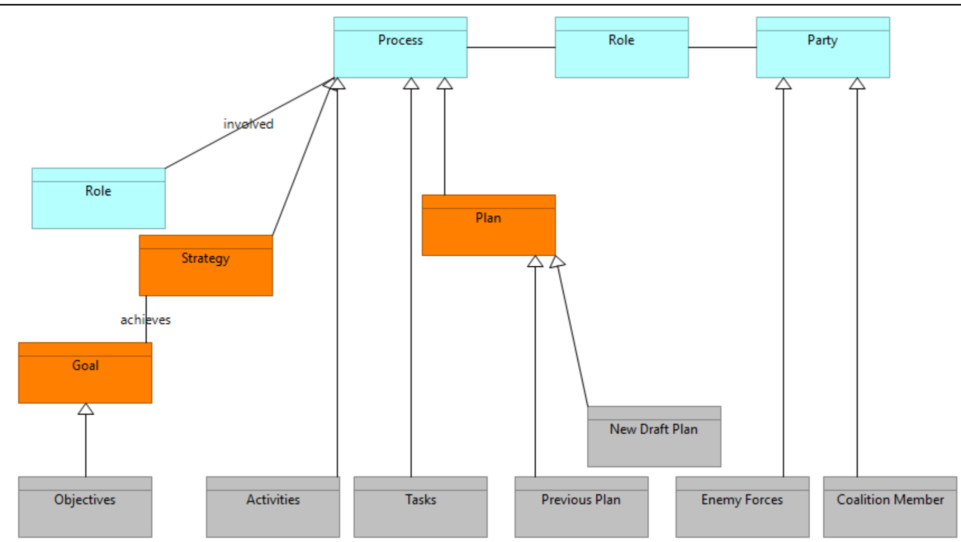
Following the definition of the business processes from the previous section, the agnostic CODPs required for each of the business processes are outlined. The agnostic CODPs (coloured shaded) and the domain specific concepts (grey shaded) are represented in diagrams using the Archimate notation (Lankhorst et al., 2009). This notation standard meets the requirements to model CODPs at its appropriate level. Each business process involved in the resolution attempt of the competency question is described in a format inspired from a CODP template proposed in (Gangemi et al., 2007). The competency question, first enunciated in the present use cases in section 3.1 Introduction, is formulated as in the following:

« What is the required logistics load and movement plan required for a given coalition force deployment and what are the factors associated with this plan?».

3.6.1 Create Draft Plan step

The first step is to create a draft plan from previous plans and from existing knowledge and assertions.

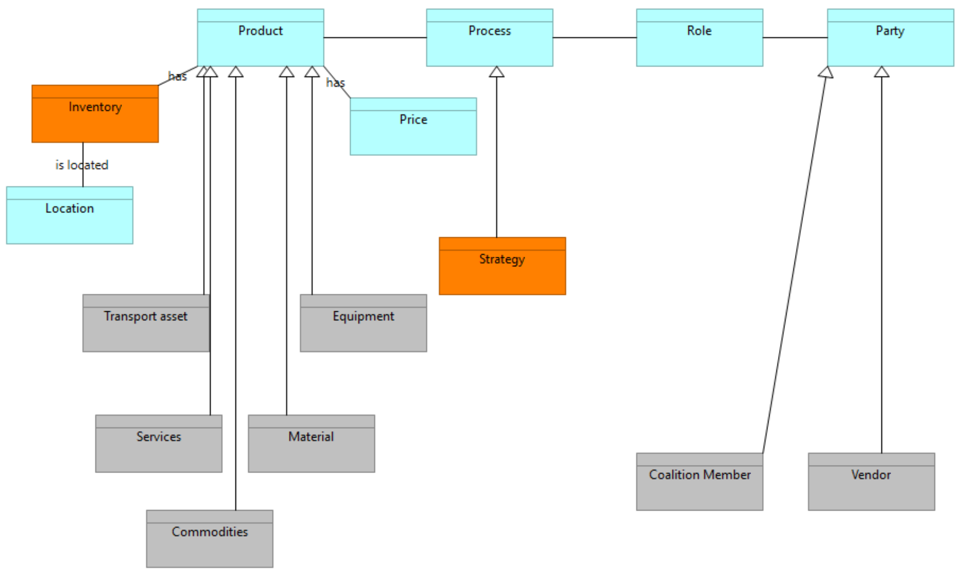
Table 3.3 Create Draft Plan

	Use of agnostic CODPs for business processes
Name	1.Create Draft Plan
Simplified UML diagram (Archimate)	

3.6.2 Determine supply opportunity

The second step consists in seeking the coalition members' supply requirements and determines any opportunity to consolidate purchases to minimize costs.

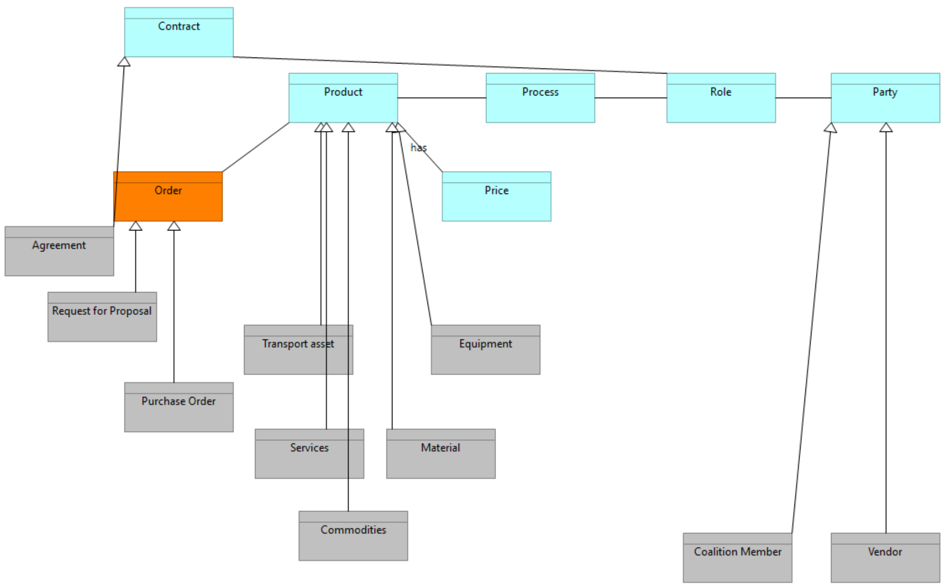
Table 3.4 Determine supply opportunity

	Use of agnostic CODPs for business processes
Name	2.Determine supply opportunity
Simplified UML diagram (Archimate)	

3.6.3 Transmit RFP and PO

The third step actions the strategy established in the previous step and issues the Request for Proposals and the Purchase Orders.

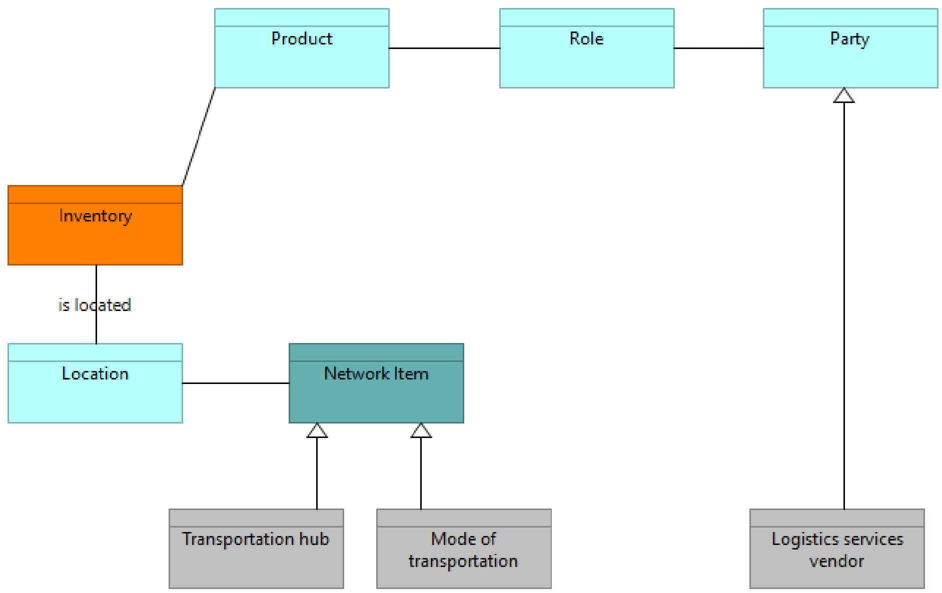
Table 3.5 Transmit RFP and PO

	Use of agnostic CODPs for business processes
Name	3.Transmit RFP and PO
Simplified UML diagram (Archimate)	

3.6.4 Establish Logistics Network

The fourth step determines the supply and transportation network for the provisioning of goods and services.

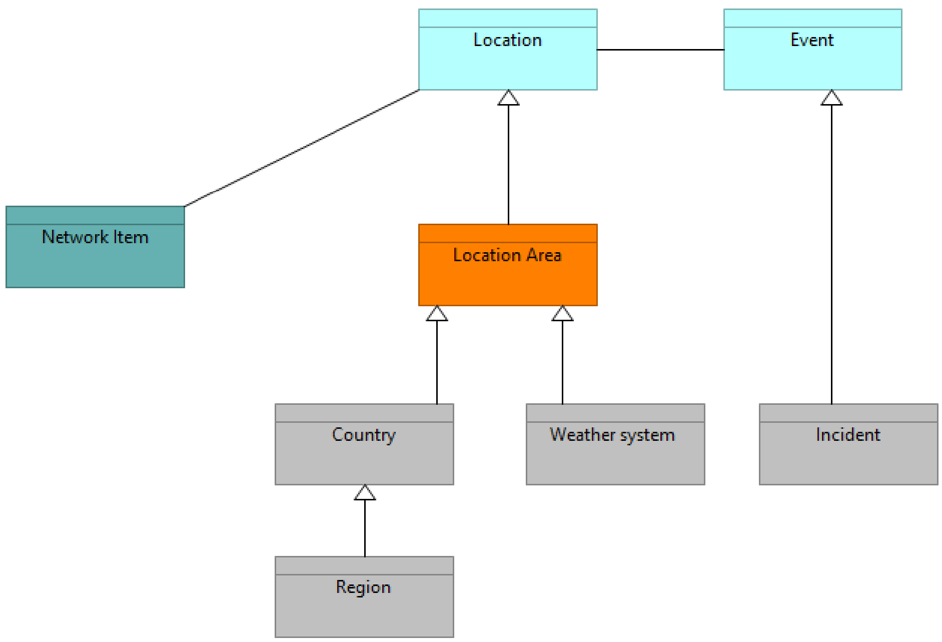
Table 3.6 Establish Logistics Network

	Use of agnostic CODPs for business processes
Name	4.Establish Logistics Network
Simplified UML diagram (Archimate)	

3.6.5 Analyze Environment/Weather

The fifth step involves the study of any weather, incidents, geological anomalies and others to determine any adverse effects on the transportation and supply network.

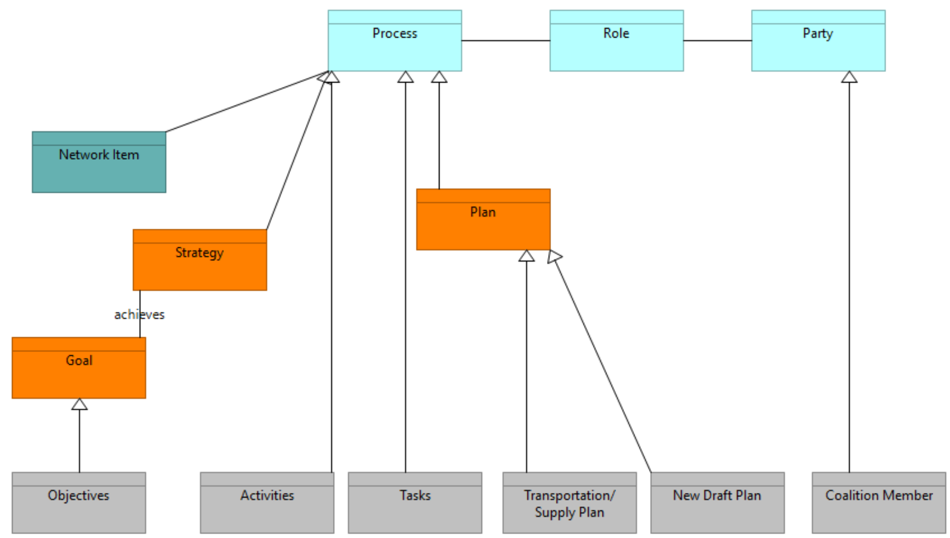
Table 3.7 Analyze Environment/Weather

	Use of agnostic CODPs for business processes
Name	5. Analyze Environment/Weather
Simplified UML diagram (Archimate)	

3.6.6 Formulate Transportation/Supply Plan

The sixth step produces the refined transportation and supply plan. It is generated from the draft plan produced in step 1 and considers all other factors determined from steps 2 through 5.

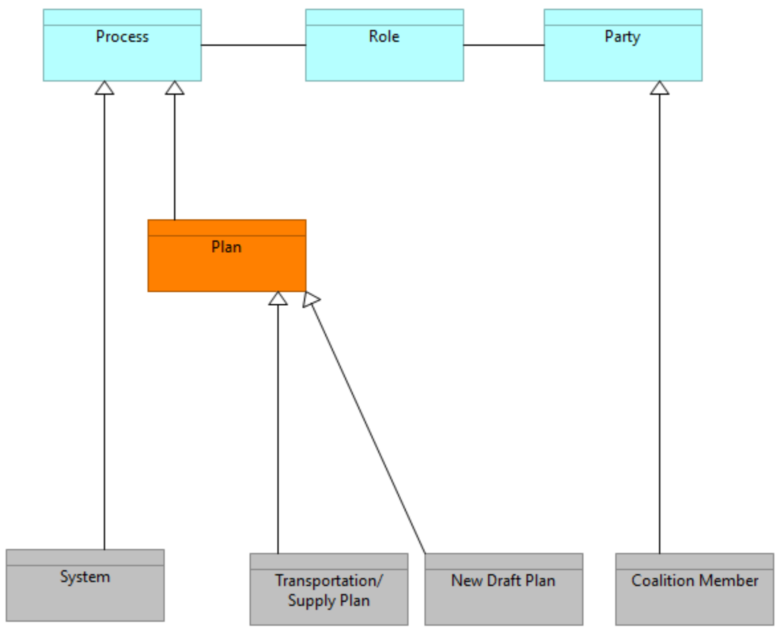
Table 3.8 Formulate Transportation/Supply Plan

	Use of agnostic CODPs for business processes
Name	6.Formulate Transportation/Supply Plan
Simplified UML diagram (Archimate)	

3.6.7 Socialize and synchronize Transportation Plan

The seventh step allows the transportation and supply plan to be socialized with all coalition members. It also involves the various application systems in the greatest coalition network to be updated with information on a need to know basis.

Table 3.9 Socialize and synchronize Transportation Plan

	Use of agnostic CODPs for business processes
Name	7.Socialize and synchronize Transportation Plan
Simplified UML diagram (Archimate)	

3.7 Conclusion

The competency question resolution illustrates the use of agnostic CODPs for each step and represented the mappings between the domain specific concept and the agnostic CODPs. This allows determining to what extent the multi-domain ontology, and its included set of patterns can support the various and numerous domain ontologies involved in the collaborative logistics planning processes.

The agnostic CODPs allows generalizing several of the domain specific concepts into a smaller set of agnostic CODPs. For example, in section 3.6.6, pertaining to the formulation of a transportation and supply plan, the agnostic concepts process and plan can subsumed several lower abstract concepts that are domain-specific ontologies surveyed in section 3.3. In section 3.6.5, the processes for the analysis of the environment and weather uses almost exclusively the location agnostic CODP, which can subsumed a significant number of geographical and weather related concepts, such as country, city, river, ocean and meteorological system.

This use case intended to demonstrate the transferability, the equivalent of generalizability for qualitative research (Anney, 2014) in respect to the set of elicited agnostic CODPs from the SLR. Upon completion of the project, further research is planned work on the multi-domain ontology for possibly reaching a higher level of theoretical saturation and eventual design, development and test experiments.

CHAPTER 4

A USE CASE OF A MULTI-DOMAIN ONTOLOGY FOR COLLABORATIVE PRODUCT DESIGN

Daniel Fitzpatrick[1], François Coallier[1], Sylvie Ratté[1]

[1]Department of Software Engineering & Information Technology, École de technologie supérieure,
1100 Notre-Dame West, Montréal, Quebec, Canada H3C 1K3

Paper submitted for publication to the International Journal of Product Lifecycle Management in April 2018

Abstract

New approaches to design manufactured products are proposed to allow product manufacturers to be more competitive: Set-Based Design (SBD) (Kerga et al., 2016), a new product development process proposed in (Belay et al., 2014) and the modular approach (Buergin et al., 2018). The SBD approach, for example, can contribute to reducing in average by 25% the project duration and by 40% the total project costs as demonstrated in laboratory simulations (Kerga et al., 2016). These new product design approaches require that the Product Lifecycle Management (PLM) application systems interoperate (Daniel Fitzpatrick et al., 2013). Semantic heterogeneity adversely affects system interoperability thus hindering efforts to execute the new product design methodologies.

To address the semantic heterogeneity problem, we propose a use case using the formal multi-domain ontology to perform data integration, thus allowing the required ontology based system interoperability. This paper uses a set of agnostic Content Ontology Design Patterns or CODPs grouped as a specific type of mid-level ontology called a multi-domain ontology (Fitzpatrick, Ratté, et al., 2018a). We believe that the use case described in this paper demonstrate the compliance to the transferability criterion to establish the trustworthiness (Anney, 2014) of qualitative Systematic Literature Review (SLR). Furthermore, this use case aims for the same research objective as (Fitzpatrick, Coallier, et

al., 2018), which pertains to collaborative logistics planning for coalition force deployment. The concept of multi-domain ontology was previously discussed in (Daniel Fitzpatrick et al., 2012, 2013; D. Fitzpatrick et al., 2013). The agnostic CODPs constitutes a conceptualization that covers real world concepts usable across all industries. These agnostic CODPs were elicited from data models and other semi-formal ontologies. Once transformed as axioms, these patterns would form together the multi-domain ontology.

Keywords: Content ODP, RA-EKI, Ontology Design Patterns, Ontology, inference application, multi-domain ontology, PLM, Product Lifecycle Management,, collaborative product design, SBE, PD, qualitative research, trustworthiness, constructivism.

4.1 Introduction

New approaches to design manufactured products are proposed to allow product manufacturers to be more competitive: Set-Based Design (SBD) (Kerga et al., 2016), a new product development process proposed in (Belay et al., 2014) and the modular approach (Buergin et al., 2018). The SBD approach, for example, can contribute to reducing in average by 25% the project duration and by 40% the total project costs as demonstrated in laboratory simulations (Kerga et al., 2016).

These new product design approaches require that the Product Lifecycle Management (PLM) systems interoperate (Daniel Fitzpatrick et al., 2013). Semantic heterogeneity adversely affects system interoperability thus hindering efforts to execute the new product design methodologies. To resolve the semantic heterogeneity problem, a SLR contained in (Fitzpatrick, Ratté, et al., 2018a) propose a multi-domain ontology composed of a set of agnostic CODPs.

This use case attempts to answer a competency question, using the agnostic CODPs, as previously executed in a use cases (Daniel Fitzpatrick et al., 2013) related to the application of the Reference Architecture of an Enterprise Knowledge Infrastructure (RA-EKI) for product design. For product design, the competency question is now reformulated as:

«*What are the factors for each phase or business process of product design, which may influence the financial, customer and environmental value of the new product currently under development?*».

Section 4.2 provides a definition of important concepts for this research. Section 4.3 Related work describe similar research initiatives. Section 4.4 outlines the multi-domain ontology modules used in this use case. Section 4.5 illustrates and defines the business processes for collaborative product design. Section 4.6 describes the application of the agnostic CODPs to answer the competency question. Section 4.7 concludes the paper with a discussion on the present use case.

4.2 Definition of terms

The following definitions are extracted from the SLR (Fitzpatrick, Ratté, et al., 2018a) for this use case. The SLR's provided these definitions to establish a conceptual foundation to this research. The original citations are also provided.

4.2.1 Conceptualization

«Conceptualization is defined here as a process that implicitly creates semantic structures. Semantic structures establish the meaning of things. Semantic structures are set of concepts, properties and their relationships» (Giaretta & Guarino, 1995), (Nicola Guarino, 1998).

4.2.2 Representation

«It is an externalized depiction, or specification, of concepts that can be shared amongst people or machines. Representing concepts involves converting implicit concepts lodged in a person's brain into explicit concepts using a language» (Nicola Guarino, 1998).

4.2.3 Ontology

Gruber defines an ontology as an «explicit specification of a conceptualization» (Thomas R. Gruber, 1993). «Guarino stresses that ontologies only approximate a conceptualization. He also indicates that the only way to enhance the representation is to develop a richer set of axioms (N. Guarino, 1998).

Figure 4.1 illustrates the two basic facets of the ontology concept: language dependent, the representation, and language independent the conceptualization (Nicola Guarino, 1998). Figure 4.1 also illustrates the four ontology levels (Héon, 2010):

- Informal: e.g. natural text;
- Semi-informal: e.g. concept maps;
- Semi-formal : e.g. data models, canonical models, XSDs;
- Formal: a set of logical rules that can be processed by an inference engine for cognitive applications.

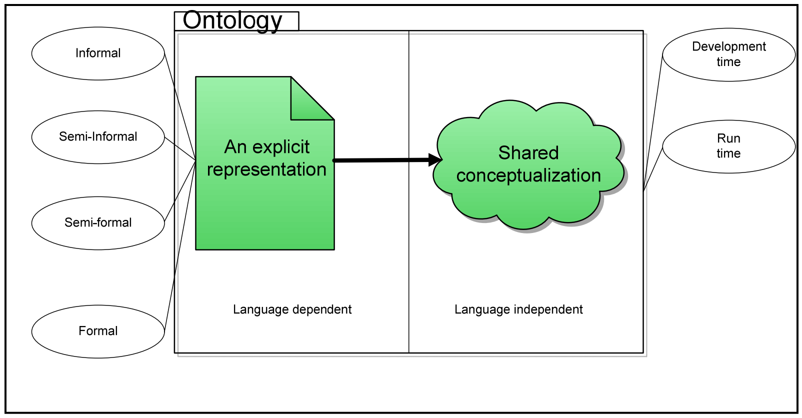


Figure 4.1 Summarized definition of an ontology
(Fitzpatrick, Ratté, et al., 2018a)

4.2.4 Ontology Pattern

Blomqvist describes an ontology pattern as *«a set of ontological elements, structures or construction principles that intend to solve a specific engineering problem and that recurs, either exactly replicated or in an adapted form, within some set of ontologies, or is envisioned to recur within some future set of ontologies»* (Blomqvist, 2010).

4.2.5 Ontology Design Pattern (ODP)

An Ontology Design Pattern is a «an ontology design pattern is a set of ontological elements, structures or construction principles that solve a clearly defined particular modelling problem» (Blomqvist, 2010).

4.2.6 Content ODP

Based on (Gangemi & Presutti, 2009) (Blomqvist, 2009a), a CODP represents a design pattern that describes business concepts found in a domain ontology. This use case provides CODPs that represents business concepts that are relevant across all domains and industries.

4.2.7 Enterprise

The Open Group Architecture Framework (Anonymous, 2009) defines an enterprise as a commercial profit driven entity or a no-profit organization or a government agency. An enterprise is also defined as a partnership or a virtual enterprise, a group of companies joining up to develop a new product. A subdivision of another enterprise such as a subdivision of a company or of a government constitutes an enterprise.

4.2.8 Domain

«A domain is defined as set of knowledge and know-how shared by a community, an enterprise or an industry sector» (Tennis, 2003).

4.2.9 Agnostic concept

An agnostic concept is defined here as an abstract concept that possesses a distinct definition amongst other concepts. Thomas Erl defines the term Agnostic in the context of Service Oriented Architecture software component logic as logic that is reusable across all contexts and domains in the enterprise. Furthermore, it is implied here that an agnostic concept is defined in such a way that it cannot be confused with another agnostic concept (Erl et al., 2017).

4.2.10 Multi-domain ontology

«A mid-level formal ontology composed that comprises a collection of interrelated agnostic CODPs that allows a cross-industry conceptualization. Concepts related to any industry may be represented using the multi-domain ontology» (Daniel Fitzpatrick et al., 2012).

The definitions contained in this section allow a better understanding of the present use case, particularly in the execution of the competency question. In the next section, a literature review is performed first on the emerging product design approach such as Set-Based Design

(SBD) (Kerga et al., 2016), a new product development process proposed in (Belay et al., 2014) and the modular approach for customized product design (Buergin et al., 2018).

4.3 Related work

As stated previously, this use case aims to demonstrate the capacity of the set of agnostic CODPs, forming the multi-domain ontology, elicited in the SLR contained in (Fitzpatrick, Ratté, et al., 2018a), to support knowledge sharing that is critical for collaborative product design. In the first part of this literature survey, we will investigate new product design approaches. This will allow proposing a set of business processes used to represent the main activities involved in product design. This set of business processes, listed and defined in section 4.5, is then used in the execution of the competency question in section 4.6. The set of business processes summarized here doesn't mean to be exhaustive and complete but to allow a sufficient context to demonstrate the adherence to the transferability criterion (Anney, 2014) of the multi-domain ontology's set of agnostic CODPs elicited in the SLR as described by Fitzpatrick et al.

In the second part of this section, a survey elicits ontologies that are specifically designed for product design and product development. Concepts drawn from the surveyed publications are included and represented in the light UML diagrams of section 4.6. The execution of the competency question intends to demonstrate that agnostic CODPs can subsume domain specific (low abstract) concepts in ontologies designed to support product design.

Industry, notably manufacturers, depends increasingly in Collaborative Product Design (CPD) to diminish costs and time-to-market and to increase quality. CPD leverages the optimization of the production and business processes of the enterprises and the virtual enterprise, a group of business units manufacturing together (Abadi, Ben-Azza, & Sekkat, 2017). Under the pressure of a highly competitive market, the manufacturers need to implement CPD to reduce design time. In order to achieve the necessary design time reduction, the virtual enterprises must support knowledge sharing. The virtual enterprises must perform their business processes in an agile, robust and flexible manner. To achieve these latter requirements, an ontology-based data integration function may allow system interoperability amongst the units of the virtual enterprises (Abadi, Ben-Azza, & Sekkat, 2016).

With similar goals, Lean Product Development (LPD) attempts to reduce unnecessary effort to design and market valued and environmentally friendly products. The SBD approach, used concurrently with LPD, accelerates the initial stage of the product development process, mainly design, and reduces the uncertainty with prototyping (Kerga et al., 2016). Systems interoperability constitutes a requirement to the virtual enterprise to outperform the competition (Belay et al., 2014). The SBD, also called Set-Based Concurrent Engineering (SBCE) as described by (Belay et al., 2014), can be defined as an approach that «allows more of the design effort to proceed concurrently and defers details specifications until tradeoffs are more fully understood» (Singer, Doerry, & Buckley, 2009). SBD differs with traditional design processes, also referred to as the «design spiral» from Evans (Evans, 1959). The traditional design approach, also called Point-Based Design (PBD) approach, is inadequate to handle large complex product developments. The PBD approach tends to signal a product design effort as complete on the basis of budget and time limitations, and not on the actual fulfillment of the product design requirements (Singer et al., 2009). By contrast to PBD, SBD engages multiple concurrent design processes.

(Buergin et al., 2018) describe an approach to address the rising requirement for customized products. This approach that consists in compartmentalizing product development in modules also leverages the concept of collaborative product design. This modular approach effectively breaks down holistic product target architecture in relatively independent major components.

(Singer et al., 2009) cites (Womack, Jones, & Roos, 1990) and (Ward, Liker, Cristiano, & Sobek, 1995) in describing a study on Toyota's automobile design approach that designs quality products in a significantly shorter time than other automobile manufacturers. Toyota's design approach, later referred to as SBD consists in four fundamental tenets:

- Broader sets of design requirements are specified to effectively enable multiple track design processes;
- The sets of design requirements are allowed longer treatment to converge to more accurate product specifications;
- The design sets evolve more accurately until a holistic solution emerges that meets the requirements;
- Finally, as the solution emerges, the design gains in detail (Singer et al., 2009).

The SBD approach, compared to PBD, has demonstrated in research and simulations a reduction of between 20% and 25% in average project duration and between 40% and 50% in total project costs (Kerga et al., 2016), (Belay et al., 2014). Kerga et al formulate the two following principle (Kerga et al., 2016) that summarizes the essence of what is SBD:

Principle #1: «*When designing, always work on several alternative solutions at the same time»*;

Principle #2: «*Instead of selecting between alternatives, proceed by elimination»*.

A set of business processes is listed and defined in section 4.5. These business processes, derived from this section's first part survey, are used in the execution of the competency question in section 4.6. Figure 4.2 represents some of the key concepts in the aforecited related work. The relationships in the light UML diagram using Archimate notation standards are read from left to right. (Lankhorst et al., 2009). Table 4.1 describes the concepts represented in figure 4.2 in more detail.

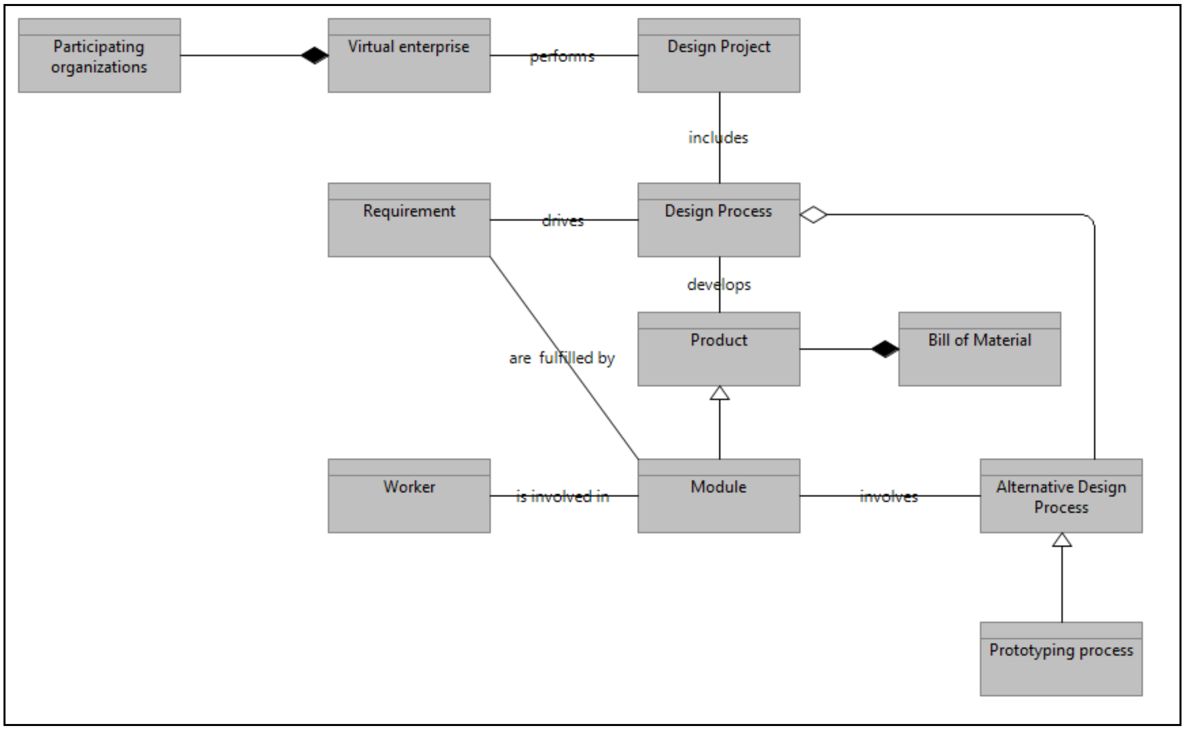


Figure 4.2 Key product design concepts based pertaining to the SBD, CPD and modular approaches

Table 4.1 Description of the product design concepts based on the SBD, CPD and modular approaches

Concept name	**Concept description**
Design project	A concerted planned and managed effort to develop a product.
Design process	A set of sub-processes intended to develop a product.
Product	A manufactured good for which the design project intends to develop.
Module	A distinct major component that can be independently designed and manufactured for the most part.
Worker	An individual that participates in the design of a module.
Alternative design process	A set of activity that can execute concurrently to another for the design of a given module or major component of a product.
Prototyping process	A type of alternative design process that involved building a working replica of the intended module strictly for design purposes.
Bill of Material	A named list of parts that composes the product, including the modules.
Requirement	A description of the intended function constraint.

Table 4.1 Description of the product design concepts based on the SBD, CPD and modular approaches (continued)

Concept name	**Concept description**
Virtual enterprise	A collection of independent organizations, mostly suppliers and manufacturers that collaborate to develop a product.
Participating organization	An organization that participates in some capacity to the development of a product.

At this point, concepts from ontologies designed to support interoperability and knowledge sharing for product development will be elicited from the survey. Formal ontologies executed in inference-capable cognitive applications can contribute to solving the problem of semantic heterogeneity. (Fortineau, 2013) and (Abadi et al., 2016) assert that formal ontologies can perform the following functions: integrate data, execute explicit knowledge for various applications and provide natural language flexible queries. In the context of collaborative product design, ontology-based applications constitute an important enabling technology especially for knowledge sharing, crucial to semantic interoperability necessary to collaborative product design (Abadi et al., 2017).

(Abadi et al., 2017) propose the Collaborative Product Design Ontology, or CPD-Onto, to address the knowledge management and sharing requirements of CPD. CPD-Onto conceptualizes the domain semantics by using generic concepts. CPD-Onto development involved using a semi-formal ontology, i.e. a data model, iteratively to properly support CPD. Figure 4.5 illustrates CPD-Onto main concepts, which are summarily described in table 4.2. The CPD-Onto ontology intends to conceptualize not only collaborative product design, but the manufacturing and supply chain processes as well. The concepts represented in this model originate from the authors' experience. The relationships in the light UML diagram using Archimate notation standards (Lankhorst et al., 2009) are read from left to right and from top to bottom.

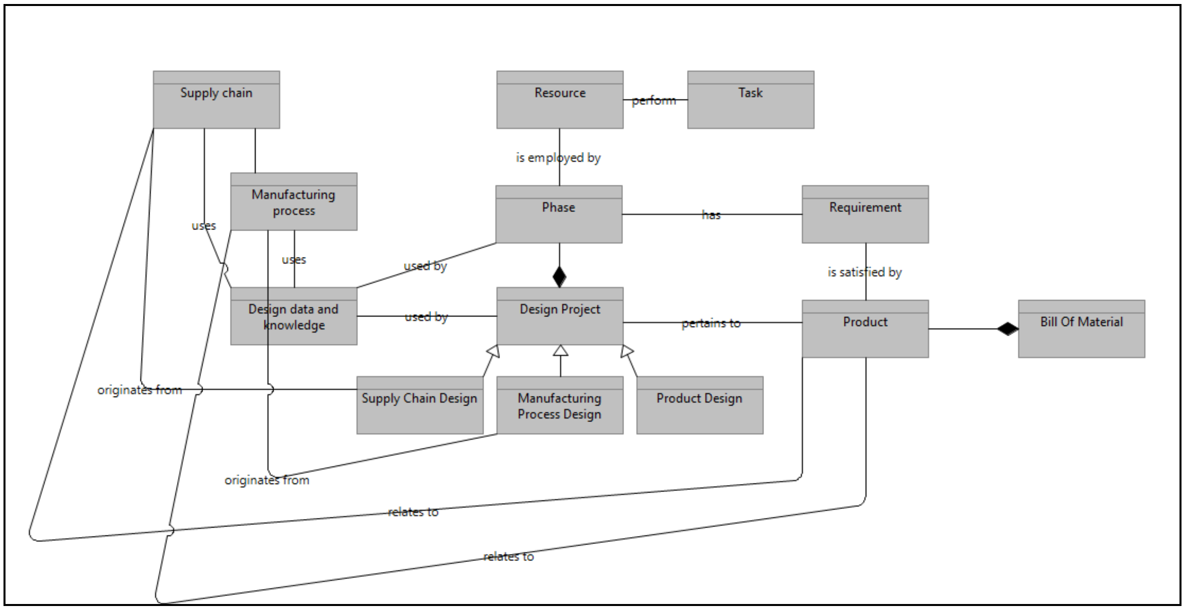


Figure 4.3 The generic conceptual model of the Collaborative Product Design ontology CPD-Onto (Abadi et al., 2017)

Table 4.2 Description of the CPD_Onto main concepts (Abadi et al., 2017)

Concept name	**Concept description**
Design Project	An initiative to develop a product to fulfill requirements and broken down in phases.
Supply Chain Design	A type of design project that specializes in implementing a supply chain, a set of processes and actors, to provision material for manufacturing.
Manufacturing Process Design	A type of design project that specializes in implementing a manufacturing process to produce the desired product.
Product Design	A type of design project that specializes in developing the actual product.
Phase	A division of the design project that represents a distinct stage in the implementation of the supply chain and the manufacturing process, and actual development of the intended product.
Resource	A thing that is involved in a phase that is involved in a task, either money, material or people.
Task	An element of work performed with the use of a resource.
Requirement	A specification of the desired functionality or of a constraint.
Product	A good that is manufactured to satisfy requirements. It can be an assembly or a part.
Bill Of Material	A named list of products, or parts, composing another product called an assembly that represents the final product.

Table 4.2 Description of the CPD_Onto main concepts (Abadi et al., 2017)
(continued)

Concept name	**Concept description**
Supply chain	A set of processes and actors of the virtual enterprise involved in supplying the required resources to manufacture the product.
Manufacturing process	A set of activities involved in producing the desired good to fulfill the requirements.
Design data and knowledge	A set of factual symbols and actionable information to be used by processes involved in the design project.

The authors of the conceptualization represented in figure 4.5 intended their ontology to be generic to the manufacturing industry. The authors also covered a wide set of processes by conceptualizing the supply chain, the manufacturing and the design processes. The product concept is a manufactured good such as equipment.

(Abadi et al., 2016) proposes an ontology to support interoperability within systems in a virtual enterprise. In the context of collaborative product development, a virtual enterprise may encompass several distinct commercial or other types of organization that collaborate for the development of a product. The authors propose an ontology for integration and interoperability purposes. Figure 4.6 represents the proposed ontology in light UML using the Archimate notation standards (Lankhorst et al., 2009). The relationships in the diagram are read from left to right and from top to bottom. Table 4.3 describes the concepts proposed by the authors in (Abadi et al., 2016) intends to cover the entire product lifecycle including the stakeholders.

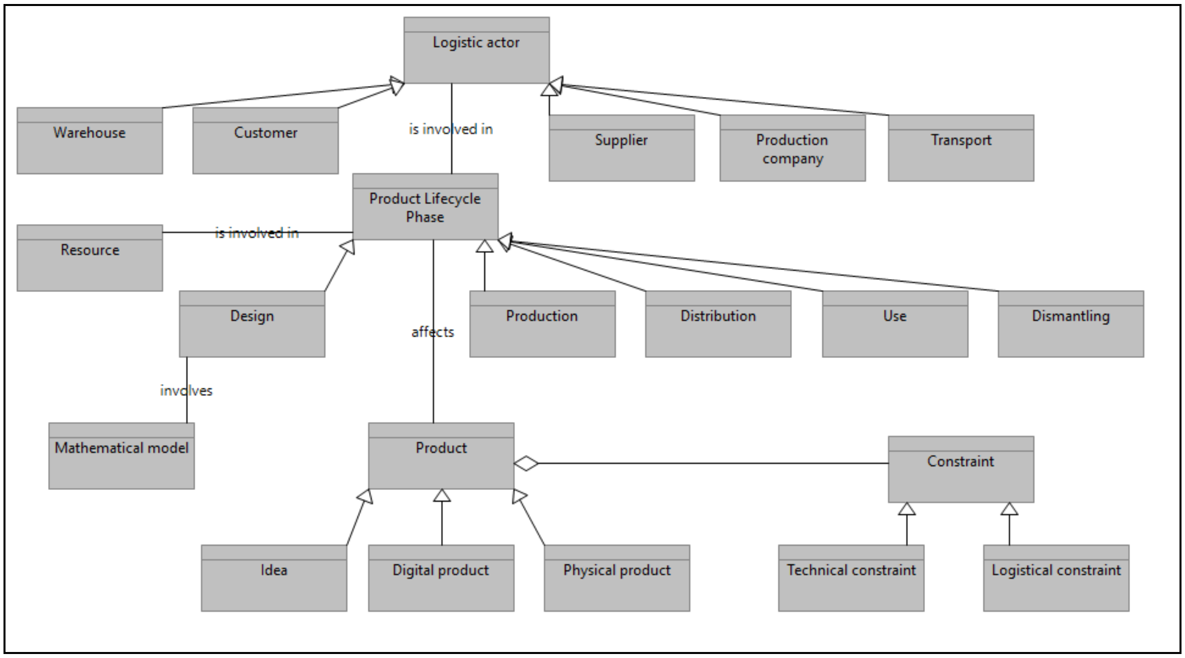


Figure 4.4 The proposed ontological meta-model by (Abadi et al., 2016)

Table 4.3 Description of ontological meta-model

Concept name	**Concept description**
Logistic actor	A stakeholder that is involved in the product lifecycle management. Includes the customer, the warehouse, the supplier, the production company and the transport organization.
Product Lifecycle Phase	A stage of evolution of the product.
Product	An offering in the form of a manufactured tangible good.
Resource	A financial, material, personnel or software concept involved in product lifecycle management.
Mathematical model	An algorithm to optimize design aspects.
Constraint	A logistical or functional limiting factor.

(Daniel Fitzpatrick et al., 2013) posits that various ontology approaches are used in PLM to provide a formal vocabulary to their semantic applications. (Daniel Fitzpatrick et al., 2013) indicates that most models use widely known ontologies such as STEP, CPM, Onto-PDM and TOVE, citing (Khedher, Henry, & Bouras, 2012; Lu et al., 2013; Marchetta, Mayer, & Forradellas, 2011; TERKAJ, PEDRIELLI, & SACCO, 2011) and (Terkaj, Pedrielli, & Sacco, 2012). These aforecited ontologies conceptualize notions that are unrelated to Product

Lifecycle Management (PLM) such as customer, human resource and financial data. Also, (Daniel Fitzpatrick et al., 2013) stress the pervasiveness of concepts in citing (Terzi, Bouras, Dutta, Garetti, & Kiritsis, 2010). (Daniel Fitzpatrick et al., 2013) raise the importance of the dynamic nature of PLM and all the other process-centric paradigms, such as Customer Relationship Management (CRM), Enterprise Resource Planning (ERP) and others.

4.4 Multi-domain ontology modules

This section intends to introduce the revised modules and definitions that compose the multi-domain ontology. These modules and their associated agnostic CODPs are used in section 4.6 for the intended resolution of the selected competency question.

It is worth noting that the same modules are reused in a use case formulated for product design. The present use case and the product design use case means to fulfill the qualitative research trustworthiness's transferability criterion (Anney, 2014).

The module descriptions contained in table 4.4 are drawn from (Fitzpatrick, Ratté, et al., 2018a). The reader will find more details relevant to the agnostic modules and the CODPs they contain in the project's SLR. The SLR comprises the agnostic CODPs for all modules with definitions.

Table 4.4 Descriptions of the revised agnostic multi-domain modules
(Fitzpatrick, Ratté, et al., 2018a)

Module name	Module description
Party	«The Party CODP allows the conceptualization of the nature of a person and an organization».
Product	«A good or service resulting from a process. The UN PCS and NAPCS classification schemes can notably be used as taxonomies for products. The concept of bill of material allows to package products».
Contract	«The Contract CODP allows the conceptualization of an agreement between parties playing roles».

Table 4.4 Descriptions of the revised agnostic multi-domain modules
(Fitzpatrick, Ratté, et al., 2018a) (continued)

Module name	**Module description**
Price	«The Price CODP allows the conceptualization of the notion of rates, rate packages, fees, penalties, pricing curve (time varying cost structure) applicable to the consumption of products».
Event	«The Event CODP allows the conceptualization of the notion of a spatiotemporal occurrence that may affect a thing by changing its state».
Document	«The Document CODP allows the conceptualization of physical or electronic representation of a body of concepts in a context».
Network	«The Network CODP allows the conceptualization of a Petri-like structure composed of two nodes and a segment linking the nodes for the purpose of transport of: energy, cargo, people, voice, data, etc. A grouping of networks is also a network».
Account	«The Account CODP allows the conceptualization of a thing used for recording transactions for the purpose of procuring products or tallying quantities for financial statements».
Concept	«The Concept CODP allows the conceptualization of a man-made imaginary construct that corresponds to real life imaginary or physical things».
Context	«The Context CODP allows the conceptualization of a set of things such as location, parties, products and events that grouped together may influence the use of vocabularies, chain of future events».
Location	«The Location CODP allows the conceptualization of a thing related to a coordinate system such as Earth location systems. This includes the notion of area, segment and grid locations. Geography also includes the notion of street addresses and electronic locations such as email and IP addresses».
Role	«The Role CODP allows the conceptualization of a form of involvement in a process or into anything other than a role. A thing playing a role would exhibit a behaviour that may not be related to its nature».
Process	«The Process CODP allows the conceptualization of a form of a unit of work in which resources are used in the fabrication of goods or in the rendering of services. A process can be performed by humans, by nature or a mix of both».

Each of the described modules comprises primary and secondary agnostic CODPs that can used in the resolution of the competency question in section 4.6. These modules are designed to solve a specific semantic problem such as in the case of product. The product agnostic CODP can be used to conceptualize any domain-specific concepts not only associated with the PLM paradigm but all other domains or industry sector as well. In the next section, the competency question is executed to show that any of the product design domain-specific concepts can be subsumed by a agnostic CODP.

4.5 Business process definition for collaborative product design

The business processes defined in this section are derived from the papers cited in section 4.3 'Related work'. These business processes establish the need for system interoperability. These sets of business processes are designed as realistic examples only for the present use case. This paper considers the business processes for SBD are far more complex, covering for example notions such as eco-friendly product design and sustainable product development as investigated in (Perry, Bernard, Bosch-Mauchand, LeDuigou, & Xu, 2011). The purpose of this paper is to illustrate a use case for data integration using the multi-domain ontology.

Figure 4.5 illustrates the business processes in sequence for collaborative product design, although they may be sometime executed concurrently. The Archimate notation is used to represent the business processes (Lankhorst et al., 2009). Table 4.5 provides the definitions for these business processes. These business processes represented in this process model represents a realistic backdrop to the resolution of the competency question.

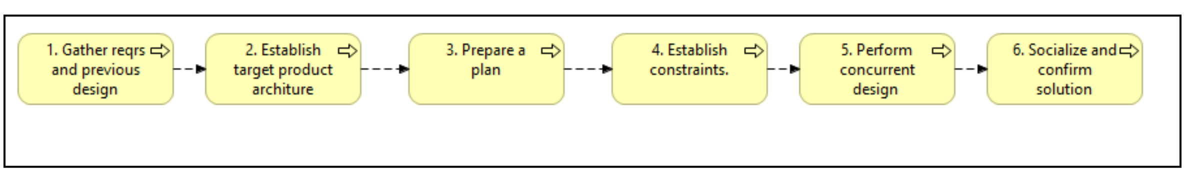


Figure 4.5 Business processes for collaborative product design

Table 4.5 Business process descriptions

Business process name	Business process description
1. Gather requirements and previous design projects data, information, knowledge and know-how.	Collect the needs relevant to the new product from the product lifecycle manager. Also, collect all available ontologies, information and data about the previous design projects that are relevant to the new project.
2. Establish target product architecture and modules.	Formulate a holistic representation of the product and determine major components as modules.
3. Prepare a plan.	Draw a named list of steps with timing and resources to design the intended product.
4. Establish constraints.	Identify the constraints for the new product.
5. Perform concurrent design and converge	Use the SBD approach concurrently with the modular approach to perform several design processes for each module of the product.
6. Socialize and confirm solution.	Present the solution to the stakeholders and get the sign-off from the product internal customer.

The next section, executes the business processes described in figure 4.5 and in table 4.5. Each business process represents the agnostic CODPs, associated with domain-specific concepts that can be used for the execution of the competency question. It is important to note, given space constraints, that only a subset of possible agnostic CODPs and low-abstract domain specific concepts are shown. The assertions used to actually perform the work are not in scope for the present use case.

4.6 Competency question resolution

As indicated also in the use case contained in (Fitzpatrick, Coallier, et al., 2018), the business processes from the previous section and the required agnostic CODPs are represented. The agnostic CODPs (coloured shaded) and the domain specific concepts (grey shaded) are in Archimate notation diagrams (Lankhorst et al., 2009). Each business process illustrated here is described using a template proposed in (Gangemi et al., 2007). The competency questions from section 4.1 Introduction is:

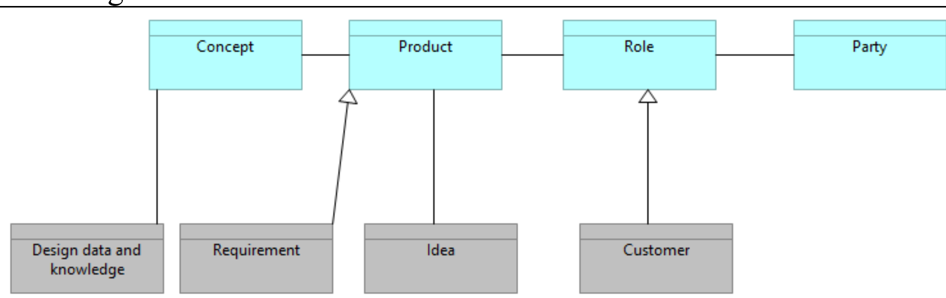
« *What are the factors for each phase or business process of product design, which may influence the financial, customer and environmental value of the new product currently under development?* ».

In the first business process, the inferential application collects knowledge and know-how relative to previous similar product design and development projects, along with the new product requirements. Then, it may infer a target architecture and module specifications. Following plan preparation, the application establishes the constraints and outlines the detail design process collaboratively. It then supports the convergence toward unique design for each module. Finally, the plan is finalized and socialized.

4.6.1 Gather requirements and previous design projects data, information, knowledge and know-how

The first step is to collect business and technical needs applicable to the new product. Also, any relevant content from previous product design processes, along with events such as new product introduction by the competition and legal cases are searched and gathered.

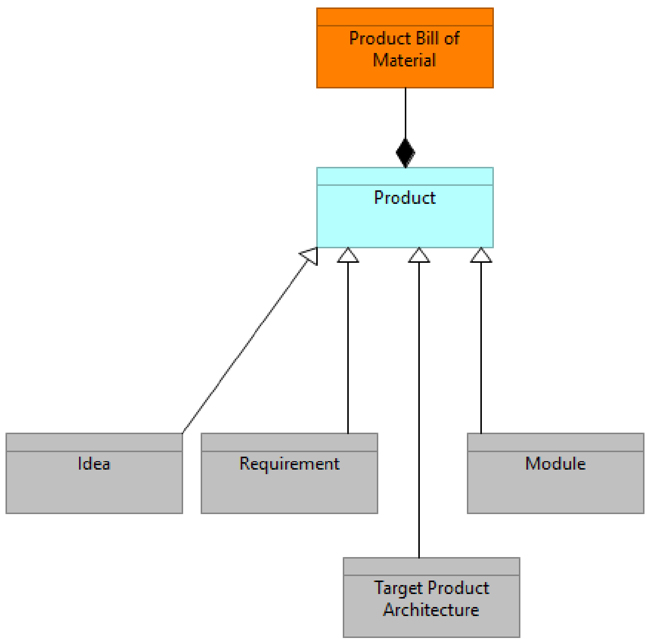
Table 4.6 Gather requirements and previous design projects data

	Use of agnostic CODPs for business processes
Name	1. Gather requirements and previous design projects data, information, knowledge and know-how.
Simplified UML diagram (Archimate)	

4.6.2 Establish target product architecture and modules

The second step consists in formulating a product vision global vision and breaks it down in modules.

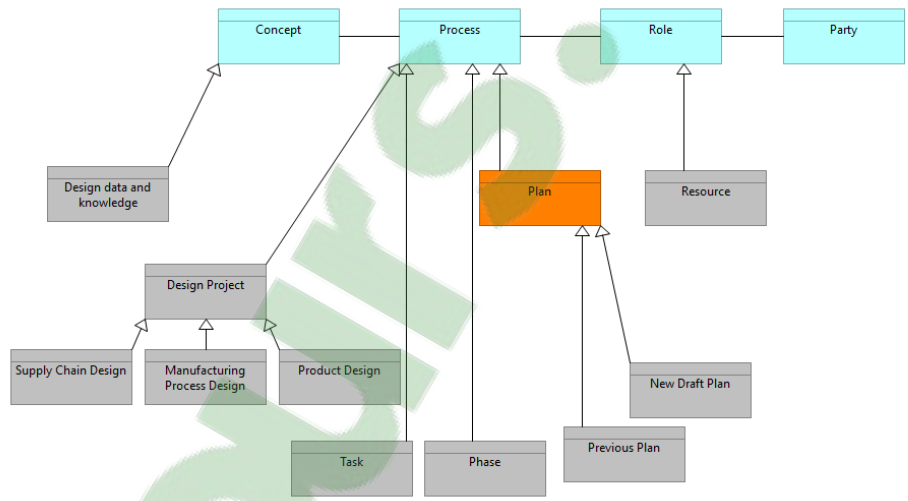
Table 4.7 Establish target product architecture and modules

	Use of agnostic CODPs for business processes
Name	2. Establish target product architecture and modules.
Simplified UML diagram (Archimate)	

4.6.3 Prepare a plan

The third step intends to elaborate the design plan.

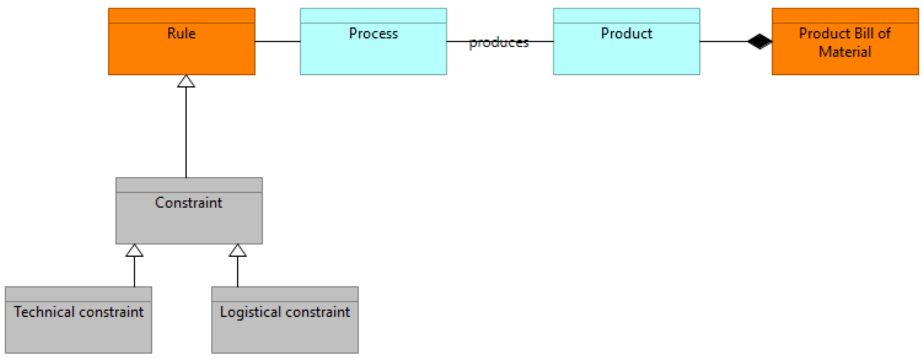
Table 4.8 Prepare a plan

	Use of agnostic CODPs for business processes
Name	3. Prepare a plan.
Simplified UML diagram (Archimate)	

4.6.4 Establish constraints

The fourth step identifies the constraints to be considered during the product development process.

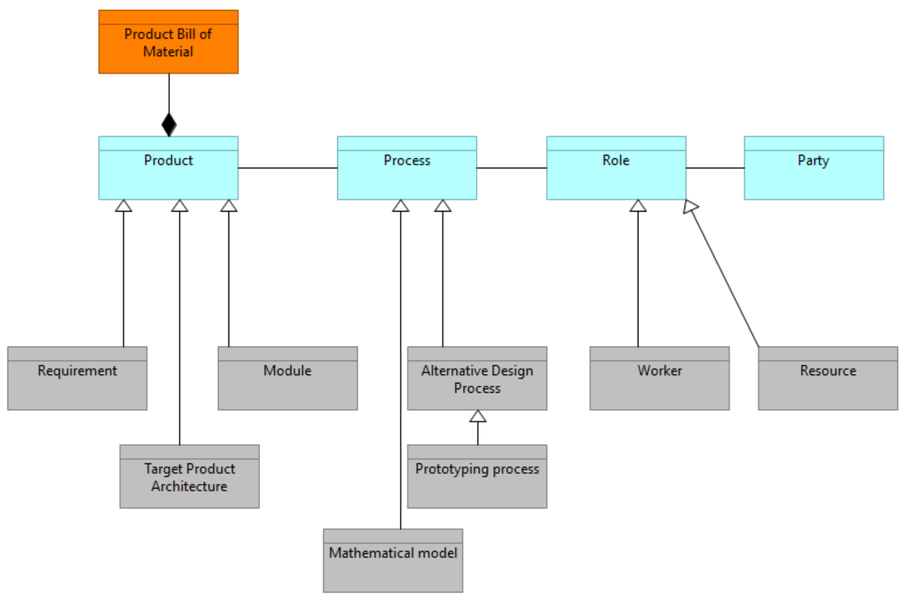
Table 4.9 Establish constraints

	Use of agnostic CODPs for business processes
Name	4. Establish constraints.
Simplified UML diagram (Archimate)	

4.6.5 Perform concurrent design and converge

The fifth step involves the execution of concurrent design processes and their convergence based on efficiency.

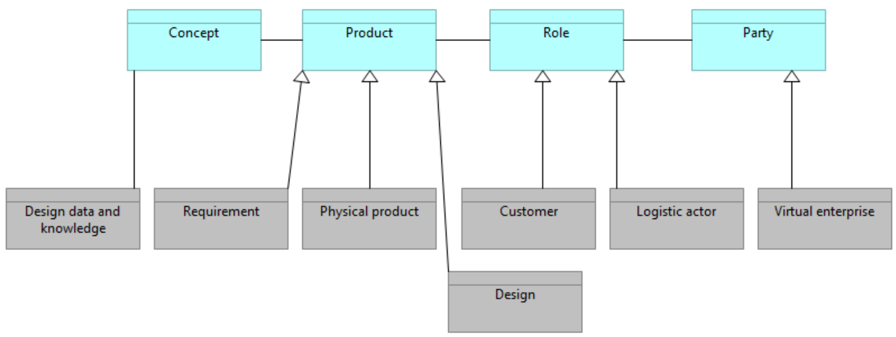
Table 4.10 Perform concurrent design and converge

	Use of agnostic CODPs for business processes
Name	5. Perform concurrent design and converge
Simplified UML diagram (Archimate)	

4.6.6 Socialize and confirm solution

The sixth step involves exposing the product design to the virtual enterprise's stakeholders involved in the project and obtaining a sign-off from the business (internal) customer.

Table 4.11 Socialize and confirm solution

	Use of agnostic CODPs for business processes
Name	6. Socialize and confirm solution.
Simplified UML diagram (Archimate)	

4.7 Conclusion

The competency question resolution illustrates the use of agnostic CODPs for each step and represented the mappings between the domain specific concept and the agnostic CODPs. This allows determining to what extent the multi-domain ontology, and its included set of patterns can support the various and numerous domain ontologies involved in the collaborative design processes.

As indicated in (Fitzpatrick, Coallier, et al., 2018), the competency question resolution executed in section 4.6 reflects the utilization of agnostic CODPs for each business process and mapped the domain specific concepts to the agnostic CODPs. This indicates how the proposed multi-domain ontology can align with the several domain ontologies involved in the collaborative product design processes. For example, all of the planning and execution processes, actual and planned, can be conceptualized and represented while using much fewer concepts patterns with the set of CODP contained in the multi-domain ontology. The semantic structure of the agnostic CODPs are detailed in the SLR (Fitzpatrick, Ratté, et al., 2018a).

This also allows us to demonstrate the transferability of the proposed set of Agnostic CODPs (Anney, 2014) as in the case of this project other use case covered in (Fitzpatrick, Coallier, et al., 2018). This is done by showing that any domain-specific concept discussed in the present paper can be subsumed by an agnostic CODP. This demonstration also shows what additional work needs to be performed after the completion of this research to prepare the multi-domain ontology for further development and testing.

CHAPTER 5

ELICITING AGNOSTIC CONTENT ONTOLOGY DESIGN PATTERNS FOR ENTERPRISE SEMANTIC INTEROPERABILITY USING A PHENOMENOLOGICAL RESEARCH METHOD

Daniel Fitzpatrick[1], François Coallier[1], Sylvie Ratté[1]

[1]Department of Software Engineering & Information Technology, École de technologie supérieure,
1100 Notre-Dame West, Montréal, Quebec, Canada H3C 1K3

Paper submitted for publication to Engineering applications of Artificial Intelligence in April 2018

Abstract

In all private and government sectors, the semantic heterogeneity problem constitutes an important roadblock to organizations' efforts to implement systems interoperability. Semantic heterogeneity, an unnecessary ill, originates from application systems designed with different vocabularies or data models within an enterprise. Systems interoperability represents a crucial capability to the industry and government sectors. This paper is one of the deliverables in a research project that aims to contribute in building the theory needed to solve this problem. This paper's research approach draws from Clark Moustakas' phenomenological research methods. Clark Moustakas' phenomenological research methods, applied in clinical psychology, elicit theoretical material through the experience of participants Moustakas referred to as co-researchers. The concept of abstract, or agnostic, concepts used for data integration represents the studied phenomenon. A series of twenty-two semi-structured interviews are held to elicit co-researchers' beliefs in relation to agnostic concepts that can be used across all industry or government sectors. The co-researchers are experienced professionals with over eight years experience in conceptualization.
The analysis involves extracting the sought meaning units: the agnostic concepts, their definitions and relationships. The "low-abstract" domain specific concepts and the subsumption relationships are also elicited. Once the analysis step is completed, the emerged meaning units from the transcripts are coalesced into integrated structures. The outcome of

the synthesis phase is the set of agnostic CODP templates that are significantly similar to the set of agnostic CODPs elicited in this paper's companion publication (Fitzpatrick, Ratté, et al., 2018a) a Systematic Literature Review (SLR). The establishment of such similarity in the outcome of both publications constitutes a triangulation, a key criterion to determine the trustworthiness of the current qualitative research methodology.

Keywords: Content ODP, Ontology Design Patterns, Ontology, inference application, multi-domain ontology, phenomenological research method, trustworthiness, constructivism.

5.1 Introduction

In all private and government sectors, the semantic heterogeneity problem constitutes an important roadblock to organizations' efforts to implement systems interoperability. Semantic heterogeneity, an unnecessary ill, originates from application systems designed with different vocabularies or data models within an enterprise. Systems interoperability represents a crucial capability to the industry and government sectors. Also, since life science research needs interoperability between its systems as well, there is logically a cost in human lives stemming from valuable medical and pharmaceutical research funds wasted in addressing semantic heterogeneity (Lenz et al., 2012). In (Williams et al., 2012) and (Mirhaji et al., 2009) the authors stress that efforts in deploying data integration pose significant challenges in biomedical research and hinders knowledge discovery critically needed to develop new drugs. Either academia or the industry has resolved the semantic heterogeneity problem (Doan et al., 2012) (De Giacomo et al., 2018).

This paper is one of the deliverables of a research project that aims to contribute in building the required theory needed to solve the problem. This paper's research approach draws from Clark Moustakas' phenomenological research methods. Clark Moustakas' phenomenological research methods, applied in clinical psychology, elicit theoretical material through the experience of participants Moustakas referred to as co-researchers.

In this paper, the concept of abstract, or agnostic, concepts used for data integration represents the studied phenomenon. A series of semi-structured interviews elicited co-researchers' beliefs in relation to agnostic concepts that can be used across all industry or government sectors. The co-researchers are experienced professionals with over eight years experience in conceptualization, as proposed by (S. Ahmed et al., 2005). The co-researchers were interviewed to provide knowledge, in addition to agnostic data model patterns, such as their appreciation on the involvement of non-technical business stakeholders in designing data integration platforms. This paper richly describes a qualitative research approach to elicit from experienced professionals a set of agnostic patterns to design a multi-domain ontology, as first proposed in (Fitzpatrick, 2012). The concept of multi-domain ontology, a type of mid-level ontology, has also been proposed previously in (Daniel Fitzpatrick et al., 2012, 2013; D. Fitzpatrick et al., 2013). As proposed by (Gangemi & Presutti, 2009), (semi-formal) data model and UML patterns can serve as the basis for creating a formal ontology. Such data model and UML patterns can then be transformed into (formal) Content Ontology Design Patterns or CODPs (Blomqvist 2010).

This paper's phenomenological research approach collects agnostic concept patterns from experienced practitioners. These practitioners have conceptualized in their careers to produce data models, domain models and other types of schemas (semi-formal ontologies) usually applied in (non-cognitive) contemporary information technologies, such as relational databases. The axiomatic form of these patterns would constitute collectively the multi-domain ontology as defined in (Daniel Fitzpatrick et al., 2013).

In section 5.2, we start with Related work. Section 5.3 provides the Definition of terms section from (Fitzpatrick, Ratté, et al., 2018a) that describes the fundamental concepts of this project. Section 5.4 Problem Statement formulates the project's primary uncertainty that it intends to address. Section 5.5 formulates the objective of this research. Both sections 5.4 and 5.5 are also drawn from this project's Systematic Literature Review (SLR) (Fitzpatrick, Ratté, et al., 2018a) since this project uses a dual research method approach, i.e. SLR and the current paper's phenomenological method, to establish triangulation. Section 5.6, Research

Method, comprises subsection 5.6.1, Research Protocol, which describes the phenomenological methodology used in this paper. Section 5.7, Research Question, describes the intended inquiry at the heart of this paper, also drawn from and shared with the SLR. Section 5.8, Content Analysis, describes the findings from the systematic examination of the semi-structured interviews' recording. Section 5.9, Content Synthesis, presents statistical information and light UML (Archimate notation) diagrams with accompanying descriptions for each derived agnostic CODP. Section 5.10 concludes the paper with a discussion on the executed phenomenological method's outcome and the research project's next steps.

5.2 Related work

In (Diego Calvanese, De Giacomo, Lembo, Lenzerini, & Rosati, 2009), the authors propose a data integration approach based on «the global schema (that) provides a conceptual representation of the application domain … as presented to the client». An enterprise may comprise several domains (Anonymous, 2009). Each domain, as «*separate islands of data*» comprises several applications and services its own (internal) clients (Rosenthal, Seligman, Renner, & Manola, 2001). Also, each domain has its own vocabulary possibly different from other domains (Corry, Coakley, O'Donnell, Pauwels, & Keane, 2013) . Although it may cover several applications systems, a domain still constitutes a silo (Malan & Bredemeyer, 2002). A data integration approach based on a conceptual representation of an application domain as advocated by (Diego Calvanese et al., 2009) would still foster semantic heterogeneity. A different approach based on a broader conceptualization, i.e. cross-industry, offers potentially a more effective solution path to semantic heterogeneity.

Other research efforts, such as in (Simsion et al., 2012) and (Anglim et al., 2009) involve interviews or surveys to acquire knowledge from data modelers. Both these studies use qualitative research in a similar fashion as performed in the present paper. (Anglim et al., 2009) cover the practice of data modeling specifically in respect to current and future trends by interviewing twenty-two experienced data modelers. The latter research reached out to the practitioners by contacting professional associations. (Simsion et al., 2012) use both surveys, with practitioners, and semi-structured interviews with named data modeling "thought leaders". The latter research elicited practitioners' insight to determine if data modeling was performed to either describe business concepts or to design databases. Following the synthesis of the survey and interview data (Simsion et al., 2012) concluded that data modeling was better characterized as design.

This paper is one of the deliverable of a project, which for the first time uses concurrently two qualitative research methods: SLR and phenomenological. This approach intends to demonstrate the research methodological trustworthiness. Also, this research also is the first to elicit agnostic CODPs for a multi-domain ontology.

5.3 Definition of terms

The following definitions are taken from the present paper's companion SLR method publication (Fitzpatrick, Ratté, et al., 2018a), with the exception of definition 2.2 on data integration that is native to this paper. The following definitions provide a better understanding of the underpinnings to this research.

5.3.1 Conceptualization

Conceptualization is defined here as a language-independent process that implicitly creates semantic structures. Semantic structures establish the meaning of things. Semantic structures are a set of concepts, properties and their relationships. Pierdaniele Giaretta and Nicola Guarino define conceptualization as «*an intensional semantic structure which encodes the implicit rules constraining the structure of a piece of reality*» (Giaretta & Guarino, 1995). Guarino also refers to a conceptualization as an «*intended meaning of a formal vocabulary*» (Nicola Guarino, 1998).

5.3.2 Data Integration

The elusive notion of data integration represents a challenge to both scientific and industry realms along the great difficulty to develop it (Doan et al., 2012). In (Bennett & Bayrak, 2011), the authors define a data integration system as a «*general-purpose (application) used to provide interoperability among autonomous heterogeneous database systems*». Later in the same article, the authors refer to data integration as a «problem». In (Lenzerini, 2002), the authors define data integration as «*the problem of combining data residing at different sources, and providing the user with a unified view of these data*».

This paper's project defines data integration as a software application that intends to solve the semantic heterogeneity problem in allowing an enterprise's systems to interoperate. In other words, the problem is semantic heterogeneity, the affected capability is interoperability and the solution is data integration. Since semantic heterogeneity is not currently solved, data integration is considered here as a palliative measure. Current scientific research on data integration aims to develop data integration as a commoditized technology (Doan et al., 2012).

5.3.3 Representation

It is an externalized depiction, or language-dependent specification, of concepts that can be shared amongst people and machines. Representing concepts involves converting implicit concepts lodged in a person's brain into explicit concepts using a language. For example, domain ontologies that are created to share a vocabulary amongst a community are represented using one or several of the following languages: natural, concept map, SQL, XSD, OWL, etc. The represented domain ontology is submitted to the members of its community through a consensus-building process to be officially recognized and used accordingly. Nicola Guarino defines a representation or a specification of an ontology as «a logical theory accounting» (Nicola Guarino, 1998).

5.3.4 Ontology

Gruber defines an ontology as an «*explicit specification of a conceptualization*» (Thomas R. Gruber, 1993). It aims in providing a shareable and reusable knowledge to be used by people and computer systems. Ontologies would favor the trend toward a greater universal interoperability across all industries. Conceptualization is independent of the notional language. However, an ontology's specification, or representation, is dependent of a language. An ontology is a logical theory that describes the intended meaning to its defined vocabulary, in other words, using the committed concepts to a particular conceptualization of the real world. Guarino stresses that ontologies only approximate a conceptualization. He also indicates that the only way to enhance the representation is to develop a richer set of axioms (N. Guarino, 1998). The search for a richer set of axioms explains this research project's interest for data model patterns for multi-domain data integration developed in the industry for acquiring the sought semantic richness.

All ontologies may be classified in five types:

- Top level or foundational ontologies, such as Cyc, SUMO and Proton describe some of the basic objects of reality such as time, matter, action, etc. These concepts are independent of a particular problem or domain. This type of ontology supplies the fundamental concepts serving as the basis to define the other type of ontologies;
- Mid-level ontologies such as the multi-domain ontology as proposed by (Daniel Fitzpatrick et al., 2012), are described by (Obrst et al., 2012) as being «less abstract (than foundational ontologies) and span multiple domain ontologies. Mid-level ontologies also encompass core ontologies that represent commonly used concepts, such as Time and Location». Core ontologies may be voluminous and can be more difficult to develop (Gangemi & Presutti, 2009);
- Domain ontologies represent the vocabulary of a generic domain that may exist in several organizations;

- Task ontologies describe a generic process structure that can be used to solve a certain type of problem;
- Application ontologies, which describe semantic entities that stem from a domain and task ontology or ontologies, both providing a specific function context (N. Guarino, 1998).

There are essentially three types of ontology applications:

- To support the mediation between people and ontology representing a vocabulary for the exchanges between people and organizations;
- Domain interoperability: support to develop (development time application) or to operate (run time application) systems of the same or different domains;
- Knowledge reuse: requires the highest level of rigor, in addition to axioms, other concepts and their properties; ontologies for knowledge reuse will rely heavily on constraints and other types of restrictions. Problem solving methods or PSM have the capacity to support shared knowledge. They often include generic algorithms to perform various functions within the domain. One type of application that is growing in popularity in the research domain is ontology-based information extraction through Natural Language Processing (NLP) (Navigli & Velardi, 2008; Völker et al., 2008; Wimalasuriya & Dou, 2010). In (Ratté et al., 2007), NLP processes are proposed to extract information from the organization's internal documents. These aspects constitute key elements behind the proposed Reference Architecture – Enterprise Knowledge Infrastructure (Daniel Fitzpatrick et al., 2013).

Figure 5.1 illustrates the two basic facets of the ontology concept: language dependent and language independent characteristics.

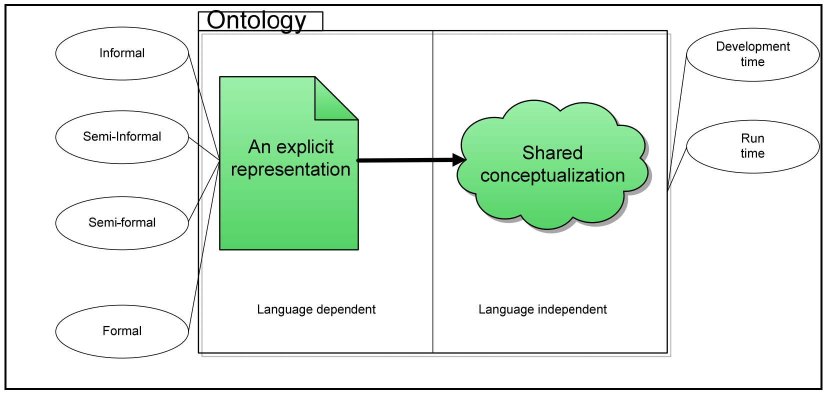


Figure 5.1 Summarized definition of an ontology

An ontology does not impose the application of properties to a given instance of a class or concept. The finality here should be to build libraries of reusable knowledge and knowledge services available on networks. Ontological commitments or agreements pertaining to classes and relationships of an ontology are discussed among software agents and knowledge bases (T. R. Gruber, 1993). A concept definition is a human readable text that in itself provides significance, meaning therefore semantically whole (Gruber et al., 2009) (Noy & McGuinness, 2001).

An effective equilibrium must be achieved in defining ontology constrains rules in order to avoid affecting the concept abstraction level in the ontology even if it supports interoperability in a more effective manner. Affecting the ontology's abstraction level may lower the robustness and flexibility of the vocabulary (Spyns et al., 2002).

Semantic relationships are categorized as synonymy, antonymy, hyponymy, meronymy and holonymy relations. Synonymy relationships relate two similar concepts. An antonymy relation indicates opposing or disjoint concepts. The Hyponymy category pertains to a generic to specific relationship between concepts. The meronymy and holonymy relationships support the build of material structure between concepts, the former indicates that a concept is included in another one, while the latter indicates that a concept includes the object of the relationship. Figure 5.2 illustrates the conceptualization aspect of an ontology that is language independent (Lacy, 2005) (Nicola Guarino, 1998).

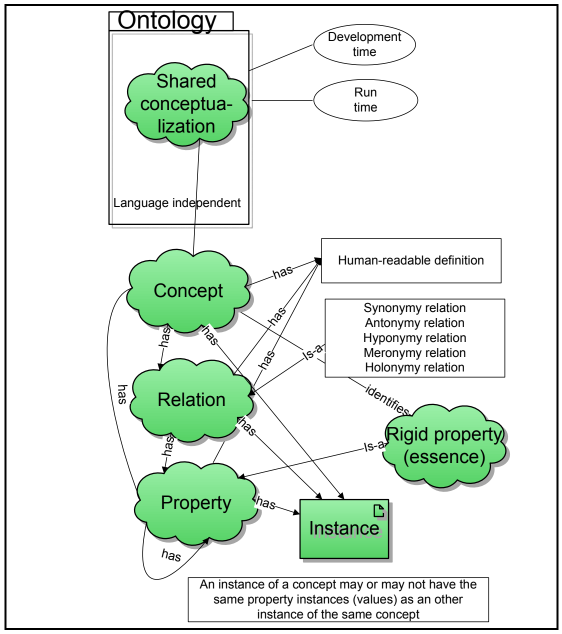


Figure 5.2 Language independent aspect of ontologies

Ontologies can be used to solve syntactic and semantic problems, and to automate data integration. Some of the ontologies are designed to be processed by inference engines and written in first-order logic-based specialized languages such as OWL, RDF, RDFS, PLIB and SWRL. Some of these formal ontologies have grown to be voluminous and are becoming difficult to execute in main memory. A hybrid solution has been proposed by both academic and industrial organizations to address the in-memory loading of voluminous ontologies (Khouri & Bellatreche, 2010).

Figure 5.3 illustrates the language dependent aspects of ontologies. In terms of their level of formalism, there are: highly informal, semi-informal, semi-formal and formal ontologies. The first level of formalism is the highly informal level. It refers to a natural language text. In the case of semi-informal ontology is represented as a restricted and structured form of natural language, such as a concept map. In a case of a semi-informal ontology, the vocabulary would be expressed in an artificial language such as pseudo-code. Semi-formal ontologies include entity relationship diagrams, UML domain models and XML Schema Definition

(XSD). Finally, at the formal level, ontologies are logical rule sets that can be processed by an inference reasoner. Such formal ontologies possess:

Meticulously defined terms with formal semantics, theorems and proofs of such properties as soundness and completeness, i.e. classes including property information, value restrictions, more expressivity, arbitrary logical statements, first order logic constraints between terms and more detailed relationships such as disjoint classes, disjoint coverings, inverse relationships, part and whole relationships, etc (Xie & Shen, 2006). An example of a commercially available semantic technology architecture, produced by Oracle, can be found in (Wu et al., 2008).

Formal ontologies can be based on first-order logic, frame-based constructs or both. (A. Gómez-Pérez et al., 2004; Lacy, 2005) The concept of multi-domain ontologies has been researched to facilitate the exchange of data, information and knowledge between domains (Jinxin et al., 2002).

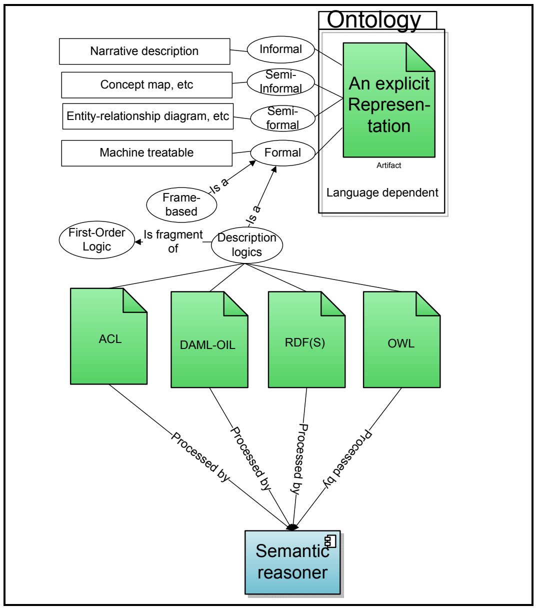


Figure 5.3 The language dependent aspect of ontologies

5.3.5 Pattern

Alexander introduces the notion of pattern in defining it as a generic solution to a recurring problem from the building architecture domain (Alexander, 1977) (Alexander, 1979). Later in 1993, the software engineering scientific community adapted the pattern concept to object-oriented design (Gamma et al., 1993). (Poveda et al., 2009) indicates that its fundamental meaning of a pattern pertains to something that can be imitated, that can serve as a starting point.

5.3.6 Ontology Pattern

Blomqvist defines an ontology pattern as «*a set of ontological elements, structures or construction principles that intend to solve a specific engineering problem and that recurs, either exactly replicated or in an adapted form, within some set of ontologies, or is envisioned to recur within some future set of ontologies*» (Blomqvist, 2010). In the present research, semantic heterogeneity constitutes the specific engineering problem.

This project excludes structural ontology patterns since foundational concepts are excluded. Also, ontology architecture patterns are excluded since the project considers concepts and relationships other than what is found strictly in a taxonomy (Blomqvist, 2009b). (Blomqvist, 2010) considers that ontology architecture patterns only cover the ontology as a whole or modules, but not specific concepts or relations. This SLR only covers ontology design patterns that are related to business concepts and that agnostic, i.e. applicable to any industry or domain.

5.3.7 Ontology Design Pattern (ODP)

An Ontology Design Pattern is a «set of ontological elements, structures or construction principles that solve a clearly defined particular modeling problem» (Blomqvist, 2010). It is a pattern used for the formulation of an ontology to be processed by a reasoning application. ODPs are represented as axioms in a specialized language such as OWL, a derivative of the XML language, for the purpose of logical, or inferential, processing. However, for the purpose of publication, an ODP can be represented in a natural language, concept map, UML, etc. This article uses the Archimate architecture modeling formalism, a simplified derivative of the Unified Modeling Language (UML), to represent the CODPs for the proposed multi-domain ontology.

5.3.8 Content ODP

According to (Gangemi & Presutti, 2009) (Blomqvist, 2009a), a content ODP, or a CODP, is a design pattern that addresses business concepts found in a domain ontology. This article represents CODPs that correspond to business concepts that are meant to be applicable to all domains.

5.3.9 Enterprise

According to The Open Group Architecture Framework (Anonymous, 2009), an enterprise can be a commercial profit driven entity, a no-profit organization or a government agency.

An enterprise can also be a group of organizations such as a coalition or a partnership. A subdivision of another enterprise such as an affiliate company or department of a government can be considered as an enterprise.

5.3.10 Domain

A domain represents a community or collection of knowledge and know-how shared by a group of individuals within an enterprise, across an industry or universally (Tennis, 2003).

5.3.11 Abstract concept

An abstract concept is defined as the quality of a general concept that can be instantiated in several forms depending on a given context. In the context of this article, the sought abstract (agnostic) concepts from the elicited data model patterns can apply to any domain.

5.3.12 Agnostic concept

An agnostic concept is defined here as an abstract concept that possesses a distinct definition amongst other concepts. Thomas Erl defines the term Agnostic in the context of Service Oriented Architecture software component logic as logic that is reusable across all contexts and domains in the enterprise (Erl et al., 2017). Furthermore, it is implied here that an agnostic concept is defined in such a way that it cannot be confused with another agnostic concept.

5.3.13 Multi-domain ontology

A mid-level formal ontology composed of a collection of interrelated agnostic CODPs that allow a cross-industry conceptualization (Daniel Fitzpatrick et al., 2012). Concepts related to any industry may be represented using the multi-domain ontology. The primary purpose is to ensure interoperability between an enterprise's application systems.

5.4 Problem statement

This problem statement is drawn from this paper's companion SLR method publication (Fitzpatrick, Ratté, et al., 2018a). Semantic heterogeneity hampers enterprise application systems' interoperability. Semi-formal and formal ontology-based data integration solutions have yet to be successful and commoditized (Doan et al., 2012). Furthermore, the ontology engineering research community, albeit significant advancements that were made, still cannot consensually formulate a single unifying definition of an ontology, the prime element of a theory (Welty, 2003).

The semantic heterogeneity problem constitutes a cost of palliative measures that do not provide any added business value. Since the life sciences' research including the medical domain is equally affected by this problem, it is reasonable to assert that quality of life and even the capacity to preserve and save lives may also be affected by this problem. In (Laínez et al., 2012), the authors raise the issue that the pharmaceutical research domain is data rich but knowledge poor. We stipulate that semantic heterogeneity may affect the pharmaceutical research domain, notably, in its capacity to convert raw data into insight.

5.5 Research Objective

The research objective is also drawn from this paper's companion publication (Fitzpatrick, Ratté, et al., 2018a), both executed for the same research. This research aims to elicit data model patterns from experienced practitioners. The data model patterns are to be re-engineered as agnostic CODPs and to compose the multi-domain ontology. Although data model patterns are only used in semi-formal ontologies, e.g. database and software design, they can contribute for building formal ontologies, such as the multi-domain ontology (Blomqvist, 2010).

This paper specifically deals with ontology patterns that can be found in the conceptualization of semi-formal ontologies, for example in an object-relational database schema or a canonical model. The sought semi-formal ontology constructs enact semantic interoperability allowing the enterprise's application systems to work jointly intra and extra organizationally. This paper's phenomenological method seeks to elicit existing conceptualization patterns that transcend any representation form (semi-formal vs. formal) and that are industry agnostic.

5.6 Research method

In their 2013 article titled «Where's the Theory for Software Engineering?», Ivar Jacobson, co-creator of the Unified Modeling Language (UML) and pioneer of the software engineering community, and co-authors reached out to researchers to « *rise from the drudgery of random action into the sphere of intentional design… We just need to subject (software engineering) to the serious scientific treatment it deserves*» (Johnson, Ekstedt, & Jacobson, 2012).

Jacobson and his co-authors also cited the «thoughtful» works by Shirley Gregor in describing the components of what constitutes a theory: descriptive, explicative, predictive and prescriptive (S. Gregor, 2006). In executing a phenomenological research method, this paper's project contributes to the descriptive and explicative components of an emerging theory.

The researcher finds that a research approach based on the phenomenological method, as pioneered by Clark Moustakas (C. Moustakas, 1994), would be the most appropriate and effective to fulfill this project's research objective and, consequently, building theory.

The phenomenology-inspired research protocol described in this paper involves a series of twenty-two semi-structured interviews (Patton 2002) to collect agnostic concept patterns related to the implementation of a data integration capability, complementing the analysis of the available technical documentation as performed in this paper's companion SLR publication (Fitzpatrick, Ratté, et al., 2018a).

In addition to allowing the extraction of more and richer pattern-like information throughout the field research part of the project, the phenomenological approach provides two other important benefits: it assists the researcher to better select the interviewees («first persons») and allows the researcher to submit himself or herself to a very rigorous and effective preparation to better conduct interviews and focus group sessions (Tesch, 1990).

(P. D. Leedy & Ormrod, 2005) states that qualitative research is needed to build theory. Although some work of scientific quality is performed, it barely scratches the surface to describe the descriptive and explicative aspects of a theory.

This field research method uses semi structured interviews, based on the phenomenological research design as practiced in the social sciences, psychology (C. E. Moustakas, 1994), in Information Systems (IS) (Bharadwaj, 2000) and in Information Technologies (IT) (Introna, 2005). A phenomenological research method involves the individual interviews of 'first-persons', persons that have actually participated in a phenomenon (Patton, 2002) (Tesch, 1990). The phenomenon here for this project is a multi-domain data integration capability, as perceived and lived by experienced practitioners.

The research protocol described in section 6.1 mirrors the research approach used in this project's SLR in several aspects. Both research methods, i.e. SLR and phenomenological, follow the same techniques for the analysis and synthesis stages. The exceptions, i.e. the differences between the SLR and phenomenological methods, are:

The techniques used to select the knowledge sources. In the case of the SLR, a practical screen is designed to systematically and rigorously select the publications to be studied to answer the research question. In the case of the phenomenological method, the selection criterion, for example, targeted practitioners with a minimum of eight years' experience in conceptualizing that speaks either French or English;

The elicitation of the knowledge performed on the knowledge sources. In the case of the SLR, a note-taking approach allows to extract the sought concepts from publications. In the case of the phenomenological method, notes are taken and the conversations are recorded.

5.6.1 Research protocol

The selection of the research method was guided by (P. Leedy & Ormrod, 2012), (Hays & Wood, 2011) and (Starks & Brown Trinidad, 2007). The phenomenological research method is selected to elicit knowledge from experienced practitioners. (P. D. Leedy & Ormrod, 2005) states that «In some cases, the researcher has had personal (professional) experience related to the phenomenon in question and wants to gain a better understanding of the experience of others. By looking at multiple perspectives on the same situation, the researcher can make some generalizations of what something is (really) like from an insider's perspective».

The four benefits of the phenomenology research method are according to (C. E. Moustakas, 1994):

- Selecting the right participants;
- Empowering (preparing and accompanying) the participants as co-researchers;
- Extracting & processing rich information;
- Preparing the most important research instrument in this specific qualitative study: the researcher. Especially a 35-year IT veteran, i.e. the researcher, who has likely cumulated preconceived ideas and hardened beliefs over the years. Such bias can adversely affect the trustworthiness of the design and the execution of the research protocol.

In this case, the phenomenology research method constitutes the best-suited approach for the researcher. Although, extensive experience on the subject matter can help the researcher, it can also hinder the objectivity and impartiality required to perform the research protocol. The phenomenology approach allows the researcher to improve interview skills and extensively prepare the rigor, neutral stance and set aside any emotional or other thoughts that may impede on objectivity and impartiality. It allows the researcher to become, on a best-effort basis, a cutting-edge research instrument as much as time and resources permit (P. Leedy & Ormrod, 2012). On the other hand, the researcher must learn to provide the co-researcher a pleasant, relaxing but educative experience.

Researchers typically conduct semi-structured interviews with between 5 and 25 participants when using the phenomenological research method (P. Leedy & Ormrod, 2012). The phenomenology approach seeks to collect data from first persons. First-persons are individuals that have not only first-hand witness the phenomenon but, in the case of the research project, have actually contributed directly and gain the invaluable knowledge and know-how sought in this research not from others but actually performed architecture, design or development work on a multi-domain data integration capability either within a data warehouse environment, a SOA infrastructure or any other architecture style.

Figure 5.4 provides an overview of the research protocol. This overview diagram illustrates using the Archimate notation (Lankhorst et al., 2009) stakeholders, the researcher and co-researchers, and the protocol's processes. This protocol is based on the works of (C. Moustakas, 1994) (Tesch, 1990) and (Patton, 2002).The protocol does not include a pilot project performed previously that allowed to fine-tune the questionnaire. The Preparation step allows designing the questionnaire, locating potential co-researchers and contacting them. The Bracketing step consists in the researcher to explicitly express own beliefs in answering the questionnaire using text and diagrams.

The Interview step involves the researcher and a single co-researcher having a pleasant telephone conversation, for approximately one hour, on the questions listed in the questionnaire. The Transcript step includes note-taking performed during the interview and done afterward from the session recording. Content Analysis consists in breaking down in each transcript the meaning units dissociating them from the conversation's text. The meaning units are classified as one of the following: the main agnostic concept, the subsumed subordinate concepts, the definitions and relationships. The Content Synthesis step integrates the meaning units elicited in each interview following a chronological order, i.e. from co-researcher “CR01”, the first participant to co-researcher “CRnn”, the last participant. Concepts are integrated around the following axes: the main agnostic concept, the subsumed subordinate concepts, the definitions and relationships. When completed, the individual transcripts are sent to the co-researchers for their approval during the Transcript step.

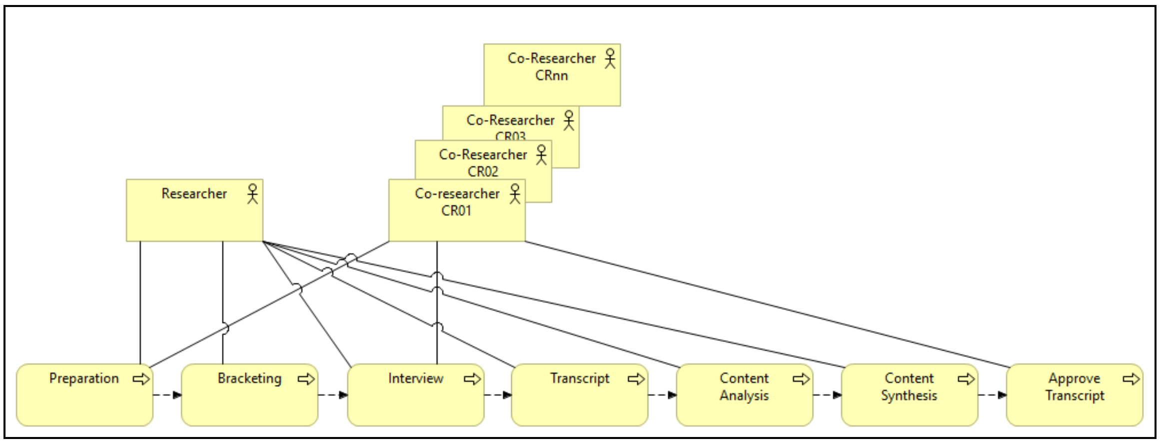


Figure 5.4 Overview of the phenomenological research protocol

The protocol steps, illustrated above, are detailed in the following sections.

5.6.1.1 Preparation.

This protocol step sees the design of the questionnaire. The first set of questions intends to outline the contextual aspect, i.e. the background, of the co-researcher, notably the number of years the participant had experience in conceptualizing as a data modeler, data architect, software engineer, developer, etc. The question about the years of experience allows the researcher to verify that the potential co-researcher meets the minimal years of experience criterion of eight years. The other background question indicates the various industry sectors the practitioner has performed conceptualization. (Suri, 2011) refers to this purposeful sampling approach as criterion sampling. Co-researchers are asked to introduce other potential participants on a voluntary basis, which Suri refers to as snowball sampling.

Snowballing consists in the co-researchers reaching out to the referred potential participants and asked permission to be contacted by the researcher or invited to contact the researcher directly.

The questions listed in table 5.1 pertain to the phenomenon itself, i.e. the concept of agnostic concepts used for data integration and peripheral issues that are often raised in the researcher's experience as an experienced practitioner. The researcher's experience does not influence in any way the outcome of this study, complying to (Bevan, 2014) citing (Husserl, 1970) in refraining in using the researcher's personal knowledge in a phenomenological research method. However, the researcher's knowledge of the phenomenon allows determining peripheral issues such as defining the notions of accuracy and quality of a data integration model. The notion of accuracy and efficiency should be logically examined in a future phase of the project to elicit knowledge on subject-related measures. The measures related to data integration may be the subject of further investigation using a metrology approach proposed by (Abran, 2010). The approach used here to effectively target the phenomenon is done first by having the co-researcher list and describe agnostic concepts, and their relationships, that can apply to any private industry and government sector. Then, a question addresses the same descriptions but for domain specific or "low-abstract" concepts that can apply specifically to a maximum of three industry sectors that the co-researcher has experienced. Also, the participants are asked to relate the domain specific concepts to the agnostic concepts previously described. This allows the participants to identify additional agnostic concepts that may have been previously missed during the interview.

A question explores the co-researcher's beliefs in respect to have agnostic concepts in a data integration model. Another question inquires about having "low-abstract" domain specific concepts in a data integration model. Other questions explore the co-researcher's notions of efficiency and quality of a data integration model. The co-researcher is also questioned about having ever observed business representatives influence in any capacity the design of a data integration platform. Finally, the co-researcher is solicited, as the last item on the questionnaire, to optionally reach out to a colleague for recruiting other co-researchers thus performing snowballing.

Table 5.1 Questions used for the semi-structured interview

Question no.	**Question formulation**
Q01	How many years have you performed conceptualization, e.g. data models, canonical model, domain model, XSD, etc?
Q02	What are the industry and government sectors have you performed conceptualization?
Q03	Name and describe abstract (agnostic) concepts that you believe may apply to any industry and government sector.
Q04	Indicate relationships between these abstract concepts.
Q05	For a maximum of three industry or government sectors, list domain specific (low abstract) concepts and identify to which abstract concept they relate to (generalization specialization only).
Q06	Do you believe that a data integration function should be designed using abstract (agnostic) concepts as you indicated in question 3? Provide a score from 1 to 10. Please comment.
Q07	Do you believe that a data integration function should be designed using low abstract (domain specific) concepts that would be understandable by business users? Provide a score from 1 to 10. Please comment.
Q08	Do you believe the problem of semantic heterogeneity (see the introduction deck) should be addressed by scientific research?

Table 5.1 Questions used for the semi-structured interview (continued)

Question no.	Question formulation
Q09	Have you participated as a designer, architect, developer or software engineer in the development of a data integration core structure for a data warehouse or of a canonical model? This question does not constitute a precondition for the continuation of the interview.
Q10	Did you ever observe line of business influence on the design of a data integration platform? Please comment.
Q11	How do you or would you define and measure the efficiency of a data integration model?
Q12	How do you or would you define and measure the quality of a data integration model?
Q13	Optional snowballing: If willing, could you please refer one or two persons, with conceptualization experience (8yrs+).

Following the design of the questionnaire, current and former colleagues were contacted by the researcher through personal email, personal telephone and social media. Twenty-two qualified practitioners accepted the invitation to be co-researchers in the present phenomenological research. An introduction document is sent, explaining the research and containing inform consent information along with the questionnaire. The co-researchers were informed of their fundamental rights as research participants to withdraw from the process without constraint at any moment and that their identities are kept confidential. Any direct quote from the co-researchers would be identified by a code such as “CR01” assigned in chronological order of the interview. The information package was sent at least two days before the telephonic interview.

5.6.1.2 Bracketing

This step consists in the researcher to explicitly express own beliefs in answering the questionnaire using text and diagrams. Before the start of the first interview, with co-researcher CR01, the researcher answers in writing the questionnaire. The researcher also drew light UML diagrams to represent the agnostic concepts, relationships and associated definitions. Furthermore, the researcher opted not to participate in the phenomenological approach. These measures, the bracketing and abstaining from participation, aim to preserve the integrity of the research process (Bevan, 2014), (C. Moustakas, 1994), (Hays & Wood, 2011).

5.6.1.3 Interview

At the scheduled time, the researcher contacted by telephone the co-researcher to begin the interview. After explaining how the interview would proceed, the researcher requests the permission of the co-researcher to record the conversation. In a very informal setting, the researcher asks the questions and accompanies the co-researcher by clarifying in rephrasing when needed. Furthermore, the researcher performed imaginative variation. Imaginative variation consists in providing a context or adding detail considerations to a question. For example, when asking question Q10 about the influence of business representatives on the design of a data integration platform, the researcher complements the question in asking an immediate follow-up question about a potentially or actually positive and negative positions that the co-researcher may have about “the business getting involved in the design of a data integration platform”. Additionally, when asking question Q06 on using agnostic concepts to design a data integration model, the researcher clarified for some co-researcher that it is assumed that there is no constraint, no politics and no pressure whatsoever. In other words, the co-researcher has complete control over the design of the data integration platform. The imaginative variation technique, widely recognized as a trademark component of the phenomenological research methodology (C. Moustakas, 1994) (Wertz, 2005).

At the end of the conversation, which lasts in almost all cases one hour, most co-researchers agreed that the conversation was pleasant and were looking forward to receiving the summary transcript and the draft article. In all cases, the co-researchers accepted to complete a 15 to 20 minutes follow-up survey, which will be done in a subsequent project. The positive reaction of the co-researchers in the aftermath of the interview is crucial to encourage experienced professionals to participate in such research. (Bevan, 2014) states that «*being in natural attitude is effortless*». The researcher and co-researchers engaged in what amounts to be an effortless, educative and pleasant discussion outside of work settings.

5.6.1.4 Transcript

During the interview, the researcher takes note even during recording. During this note-taking, the researcher noted the agnostic concepts, their relationships, and the domain specific concepts with generalization-specialization relationships with agnostic concepts, along with a summary of the responses from the other questions (Q06 through Q13).

Following the interview, the sought material from the recording was extracted and written in transcript documents. The extraction of material consists in the use of dictation software where speech is converted into text and inserted in a document. This activity ensures the accuracy and the richness of the notes taken during the interview and allows eliciting the most difficult data to collect such as comments to questions and the concept and relationship definitions (Bevan, 2014). When ready, the transcripts are sent to the co-researchers who have 24 hours, the allotted period, to return comments and corrections. The transcript is deemed accepted if no comment is received in the allotted period.

5.6.1.5 Content Analysis

The researcher extracts the sought meaning units: the agnostic concepts, their definitions and relationships. The low-abstract domain specific concepts and the subsumption relationships are also elicited. Spreadsheets are used to contain the meaning units in various forms, such as comparative series of scoring with questions Q06 and Q07, comparing the average and standard deviation of the numeric responses. The domain specific concepts are to be used in future use case reports that would comprise a competency question directed to a given industry or government sector.

A meaning unit, as defined by (Hycner, 1985), is «*crystallization and condensation of what the participant (co-researcher) has said, still using as much as possible the literal words of the participant. This is a step whereby the researcher still tries to stay very close to the literal data. The result is called a […] meaning [unit]*». In other words, it is the essence of what emerges from the transcripts, deliberately or coincidentally, and will be coalesced during the synthesis step.

The types of meaning units identified ex post facto in the present paper are:

- Years of experience of the co-researcher;
- The industry or government sectors that the co-researcher performed conceptualization. It is important to note that the industry sector terms that were provided by the co-researcher is usually converted into the North American Industry Classification System designation the closest to the one provided by the participant. This is one instance where the researcher opted not to comply with the definition of meaning units;
- The agnostic concepts;
- The subsumption and other relationships between the agnostic concepts;
- The definition or description of the agnostic concepts; and
- The de facto agnostic CODPs derived for the above-mentioned meaning units obtained by executing the synthesis step.

Meaning units 1 and 2 represents the contextualization meaning unit that provides the needed backdrop to enhance the phenomenological insight elicitation. Meaning units 3 through 6 constitute the phenomenological meaning units that are at the heart of this research. (Simsion et al., 2012) indicated that participants of most similar studies were students. In the case of this paper's research, the average co-researcher experience in conceptualization is 21.19 years, more than double the threshold defined by (S. Ahmed et al., 2005) for a professional to be considered an "expert".

5.6.1.6 Content synthesis

The emerged meaning units from the transcripts are coalesced into integrated structures. The following rules listed in Table 5.2 and established by this paper's project are applied to produce the intended results for each type of meaning units:

Table 5.2 Meaning unit coalescence rules

Meaning unit number	**Meaning unit type description**	**Meaning unit coalescence rule description**
1	Years of experience of the co-researcher.	Basic aggregating statistical functions such as average and standard deviation.
2	The industry or government sectors that the co-researcher performed conceptualization.	Basic aggregating statistical functions such as average and standard deviation.
3	The agnostic concepts.	Concepts defined in the same manner are retained if it was identified by at least two co-researchers; In the case of synonyms, only the term with the greatest selection by co-researchers is retained. In case of equal number of selections, the researcher makes the final decision; In the case of concepts that have been defined in more than one way, the same rule as in the case of synonyms applies.
4	The subsumption and other relationships between the agnostic concepts.	The relationships need to be selected only once to be retained. In case of conflicting relationships, only the one with the greatest number of selections is retained.
5	The definition or description of the agnostic concepts.	The texts are integrated by the researcher.

Table 5.2 Meaning unit coalescence rules (continued)

Meaning unit number	**Meaning unit type description**	**Meaning unit coalescence rule description**
6	The de facto agnostic CODPs derived for the above-mentioned meaning units.	The aforementioned meaning units are then integrated in distinct modules using the SLR's module structure as a starting point. The researcher may decide to diverge from the SLR's architecture on a case-by-case basis. The researcher, for example, may opt to rename and redefine the Contract module to Agreement if the phenomenology research reverses the subsumption relationship between Contract and Agreement. The researcher the names the module Agreement.

It is noteworthy to mention that some these rules may allow consistent reproducible outcomes should the data be provided to different researchers for at least some of the questions, which could contradict (Okoli, 2015) position on the irreproducibility of the synthesis phase in the context of the SLR research method. Albeit the fact that they are different methods, the SLR and phenomenological research methods used in this project are qualitative research methods. Qualitative research methods such as phenomenological, grounded theory and discourse analysis share the analysis step's decontextualizing of collected data and also the re-contextualizing of data performed in the synthesis step (Starks & Brown Trinidad, 2007). The aspect of reproducibility represents a requirement for investigation in an upcoming project.

The last step of the research protocol consists in producing a draft report of the phenomenological study, the individual summary transcripts and to transmit them to the co-researchers for comments for the draft report and their approval of the interview transcripts. The co-researchers have 24 hours, the allotted period, to return comments and corrections. The transcript is deemed accepted if no comment is received in the allotted period.

5.7 Research question

As shown earlier in section 5.3.4 and indicated in this project's SLR (Fitzpatrick, Ratté, et al., 2018a), Guarino in (N. Guarino, 1998), stresses that ontologies only approximate a conceptualization. He also indicates that the only way to enhance the representation is to develop a richer set of axioms, which are derived from concepts. As Guarino stipulated that conceptualization is language-independent, it can be argued here that the elicitation of richer concepts as ontology design patterns, and their conversion into axiomatic rules or axioms as proposed by (Blomqvist, 2009b), would enhance the use of inference engine technologies described notably by (McGuinness & Da Silva, 2004). Data integration, also referred to as semantic data integration by (De Giacomo et al., 2018), represents a potentially effective application for ontology-based inference technologies. As proposed by (Daniel Fitzpatrick et al., 2013), a multi-domain ontology would leverage agnostic design patterns, based on semi-formal ontologies, to perform data integration and resolve the semantic heterogeneity problem.

For this phenomenological study, the research question is formulated by the following: «w*hat are the conceptualization patterns found in semi-formal ontologies, e.g. data model patterns, software engineering patterns, etc, that can be agnostic to any domain or industry sector in the context of enterprise semantic interoperability and can be used as the basis of agnostic CODPs to resolve semantic heterogeneity in enterprise systems*?»

This research question constitutes the basis for the design of the semi-structured interview questionnaire. Following the interview, the transcripts provide the elements of the system of beliefs, the meaning units, for each co-researcher.

5.8 Content analysis

The content analysis step encompasses three distinct knowledge components, i.e. knowledge being actionable information as defined in (Fitzpatrick, 2012):

- The contextual knowledge: responses to questions Q01 and Q02 in respect to the number of years of experience of the co-researcher;
- The phenomenon knowledge: the essential set of questions for this study, Q03 through Q05, which aims to elicit the sought concepts to respond to the research question;
- The peripheral knowledge: questions Q06 through Q12 that provide more context and material to prepare for the subsequent phases of this project, notably on determining metrics pertaining specifically to data integration.

5.8.1 Contextual knowledge

The first question Q01 is formulated as “How many years have you performed conceptualization, e.g. data models, canonical models, domain model, XSD, etc?” Figure 5.5 shows the distribution of the number of years’ experience per 5-year range group of the twenty-two co-researchers that participated in the phenomenological study. Additional statistics are provided in section 5.9 Content synthesis. The minimum number of years of experience is eight years in compliance with the purposeful sampling criterion as explained in section 5.6.1.1.

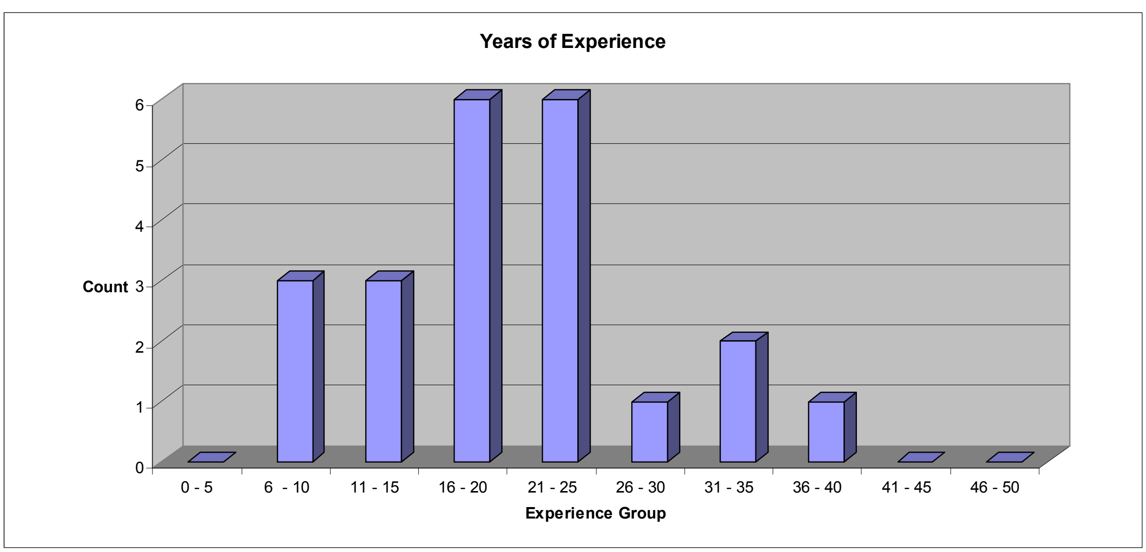


Figure 5.5 Distribution of the co-researchers’ years of experience

The second question Q02 is formulated as "What are the industry and government sectors have you performed conceptualization?". The industry sector terms that were provided by the co-researcher is usually converted into the North American Industry Classification System (NAICS) designation the closest to the one provided by the participant (President, 2017). This is one instance where the researcher opted not to comply with the concept of meaning unit by not using the direct input from the co-researcher.

Figure 5.6 outlines the number of co-researchers for each NAICS category. The twenty-two co-researchers identified a total of 138 industry sectors in which they performed conceptualization. The banking and credit union sector receive the highest number of selections, followed by retail trade, insurance and securities & commodities (Investment). During the interview, names of actual organizations were provided to help determine the industry sector but were not noted in the transcripts. Furthermore, some participants were involved in more than one sector while working for an enterprise. In such cases, usually very large enterprises, the participants work in an IT function, e.g. data architecture, which provides services to several divisions encompassing more than one industry sector. The researcher ensures that the proper industry sectors are assigned for these cases considering the nature of the projects the co-researchers were involved.

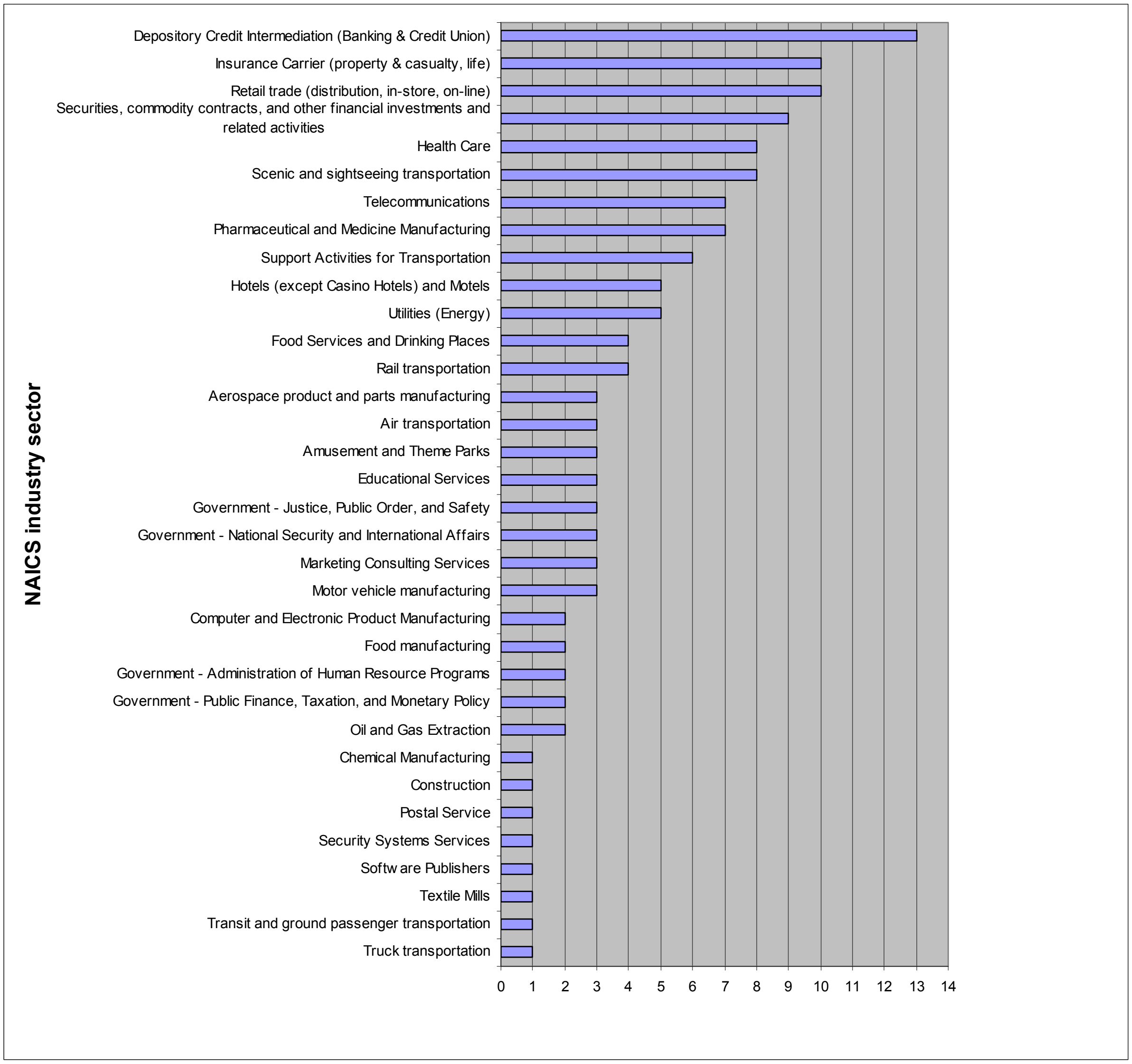


Figure 5.6 Distribution of co-researchers per NAICS industry sectors

The participating twenty-two co-researchers having experienced 138 industry sectors have cumulated a great deal of experience. This experience also covers a wide variety of private industry and government sectors. Some examples of co-researchers' diverse career paths are:

CR01: depository credit Intermediation (banking & credit union), motor vehicle manufacturing, telecommunications, pharmaceutical and medicine manufacturing, health care, insurance carrier, rail transportation; and

CR07: Aerospace products and parts manufacturing, government - national security and international affairs, amusement and theme parks, utilities (energy), depository credit intermediation (banking & credit union), retail trade (distribution, in-store, on-line), support activities for transportation, oil and gas extraction, marketing consulting services, pharmaceutical and medicine manufacturing.

5.8.2 Phenomenon knowledge

The phenomenon knowledge questions mean to elicit the agnostic CODPs for designing the multi-domain ontology. The third question Q03 is formulated as "Name and describe abstract (agnostic) concepts that you believe may apply to any industry and government sector". The co-researchers identified a total of 171 agnostic concepts. Table 5.3 outlines the twenty agnostic concepts that were the most identified by co-researchers, the top twenty selections, and the number of co-researchers that identified them.

Table 5.3 Top twenty agnostic concepts

Name of the top twenty agnostic concepts	Number of the top twenty selections
Party	19
Product	19
Service	19
Good	17
Event	16
Organization	15
Location	14
Person	13
Transaction	13
Account	12
Address	9
Bill-Of-Material	9
Building	9
Contract	9
Customer	8
Email address	8
Party role	8
Telephone	8
Agreement	7
Price	7

These concepts are part of broad domains such as party (party, party role, person, organization) that may represent any concept pertaining to people, group of persons, companies, enterprise, government agency, virtual enterprise, customer, supplier, employee, etc. The party related data model patterns have been popularized by (Hay, 1996) and (Silverston & Agnew, 2011).

The fourth question Q04, "Indicate relationships between these abstract concepts." is answered by the co-researchers, who provided these relationships most of the time while responding to the third question Q03. The co-researchers establish generalization-specialization relations and others. Table 5.4 lists examples of the responses provided by indicating the contributing co-researchers for each relationship example, the first entity, the relationship verb and the second entity of each relationship.

Table 5.4 List of examples of relationships provided by the co-researchers

List of contributing co-researchers	**First entity**	**Relationship verb**	**Second entity**
CR18	Party	Is a synonym	Thing
CR02, CR03, CR04, CR05, CR06, CR07, CR08, CR09, CR10, CR11, CR12, CR13, CR14, CR16, CR17, CR19, CR20, CR21, CR22	Party	Is-a	Thing
CR10, CR18, CR20	Role	Is-a	Thing
CR18	Thing	Can play a	Role
CR01, CR02, CR03, CR04, CR06, CR07, CR08, CR09, CR11, CR12, CR14, CR15, CR16, CR19, CR20, CR21, CR22	Good	Is-a	Product
CR05, CR10, CR13	Good	Is-synonym	Product

Table 5.4 List of examples of relationships provided by the co-researchers (continued)

List of contributing co-researchers	First entity	Relationship verb	Second entity
CR18	Party	Is a synonym	Thing
CR02, CR03, CR04, CR05, CR06, CR07, CR08, CR09, CR10, CR11, CR12, CR13, CR14, CR16, CR17, CR19, CR20, CR21, CR22	Party	Is-a	Thing
CR10, CR18, CR20	Role	Is-a	Thing
CR18	Thing	Can play a	Role
CR01, CR02, CR03, CR04, CR06, CR07, CR08, CR09, CR11, CR12, CR14, CR15, CR16, CR19, CR20, CR21, CR22	Good	Is-a	Product
CR05, CR10, CR13	Good	Is-synonym	Product
CR02, CR03, CR04, CR06, CR07, CR08, CR09, CR11, CR12, CR14, CR15, CR16, CR19, CR20, CR21, CR22	Service	Is-a	Product
CR03, CR05, CR06, CR08, CR10, CR11, CR13, CR14, CR15, CR18, CR21, CR22	Agreement	Is-a	Thing
CR03, CR06, CR08, CR10, CR14, CR18	Contract	Is-a	Agreement
CR02, CR04, CR05, CR07, CR09, CR12, CR17, CR19	Contract	Is-a	Thing

Some of these relationships, for example “Good is-a Product” and “Good is synonym Product, are in conflict and only one is retained based on the most stated. During the synthesis step, these cases are handled by the treatment defined in meaning unit 4 in Table 5.2.

The fifth question Q05, "For a maximum of three industry or government sectors, list domain specific (low abstract) concepts and identify to which abstract concept they relate to (generalization-specialization only)" generated a significant number of concepts. While providing, as requested, generalization-specialization relationships to the agnostic concepts, this allowed eliciting additional concepts to the ones already identified in question Q03. Table 5.5 enumerates examples of the responses provided by indicating the contributing co-researchers, the industry domain, domain specific concepts and relationships with agnostic concepts.

Table 5.5 List of examples of domain specific concepts with subsumed relationships with agnostic concepts

Contributing co-researchers	**Industry sector**	**Domain specific concepts**	**Subsumed relationships with agnostic concepts**
CR01, CR04, CR11, CR18	Manufacturing	Armored vehicle, tank, helicopter, fabrication plant, circuit, assembly, engineering Bill-Of-Material, manufacturing Bill-Of-Material, windshield	Armored vehicle is-a equipment, tank is-a equipment, helicopter is-a equipment, fabrication plant is-a location, circuit is-a good, assembly is-a good, engineering Bill-Of-Material is-a BOM, manufacturing Bill-Of-Material is-a BOM, windshield is-a good
CR05, CR08, CR10, CR13	Depository Credit Intermediation (Banking & Credit Union)	Interest rate, loan contract, mortgage, evaluator, transfer of funds, borrower, bank fee, banking service, branch	Interest rate is-a price, loan contract is-a contract, evaluator is-a party role, borrower is-a party role, banking service is-a service, branch is-a building

Table 5.5 List of examples of domain specific concepts with subsumed relationships with agnostic concepts (continued)

Contributing co-researchers	Industry sector	Domain specific concepts	Subsumed relationships with agnostic concepts
CR01, CR15, CR19	Pharmaceutical and Medicine Manufacturing	Drug, biological drug, chemical drug, disease, Food and Drug Administration, disease, prescription	Drug is-a good, biological drug is-a good, chemical drug is-a good, Food and Drug Administration is-a Organization, disease is-a process, prescription is-a request

This enumeration illustrates the capacity of the agnostic concepts to subsume several domain specific concepts for each industry sector. Considering that this present paper elicits thirty-four industry sectors, agnostic patterns may each apply to numerous domain specific concepts. This data can be used for industry sector specific use cases to demonstrate the transferability, a trustworthiness criterion, of the phenomenological research method.

5.8.3 Peripheral knowledge

The questions related to peripheral knowledge intend to induct more context and material to prepare for the subsequent phases of this project. With these questions, the researcher is casting a wider net to collect data for the continuation of the theory-building process. Questions Q06 and Q07 initiate the exploration into the prescriptive aspects of the design and development of the multi-domain ontology. Question Q10 inquires on the delicate subject of users' influence on the design of a data integration platform. An SLR (Bano & Zowghi, 2013) and a case study (Zowghi, da Rimini, & Bano, 2015) authored by Bano et al. concluded that users' involvement in system development tends to be positive. This present paper depicts a very different picture for the design of a data integration platform. As covered in this analysis step and the following synthesis step, the line of business influence tends to be perceived negatively by the majority of the co-researchers. Questions Q11 and Q12 investigate the notions of accuracy and quality of a data model with the intention of eventually proposing accuracy and quality related metrics.

Question Q06 is formulated as "Do you believe that a data integration function should be designed using abstract (agnostic) concepts as you indicated in question 3? Provide a score from 1 to 10. Please comment". This question is intended to inquire about co-researcher assessment on the importance of agnostic concepts, applicable to any industry or government sector, in the design of a data integration function. This question also refers to question Q03 in which the co-researcher outlines the agnostic concepts, their description or definition and their relationships. Figure 5.7 illustrates the skewed graph indicating a very strong positive response to the use of agnostic concepts in design a data integration function.

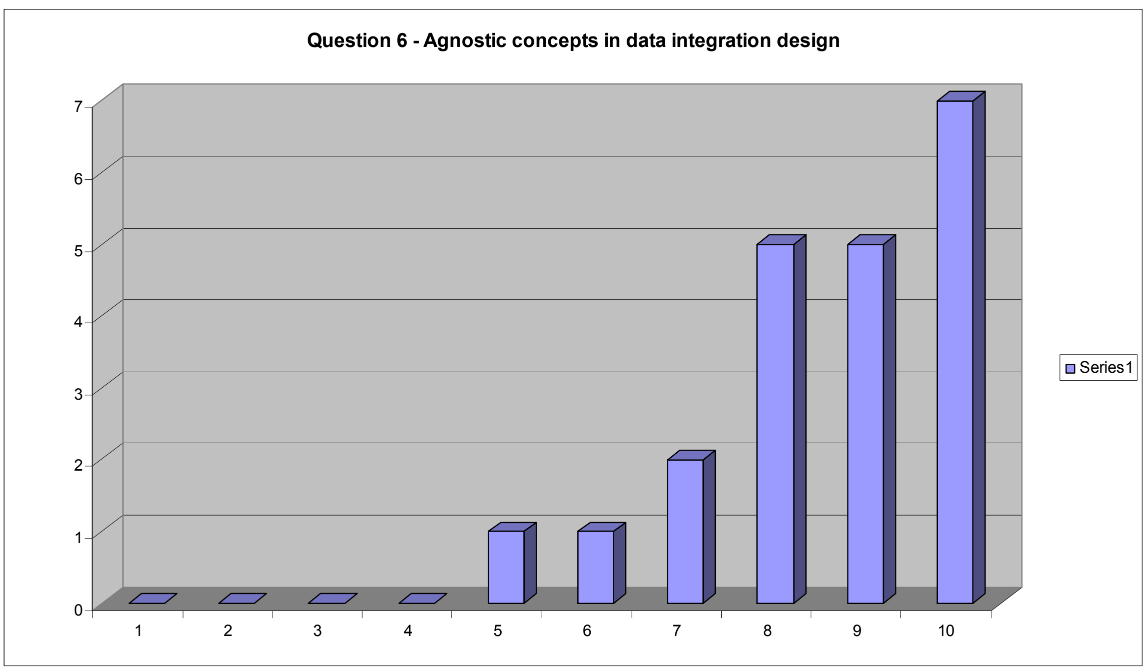


Figure 5.7 Use of agnostic concepts to the design of a data integration function

To question Q06, co-researcher CR01 responds: "Absolutely, it is a concept that I am attempting to drive". Co-researcher CR02 indicates: "score 6 for a stable environment company, score 8 for an organization that goes through a lot of changes, e.g. mergers". Co-researcher CR04 responds: "I score 10, you need the generic concepts for efficient data integration". On the other hand, co-researcher CR06 indicates: "yes but I score 7, we start with agnostic concepts but we rapidly get to details so we need the domain-specific concepts. (We) spend at least 50% of our time doing data integration. I see a lot of heterogeneous systems, with synonyms and different semantics".

Question Q07 is formulated as "Do you believe that a data integration function should be designed using low abstract (domain specific) concepts that would be understandable by business users? Provide a score from 1 to 10. Please comment". Question Q07 seems, for the co-researchers, to counterweigh against question Q06. The researcher needed to clarify, using the imaginative variation technique, which the two questions should be taken separately. This question is intended to inquire about co-researcher assessment on the importance of domain-specific concepts, applicable here to a specific domain of the industry, in the design of a data integration function. Figure 5.8 appears to be showing less decisiveness than the previous question.

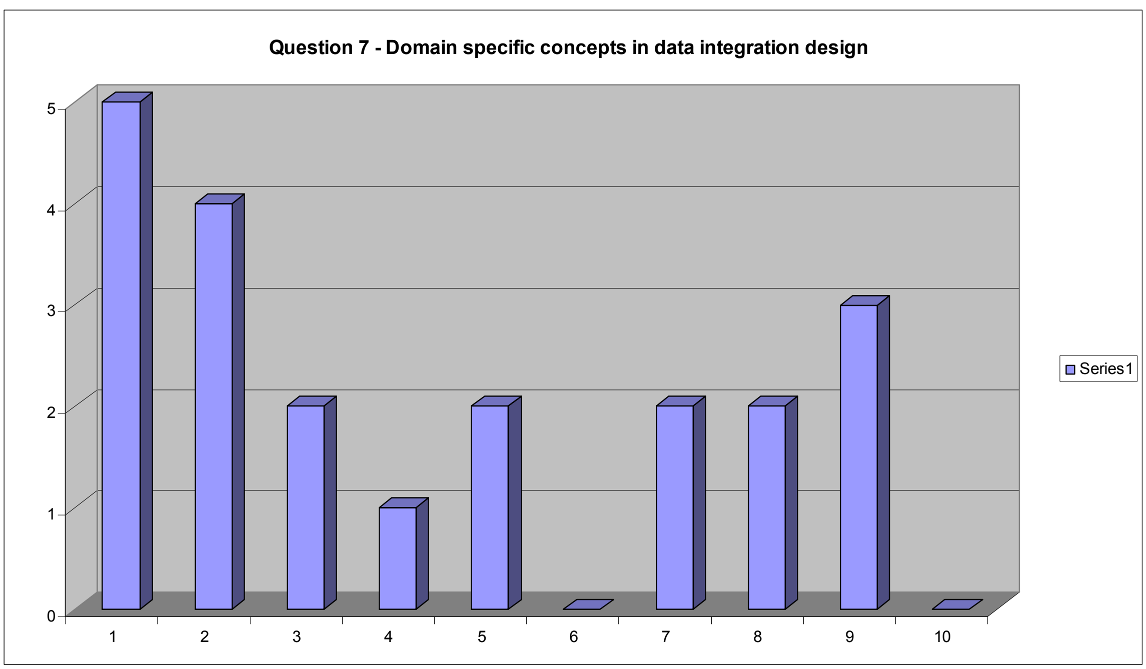


Figure 5.8 Use of domain-specific concepts to the design of a data integration function

Two sub-groups appear to be emerging from this graph. The first group of co-researchers that see less needs (scores 1,2 and 3) for domain-specific concepts. Some of the co-researchers, CR03 and CR04 indicate: "… score 1 for organizations with a lot of changes because we would have too much schema changes with "low-abstract", domain specific concepts are necessary for users. The integration function doesn't require the low-abstract concepts, only the layer through which the users access the data does". CR17 adds: "No single line of business should prejudice (by having domain specific concepts) our business (capacity to interoperate)".

The second group, on the other hand, responds in the case of co-researcher CR05, CR06 and CR22 respectively.

"I score 7 (ideally) in certain cases, low abstract concepts are required to complete the design of a data integration platform":

- "It is a 9, yes and even more, in designing you reach the detail attribute level, low-abstract concepts must be used"; and
- "Yes and I score 10. The IT specialist must not impose a vocabulary. The risk here is to disenchant the business in being involved, thus loss of financing".

Question Q08 explored the sentiment of the co-researchers in relation to scientific research's potential contribution to solve the semantic heterogeneity problem. Although the response was mostly positive, the nature of this potential contribution appeared to be ambiguous. Others, such as co-researcher CR13 considered that scientific research "would only help in performing (hypothetico-deductive) studies". Co-researcher CR06 replies: "I don't know how scientific research can help". In the case of question Q09 that asks the co-researcher if he or she has performed data integration function design, the response is unanimously positive.

Question Q10, formulated as “Did you ever observe line of business influence on the design of a data integration platform? Please comment” provoked in most cases a negative response on the line of business, or users, having an influence on the design of a data integration platform. Although co-researcher CR08 indicates:” I think it was a positive influence. The (line of) business brings clarity”. Co-researcher CR12 hypothesizes that: ”It could be positive if the they (lines of business) are supporting, not designing. The doctor metaphor applies here (the doctor not the patient decides how to perform the procedure). Roles must be clear”. Table 5.6 summarizes the negative response from the majority of co-researchers.

Table 5.6 Negative responses from co-researchers to question Q10

Co-researcher	**Reponses to question Q10**
CR01	“while there may had been valid reasons, there was an undue bias applied to the data integration platform because of the line of the business, looking for an easy solution… this ended up not working, costing the company a lot of money”.
CR05	“yes and I think that a line of business may adversely affect the design of the data integration function by pushing their own agenda, their own terminology, against the need to have reusable constructs”.
CR06	“yes increasingly… Needs are more and more expressed as technical specifications in the form of a prototype, instead of business requirements. Nowadays, everybody wants to design!”.
CR07	”Yes I have seen the business influencing the design of the data integration platform and it had an adverse effect. The platform’s design should not be based on or influenced by one specific business domain”.
CR09	”Yes, they (the business) are only interested in their data, they do not care for the other things (other areas of the enterprise)”.
CR10	”Yes I have seen such influence and it is not good. The (line of) business wants less abstraction, the model must tell them something, show their concepts (more clearly)”.
CR14	”yes and it is sometimes very negative. It (the influence of the line of business) can sometimes be very negative, affecting the reusability (of the data integration model) by introducing too many specializations, increasing time and effort on changing the model”.

Table 5.6 Negative responses from co-researchers to question Q10 (continued)

Co-researcher	Reponses to question Q10
CR16	”Yes I did observe the (line of) business influence the design of a data integration platform. And the (influence) was negative, bending best practices to suite (specific) business needs. There would be no more best practices”.
CR17	”yes and, overall, the influence was negative. It creates confusion and delay”.
CR18	”Yes and the influence is mostly negative. The (line of business) that shouts the loudest is the biggest payer (funder) or acts the fastest dictates (the design of the data integration platform)”.
CR19	”Yes and the influence was negative. They (the lines of business) negatively affect the agnostic (reusability) quality of the data integration platform”.
CR20	“yes and the influence was negative. They (lines of business) can derail the design”.
CR22	”(The line of business) should not normally be involved in the decision-making in respect to the design or architecture (of the data integration platform). It can be positive if the role of the (line of) business is to review the solution”.

Question 10 has elicited a great amount of insight in the matter of user involvement in the design of a data integration platform. The researcher considers this matter as an opportunity for further dedicated investigation. Several qualitative research techniques can be used such as semi-structured interviews, surveys and focus groups that would concentrate specifically on user involvement in the design and development of a data integration platform.

Questions Q11 and Q12 intended to elicit from the co-researchers their insight on how to define and measure the efficiency and quality of a data integration model. These questions, and their responses, did not bring the convergence the researcher was seeking. For question Q11, some co-researchers, such as for CR04, CR05, CR06, CR08 and CR14 respectively posit:

- “Efficiency is based on the amount of time to implement your first project, and then the time it takes to implement subsequent phases or modifications, which should, in proportion, progressively reduces over time. In other words, the time to deliver a solution diminishes";

- "The speed at which the organization can respond to change. We could measure the efficiency of the data integration model by considering the impact of amount of work performed on the data model, transformation and load processes, consumer and provider applications. Progressive reduction in time and effort spent on data integration. If there was never any changes in the organization, having low or high abstract concepts in the data integration model would not matter";
- "to be efficient, a data integration model would need as little attributes as possible. A faulty design of a data integration model would have a lot of redundancy. Perhaps, we could, ideally, have standard number of attributes, say 1000, which would tell us how efficient our data integration model is. Reusability is critical, the data integration model must be agnostic";
- "A data integration model must be flexible in the sense that is generic, reusable and allows rapid delivery. It must also be easy to understand"; and
- " (Efficiency is in essence) reusability. It (the data integration model) can be easily extended (to accommodate new requirements). It progressively requires less and less effort to be changed and maintained. The percentage of new concepts and properties (in the data integration model) diminishes over time".

Co-researchers in some cases equate quality to efficiency, such as for co-researchers CR03, CR05, CR07, CR10 and CR17. For co-researcher CR09 indicate that: "quality for a data integration model is more abstract, fewer moving parts". Additionally, co-researcher CR05 proposes the notion of "data-driven" with the data integration model comprising semantic ("parameters") that would allow process control, a much higher state of efficiency, instead of being "code driven". Co-researchers CR04, CR06, CR14 and CR20 stipulate that the quality of a data integration model is mostly about a good documentation, about the rigor in defining the objects.

The content analysis step's purpose is to break down the sought material for each interview. The analysis step encompasses three distinct knowledge components i.e. contextual knowledge, providing background on the co-researchers, phenomenon knowledge, related to the concepts at the center of this study and peripheral knowledge, harvesting material for future phases of the project. In the next step, the meaning units collected during the interviews, the agnostic concepts and their description and relationships, are coalesced into agnostic CODPs.

5.9 Content synthesis

In the previous analysis section, the content of the interviews is reduced to the intended material that will become meaning units. This decontextualization process on collected data leads into the recontextualization process of this research data performed in the synthesis step (Starks & Brown Trinidad, 2007). As detailed in section 5.6.1.6 in Table 5.2, the meaning unit coalescence rules establish the process from which emerge the meaning units. Through this process, the aggregated elements coalesce into ontology patterns, a «set of ontological elements, structures or construction principles that intend to solve a specific engineering problem» (Blomqvist, 2010).

Also, the notion of theoretical saturation is also examined in this section. Theoretical saturation originates from the ground theory research domain, and although it is not used as a standard in phenomenological research, this concept may shed some light on determining a relative level of maturity on the emerging theory on agnostic CODP for a multi-domain ontology. Theoretical saturation is used in the grounded theory and narratology qualitative research methods (Hays & Wood, 2011), (Stol, Ralph, & Fitzgerald, 2016). The authors in (Stol et al., 2016) define theoretical saturation as «…the point at which a theory's components are well supported and new data is no longer triggering revisions or reinterpretations of the theory».

Theoretical saturation in the context of this research is defined at which interview an agnostic concept is selected twice therefore included in the multi-domain ontology as an agnostic CODP.

In summary, the following are the rules used to synthesize the interview material into meaning units:

- Years of experience of the co-researcher: Basic aggregating statistical functions;
- Theoretical saturation: Basic aggregating statistical functions;
- The agnostic concepts: Retained concepts are selected at least twice by co-researchers. In case of conflicting or diverging definitions for the same concept, the greatest number is retained;
- The subsumption and other relationships between the agnostic concepts: only one instance expressed by a co-researcher is required. In case of conflict, the relationship with the greatest number of instances is retained;
- The definition or description of the agnostic concepts: The texts are integrated by the researcher.

The de facto agnostic CODPs derived for the above-mentioned meaning units: The selected meaning units are assembled to form the agnostic CODP and represented using the (UML light) Archimate modeling notation (Lankhorst et al., 2009).

Table 5.7 summarizes the statistics about some of the examined meaning units. Although the authors consider that this research is still in its infancy and no hypothetico-deductive techniques are considered at this point, the statistics may contribute to defining the next phases of the projects.

Table 5.7 Basic aggregating statistics about the meaning units

Name of the meaning unit	**Number of samples**	**Average**	**Standard deviation**	**Variance**	**Median**
Number of years of experience in conceptualization	22	21.1	8.1	65.6	20
Score for question Q06 about agnostic concepts in data integration function design	22	8.6	1.4	2.0	9
Score for question Q07 about domain specific concepts in data integration function design	22	4.5	3.1	9.9	3
Theoretical saturation point for agnostic concepts	83	10.6	6.4	41.4	10
Number of agnostic concepts identified by co-researchers	22	24	5.3	28.5	24

Table 5.7 statistics stand out mainly for the score for question Q06 about agnostic concepts in data integration function design. The narrow standard deviation notably suggests a high level of consensus amongst the co-researchers regarding the importance of agnostic concepts in the design of a data integration platform and most importantly in their presence as design patterns in a data integration model.

Figure 5.9 illustrates the progression of the theoretical saturation events longitudinally from the first to the last interview. Since a minimum of two selections are needed for an agnostic concept to be retained, no saturation event is recorded on the first interview.

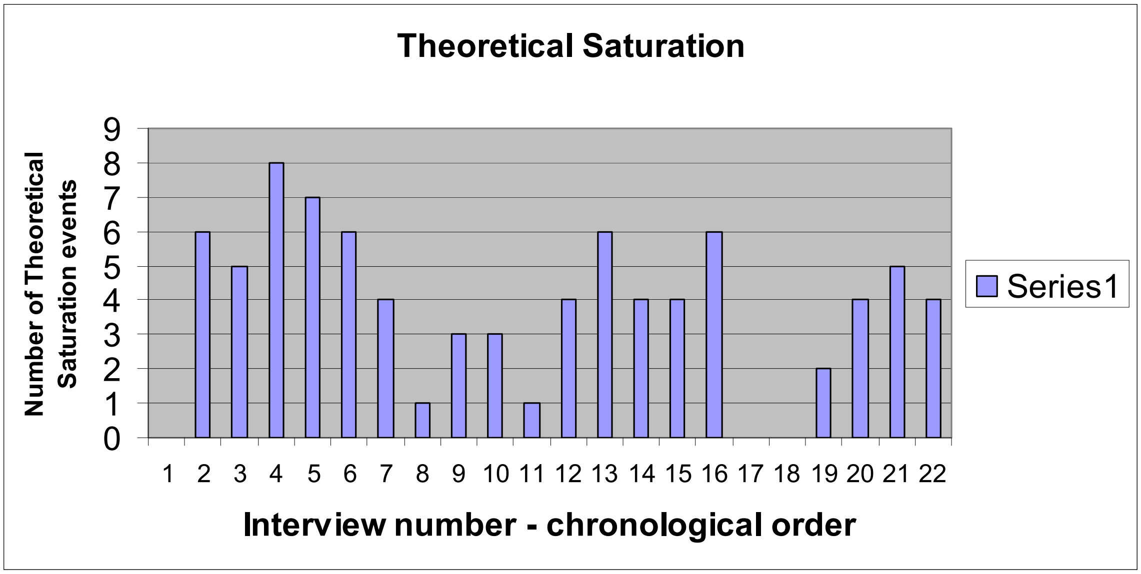


Figure 5.9 Progression of the theoretical saturation events

Albeit the diminishing trend in the graph of figure 5.9, the researcher at this point cannot conclude of any apparent behavior, notably of the sinusoidal curve and the presence of what appears to be three "waves". The researcher has not changed the questionnaire between the second and the 21st interview. Potential participants became co-researchers and were interviewed in a random fashion. Other than imaginative variation, i.e. to provide more contexts to questions to stimulate the conversation, no apparent reason may explain this sinusoidal behavior. In their qualitative study, the authors in (Guest, Bunce, & Johnson, 2006) observe complete theoretical saturation in their research at twelve interviews but also express the difficulty to conclude and generalize. The researcher plans to pursue to recruitment of twenty-five additional co-researchers. The next phase will be conducted using the same approach as described in the current paper except no imaginative variation will be done, in the attempt of achieving theoretical saturation for agnostic concepts.

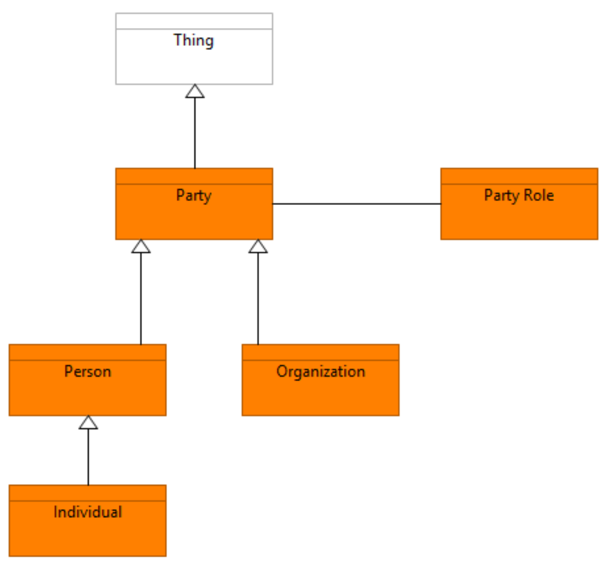
At this point of the content synthesis step, the agnostic concepts, their descriptions and relationships are synthesized from the material extracted from the co-researchers. This process parallels the synthesis process performed in the SLR, the companion paper to the present paper (Fitzpatrick, Ratté, et al., 2018a). The resulting meaning units, the agnostic

CODPs, are shown in Archimate notation diagrams, a lighter form of UML (Lankhorst et al., 2009). As in the SLR, each agnostic CODP is documented using a CODP template proposed in (Gangemi et al., 2007). The agnostic concept Thing, anything imaginary or real, is used in all diagrams. Each of the following modules are based on the ontology architecture pattern proposed in (Daniel Fitzpatrick et al., 2013), but adapted by the researcher when there is a change in the name of the module when the root entity is renamed. The root entity is the main agnostic concept that bears the same name as the module. In some cases, the definition from this project's SLR is used when the present approach has not produced a suitable definition.

5.9.1 The Party agnostic CODP

The Party CODP conceptualizes people and organizations.

Table 5.8 Phenomenological study Party CODP

Ontology Pattern Type	***Content Ontology Design Pattern***
Name	Party
General description	The Party CODP allows the conceptualization of the nature of a person and an organization.
Examples	Any physical person regardless of what role or roles may be played, e.g. John Doe. A private corporation, a job position, a government agency, a government as a whole, an informal group, a family.
Simplified UML diagram (Archimate)	
Definitions of the agnostic concepts	• Party: A thing that is either a person or an organization; • Party Class: A classification scheme for parties; • Person: A biological thing classified as a Homo Sapiens; • Organization: A group of persons; • Party Role: See the Role CODP.

5.9.2 The Product agnostic CODP

The Product CODP covers the goods and services that result from processes. It includes the notions of classification and Bill of Material.

Table 5.9 Phenomenological study Product CODP

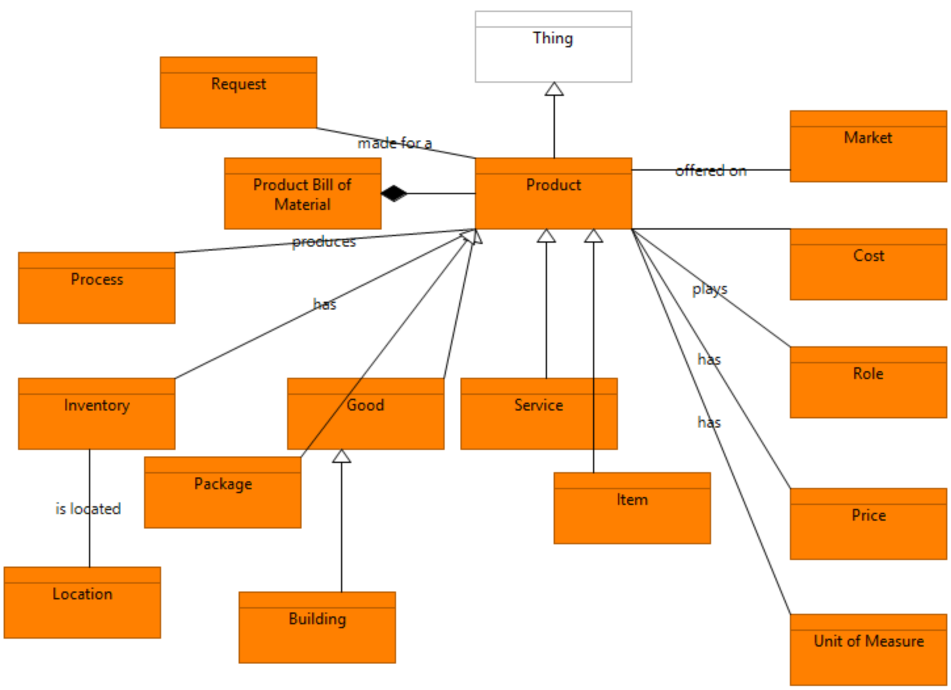
Ontology Pattern Type	***Content Ontology Design Pattern***
Name	Product
General description	A good or service resulting from a process. The UN PCS and NAPCS classification schemes can notably be used as taxonomies for products. The concept of bill of material allows to package products.
Examples	Goods are tangible products such as automobile, an electronic equipment, salt, fuel. Services are intangible services such as car rental, banking offerings, investment portfolio management.
Simplified UML diagram (Archimate)	

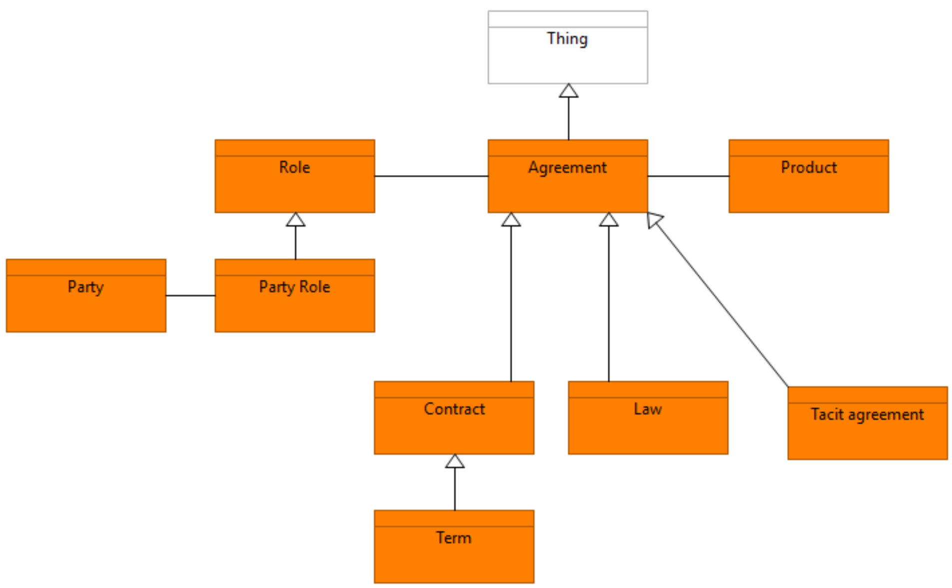
Table 5.9 Phenomenological study Product CODP (continued)

Ontology Pattern Type	***Content Ontology Design Pattern***
Definitions of the agnostic concepts	• Product: A tangible or an intangible thing offered commercially through a process. A product may comprise other products, items or parts, which are also products; • Order: A demand to obtain products; • Bill of Material: A grouping of products that is a product as well; • Inventory: A list of goods or services available at a location; • Good: A tangible thing such as a building; • Service: An intangible product offered to provide value to a customer; • Unit of Measure: A standard for establishing the quantity of a thing, e.g. Currency, weight, height, etc.; • Role: See the Role CODP; • Location: See the Location CODP; • Process: See the Process CODP; • Price: See the Price CODP.

5.9.3 The Agreement agnostic CODP

The Agreement CODP covers any form of tacit or explicit agreement between parties.

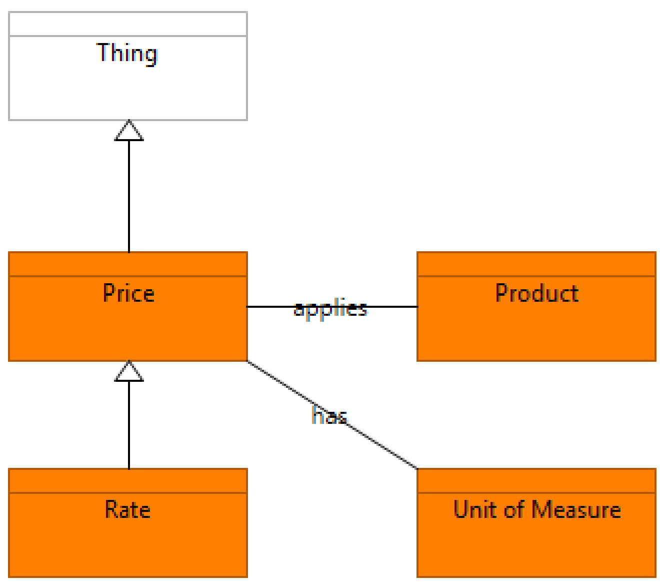
Table 5.10 Phenomenological study Agreement CODP

Ontology Pattern Type	***Content Ontology Design Pattern***
Name	Agreement
General description	The Agreement CODP allows the conceptualization of an arrangement between parties playing roles.
Examples	A legal binding contract for the sales of a house between two persons playing roles of buyer and seller. A Service Legal Agreement for procuring an infrastructure cloud service to a user from a cloud provider. The set of terms and conditions associated with a bank-checking service.
Simplified UML diagram (Archimate)	
Definitions of the agnostic concepts	• Agreement: An arrangement between parties playing roles within a context; • Contract: An explicit agreement between playing roles that is normally enforceable by a court of law in case of dispute; • Role: See the Role CODP; • Party: See the Party CODP.

5.9.4 The Price agnostic CODP

The Price CODP optionally relates to products and allows the commercial operations to generate revenues.

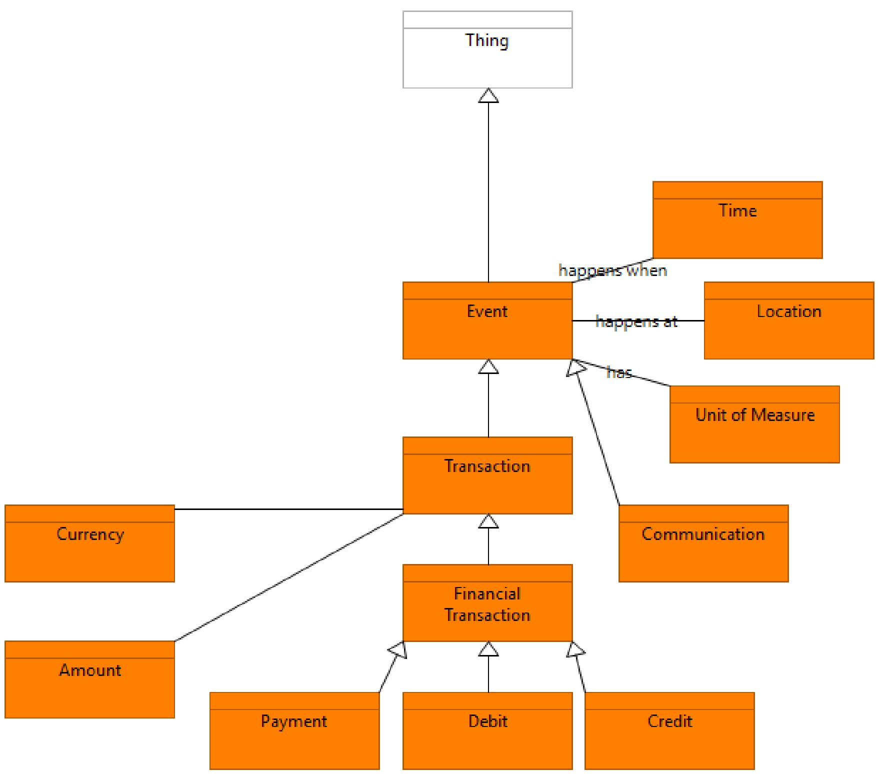
Table 5.11 Phenomenological study Price CODP

Ontology Pattern Type	***Content Ontology Design Pattern***
Name	Price
General description	The Price CODP allows the conceptualization of the notion of rates, rate packages, fees, penalties, pricing curve (time varying cost structure) applicable to the consumption of products.
Examples	A rack rate applicable for selling room nights in a hotel. A driver's licence fee for the right to drive a motor vehicle as a service dispensed by a government agency.
Simplified UML diagram (Archimate)	
Definitions of the agnostic concepts	• Price: A financial quantity associated to the selling of products; • Rate: A price measured in level of consumption; • Product: See the Role CODP.

5.9.5 The Event agnostic CODP

The Event CODP relates to occurrences in space and time that affects the state of things.

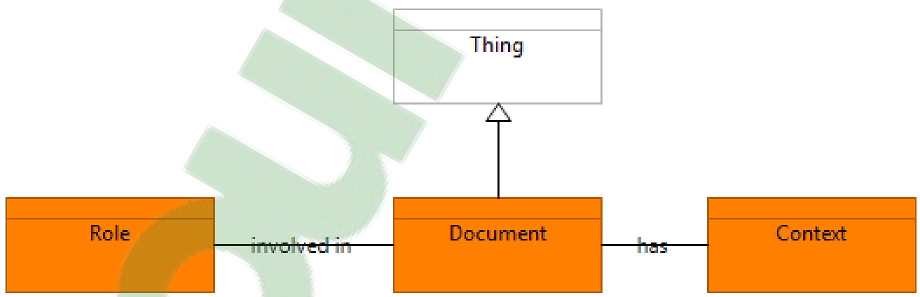
Table 5.12 Phenomenological study Event CODP

Ontology Pattern Type	***Content Ontology Design Pattern***
Name	Event
General description	The Event CODP allows the conceptualization of the notion of a spatiotemporal occurrence that may affect a thing by changing its state.
Examples	The start of a registration process for a student in a university. A financial transaction reducing a cash accounting account after the disbursement of a pay cheque.
Simplified UML diagram (Archimate)	
Definitions of the agnostic concepts	• Event: A spatio-temporal thing that affects another thing; • Transaction: An event where an exchange in money or commodity occurs; • Unit of Measure: A standard to measure a thing, e.g. Currency, weight, height, etc.; • Location: See the Location CODP.

5.9.6 The Document agnostic CODP

The Document CODP is a media containing symbolic facts that a person may bring context and acquire as knowledge and know-how.

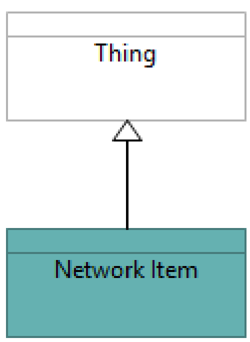
Table 5.13 Phenomenological study Document CODP

Ontology Pattern Type	***Content Ontology Design Pattern***
Name	Document
General description	The Document CODP allows the conceptualization of physical or electronic representation of a body of concepts in a context;
Examples	The Open Group Architecture Framework book purchased on the Open Group web site. This SLR will be published as a journal article.
Simplified UML diagram (Archimate)	
Definitions of the agnostic concepts	• Document: A physical or electronic media support that represents concepts; • Context: see the Context CODP.

5.9.7 The Network agnostic CODP

The Network CODP is the implementation of the Petri-network concept for conceptualization.

Table 5.14 Phenomenological study Network CODP

Ontology Pattern Type	***Content Ontology Design Pattern***
Name	Network
General description	The Network CODP allows the conceptualization of a Petri-like structure composed of two nodes and a segment linking the nodes for the purpose of transport of: energy, cargo, people, voice, data, etc. A grouping of networks is also a network.
Examples	A non-stop flight links Montreal, Canada to Chicago USA. A telecommunication channel links switching node A to switching node B.
Simplified UML diagram (Archimate)	
Definitions of the agnostic concepts	Network: A structure composed of two nodes and an edge that associates an origin and a destination for the purpose of transportation of: energy, cargo, people, voice, data, etc.

5.9.8 The Account agnostic CODP

The Account CODP is the only agnostic concept that possesses a dual nature, the Product Account, a mechanism to allow access to a product, and an Accounting Account that is used in financial recording and reporting.

Table 5.15 Phenomenological study Account CODP

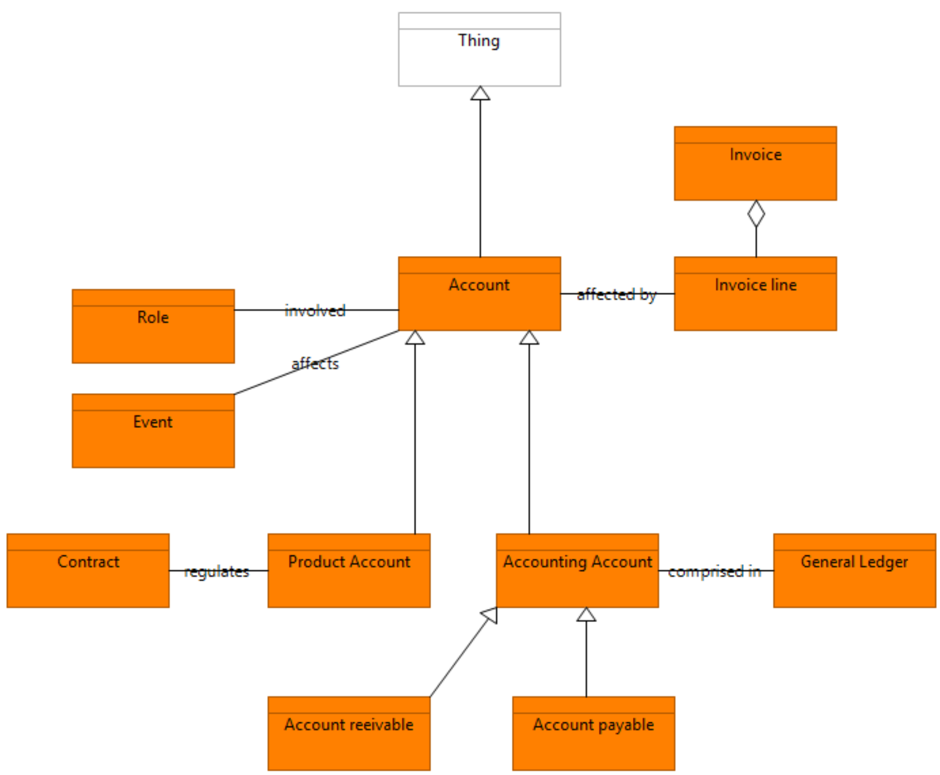
Ontology Pattern Type	***Content Ontology Design Pattern***
Name	Account
General description	The Account CODP allows the conceptualization of a thing used for recording transactions for the purpose of procuring products or tallying quantities for financial statements.
Examples	A checking account allows the customer to write cheques without fees when the balance is more than $1000 for the whole month. The Building – Asset account has been adjusted in the Consolidated Grand Ledger by a post-mortem transaction.
Simplified UML diagram (Archimate)	

Table 5.15 Phenomenological study Account CODP (continued)

Ontology Pattern Type	***Content Ontology Design Pattern***
Definitions of the agnostic concepts	• Account: A mechanism that aggregates transactions to offer products or to tally financial numbers; • Contract: See the Contract CODP; • Role: See the Role CODP; • Event: See the Event CODP.

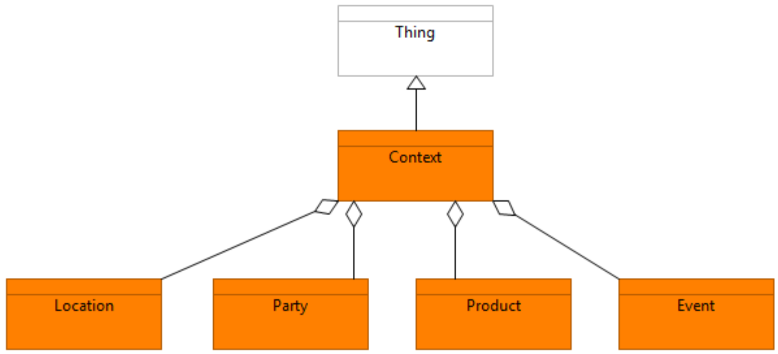
5.9.9 The Context agnostic CODP

As mentioned in the SLR, the companion paper to the present publication, the Context CODP is one of the least known of the data model patterns. This agnostic concept was confirmed in a theoretical saturation event at the 22nd interview. This pattern may be quite useful for several applications including NLP and other cognitive applications as discussed in (Daniel Fitzpatrick et al., 2013).

Table 5.16 Phenomenological study Context CODP

Ontology Pattern Type	***Content Ontology Design Pattern***
Name	Context
General description	The Context CODP allows the conceptualization of a set of things such as location, parties, products and events that grouped together may influence the use of vocabularies, chain of future events.
Examples	In the metaphor-rich American culture, an expression such as «passing the buck» may mean something quite different than when taken literately. In the context of ACME Corporation, deploying Service-Oriented Architecture (SOA) services just means implementing plain web services.

Table 5.16 Phenomenological study Context CODP (continued)

Ontology Pattern Type	***Content Ontology Design Pattern***
Simplified UML diagram (Archimate)	
Definitions of the agnostic concepts	• Context: A set of concepts that defines a situation; • Location: see the Location CODP; • Party: see the Party CODP; • Product: see the Product CODP; • Event: see the Event CODP.

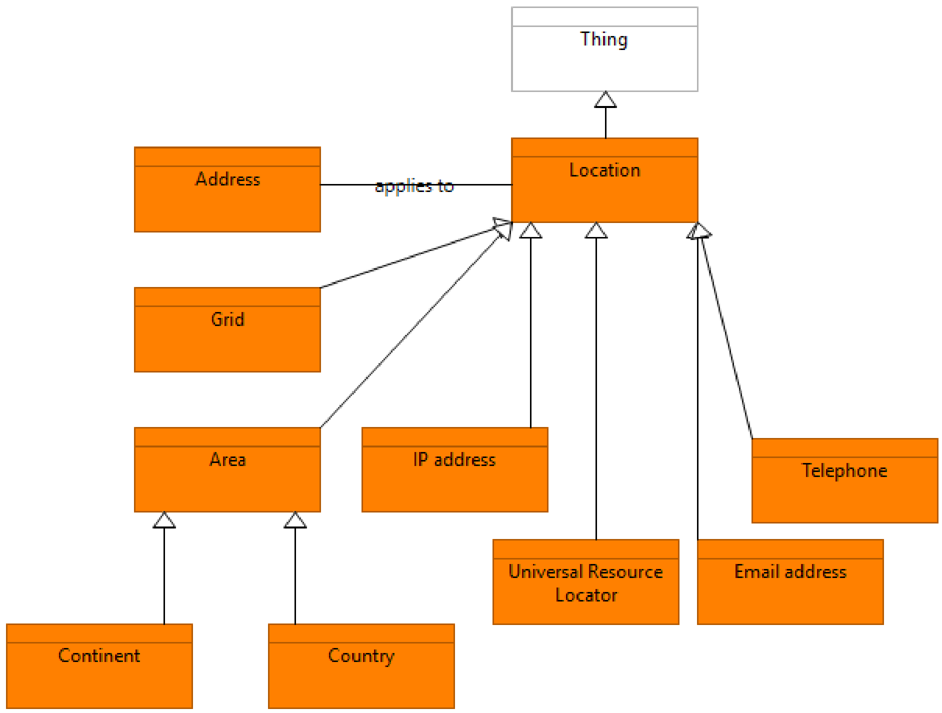
5.9.10 The Location agnostic CODP

The Location CODP covers geographical and other forms of coordinated systems.

Table 5.17 Phenomenological study Location CODP

Ontology Pattern Type	***Content Ontology Design Pattern***
Name	Location
General description	The Location CODP allows the conceptualization of a thing related to a coordinate system such as Earth location systems. This includes the notion of area, segment and grid locations. Geography also includes the notion of street addresses and electronic locations such as email and IP addresses.
Examples	The City of New York is a Location Area included in the State of New York. The address of this house is 123 Main Streer, Littletown USA and has a centroid determined by a longitude and latitude.

Table 5.17 Phenomenological study Location CODP (continued)

Ontology Pattern Type	***Content Ontology Design Pattern***
Simplified UML diagram (Archimate)	
Definitions of the agnostic concepts	• Location: An object in a coordinated system; • Location Grid: A zero-dimensioned point on a coordinate system; • Location Area: A closed surface location such as a country; • Address: A designation used as a contact mechanism.; • Electronic Location: A location used in an electronic realm.

5.9.11 The Role agnostic CODP

The Role CODP includes all types of behavior that are part of the intrinsic nature of a thing.

Table 5.18 Phenomenological study Role CODP

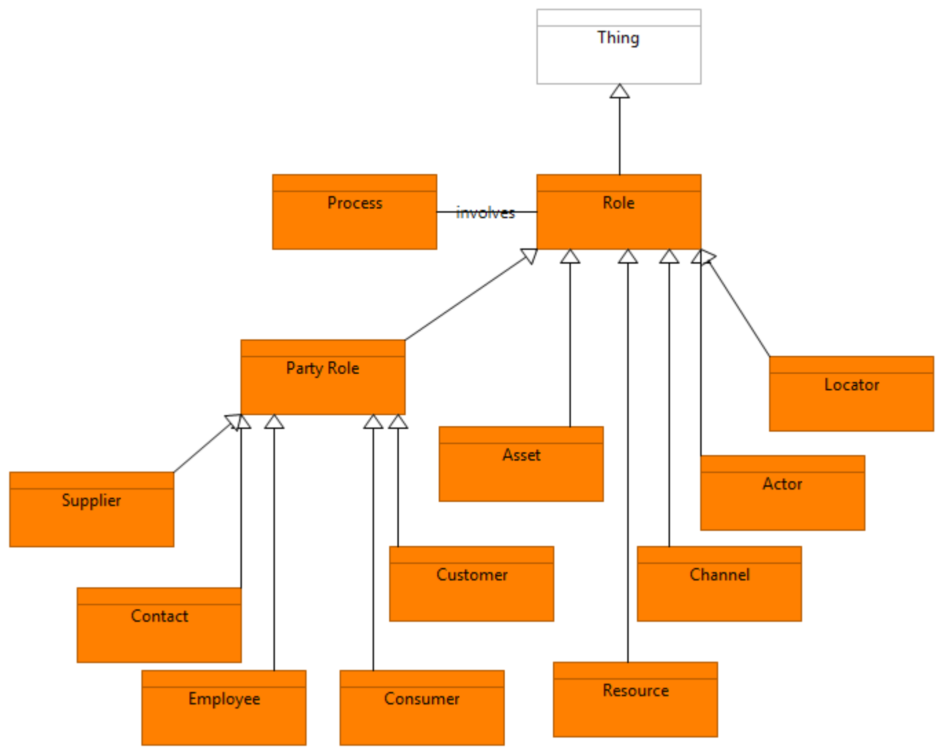
Ontology Pattern Type	***Content Ontology Design Pattern***
Name	Role
General description	The Role CODP allows the conceptualization of a form of involvement in a process or into anything other than a role. A thing playing a role would exhibit a behaviour that may not be related to its nature.
Examples	A person plays the role of an contact in ACME Corporation. This horse is an asset for this farmer and is a resource that is involved in farm processes.
Simplified UML diagram (Archimate)	

Table 5.18 Phenomenological study Role CODP (continued)

Ontology Pattern Type	***Content Ontology Design Pattern***
Definitions of the agnostic concepts	• Role: A form of relationship between things; • Identity: A Role being played by a Thing to uniquely designate another Thing; • Name: A form of Identity composed of one or more words; • Party Role: A form of Role played by a Party; • Vendor: A Party Role that involved supplying a Product; • Employee: A Party Role that involves being a full-time worker for an organization; • Customer: A Party Role that involves consuming a Product from a vendor; • Asset: A Role being played by a Thing that involves having a value for another Thing; • Resource: A Role being played by a Thing that involves participating in a Process; • Channel: A Role being played by a Thing for allowing access to another Thing; • Process: see the Process CODP.

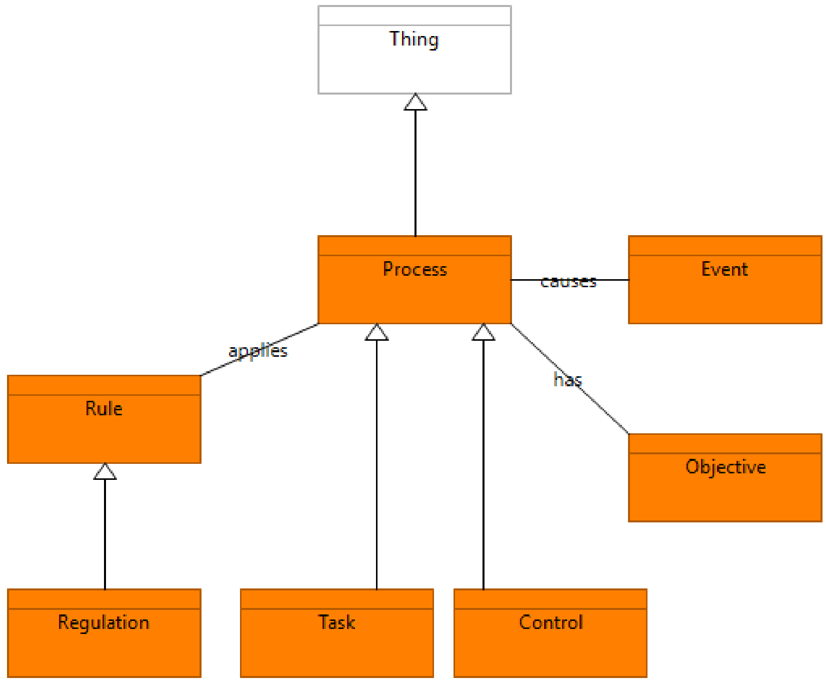
5.9.12 The Process agnostic CODP

The Process CODP covers all forms of human or natural activities.

Table 5.19 Phenomenological study Process CODP

Ontology Pattern Type	***Content Ontology Design Pattern***
Name	Process
General description	The Process CODP allows the conceptualization of a form of a unit of work in which resources are used in the fabrication of goods or in the rendering of services. A process can be performed by humans, by nature or a mix of both.
Examples	A set of activities in the manufacturing of a consumer electronic product is a Process. The growth of an animal’s fetus in an In Vitro facility is a Process.

Table 5.19 Phenomenological study Process CODP (continued)

Ontology Pattern Type	***Content Ontology Design Pattern***
Simplified UML diagram (Archimate)	
Definitions of the agnostic concepts	• Process: A form of activity in which resources are used in the fabrication of goods or in the rendering of services; • Rule: A formulated logical constraint that would be used to control the execution of a Process; • Strategy: A Process specifically designed to achieve a goal and not a Product; • Objective: A desired state at the completion of a process; • Event: See the Event CODP; • Role: see the Role CODP.

The Content Synthesis step concludes the SLR research method by providing the consolidated set of agnostic CODPs. These agnostic CODPs are drawn from the literature using a qualitative form of the SLR approach proposed by (Okoli, 2015).

5.10 Conclusion and future work

The research question formulated in section 5.7 pertains to the inquiry into the elicitation of agnostic concepts that can be used as agnostic CODPs in a multi-domain ontology. Although positivist or hypothetico-deductive criteria of validation cannot apply here in a qualitative research (Guba & Lincoln, 2001), evidences are emerging to indicate that the findings of this paper's phenomenological research method is significantly consistent in its similarity to the findings of two other sources: this paper's companion publication (Fitzpatrick, Ratté, et al., 2018a) and the best practice research on CODPs in (Blomqvist, 2010). This significant similarity in the outcome of qualitative research, as in the case of this project's two companion papers along with Blomqvist research on CODP best practices, is referred to as triangulations. Anney in (Anney, 2014) recommends that one or two such triangulations be demonstrated as a criterion to establish the research's trustworthiness. The authors posit that, although this is an initial phase of a multi-phase project, the outcome of this phenomenological study demonstrated a credible inductive process in eliciting data model patterns from experienced practitioners that may be considered as experts in twenty out of twenty-two individuals based on criteria established in (S. Ahmed et al., 2005). Furthermore, the companion SLR is also followed by two use case papers: (Fitzpatrick, Coallier, et al., 2018) and (Fitzpatrick, Ratté, et al., 2018d). These use cases allow determining the transferability of the SLR. (Anney, 2014) indicates that transferability is the equivalent of positivism's generalizability criterion for qualitative research. Anney also posit that thick description and purposeful sampling facilitates transferability. Along with the involvement of several co-researchers in the execution of the phenomenological protocol (use of peer debriefing) (C. Moustakas, 1994) (Anney, 2014), an audit trail, thick documentation and the application of Okoli's best practice approach for conducting qualitative, this research has shown evidence of trustworthiness following the guidelines established in (Guba & Lincoln, 2001).

The authors consider that the phenomenological research method has supported quite adequately their needs for eliciting agnostic CODPs and other insights, such as prescriptive directions to eventually study design methods for multi-domain ontology based applications to resolve semantic heterogeneity. While it is expected that qualitative research protocol will predominate in this research project for some time in the future, it is conceivable that, on occasions, when sample size and other conditions are met to perform hypothetico-deductive methods that theory-testing protocols may complement the current approach.

Following this phase of the project, where an SLR approach and a phenomenological research method were used, a new group of about twenty-five participants will be solicited to become co-researchers. The phenomenological research method will be executed identically as in the present study. Additional semi-structured interview questionnaire, surveys and focus group sessions will be designed to further investigate some questions studied in this paper such as additional agnostic CODPs, additional domain-specific concepts, the influence of lines of business and others. This project intends to increase the size of the co-researcher group from twenty-two to approximately 100.

CHAPTER 6

ESTABLISHING TRUSTWORHTINESS OF A DUAL METHOD QUALITATIVE RESEARCH FOR ELICITING AGNOSTIC CONTENT ONTOLOGY DESIGN PATTERNS IN A MULTI-DOMAIN ONTOLOGY

Daniel Fitzpatrick[1], François Coallier[1], Sylvie Ratté[1]

[1]Department of Software Engineering & Information Technology, École de technologie supérieure,
1100 Notre-Dame West, Montréal, Quebec, Canada H3C 1K3

Paper submitted for publication to the Journal on Data Semantics in September 2018

Abstract

All private companies and government agencies require their systems to interoperate. System interoperability facilitates crucial exchange of data to solve business problems and engage in commercial opportunities. Semantic heterogeneity consists in the phenomenon where enterprise systems are designed based on various vocabularies that render information sharing difficult or impossible without a data integration function. A data integration function represents a palliative measure that attempts to provide data seamlessly as if it came from only one source. This paper intends to establish the trustworthiness of a research project that intends to solve the semantic heterogeneity problem. Due to the theory-building role of this project, a qualitative research approach constitutes the appropriate manner to conduct research. Contrary to theory-testing quantitative methods that rely on well-established validation techniques to determine the reliability of the outcome of a given study, theory-building qualitative methods do not possess standardized techniques to ascertain the reliability of a study. This project intends to use a dual method theory-building approach to more decisively demonstrate trustworthiness. The first method, a qualitative SLR approach based mainly on the guide provided in (Okoli, 2015), induces the sought knowledge from publications using a practical screen. The second method, a phenomenological research method based on the works of C. Moustakas, elicits mainly the agnostic concepts from semi-

structured interviews involving senior practitioners with eight years or more of experience in conceptualization.

The SLR retains a set of 89 agnostic concepts from 69 publications from 2009 through 2017. The phenomenological study in turn retains 83 agnostic concepts from 22 interviews. During the synthesis stage for both studies, data saturation was calculated for each of the retained concepts at the point, publication or co-researcher sequential number, where the concepts have been selected for a second time. The saturation points are tallied and represented on a diagram for each of the two studies. Although it can be asserted that this effort of establishing the trustworthiness can be construed as extensive and this research track is promising, data saturation for both studies has still not been reached. Further work is required using exactly the same protocols for each of the methods, expand the year range for the SLR and to recruit new co-researchers for the phenomenological protocol. This work will continue until these protocols do not elicit new theory material. At this point, new protocols for both methods will be designed and executed with the intent to measure theoretical saturation.

Keywords: Content ODP, Ontology Design Patterns, Ontology, inference application, multi-domain ontology, Systematic Literature Review, phenomenological research method, trustworthiness, constructivism, dual method, qualitative research.

6.1 Introduction

All private companies and government agencies require their systems to interoperate. System interoperability facilitates crucial exchange of data to solve business problems and engage in commercial opportunities. For example, in the manufacturing sector, new innovative design methods such as Set-Based Design (SBD) (Kerga et al., 2016) and the modular approach (Buergin et al., 2018) intend to increase performance and productivity. The SBD approach can reduce in average by 25% projects' duration and by 40% projects' costs (Kerga et al., 2016). In the defense sector, government agencies deem system interoperability crucial to deploy a multinational coalition force (Egon Kuster, 2007). System interoperability allows coalitions' members to exchange vital information to cooperate on effective deployment and operation planning (J. Patel et al., 2010) (Dorneich et al., 2011). Semantic heterogeneity consists in the phenomenon where enterprise systems are designed based on various vocabularies that render information sharing difficult or impossible without a data integration function. A data integration function represents a palliative measure that attempts to provide data seamlessly as if it came from only one source. (De Giacomo et al., 2018). This research investigates semantic structures, specifically agnostic Content Ontology Design Patterns (CODP) (Blomqvist, 2010) that can be used within a multi-domain ontology executed in an inferential application to effectively perform data integration in resolve semantic heterogeneity (Daniel Fitzpatrick et al., 2012). The research question is formulated in (Fitzpatrick, Ratté, & Coallier, 2018b) as: «*what are the conceptualization patterns found in semi-formal ontologies, e.g. data model patterns, software engineering patterns, etc. that can be agnostic to any domain or industry sector in the context of enterprise semantic interoperability and can be used as the basis of agnostic CODPs to resolve semantic heterogeneity in enterprise systems?*»

Also expressed in (Fitzpatrick, Ratté, et al., 2018b), this research project argues the following two theses:

- «*There is a set of data model patterns that are applicable to any private industry or government sector that can be used as agnostic CODPs and constitute a (formal) multi-domain ontology that can be used by an inferential data integration application to resolve the semantic heterogeneity problem*»; and
- «*A dual method qualitative research approach, using trustworthy SLR and phenomenological research methods, allows to elicit the sought agnostic data model patterns to form the (formal) multi-domain ontology for an inferential data integration application*».

The first thesis is addressed using a dual method qualitative research approach to elicit the agnostic CODPs. The first method, a qualitative Systematic Literature Review (SLR), induces the CODPs from papers published between 2009 and 2017 inclusively, detailed in (Fitzpatrick, Ratté, et al., 2018a). The second method inspired from Clark Moustakas' phenomenological research approach and thickly described in (Fitzpatrick, Ratté, et al., 2018c), elicits the sought CODPs through semi-structured interviews involving experienced senior practitioners with over eight years experience (S. Ahmed et al., 2005). In addition to the aforementioned main research processes, two uses cases shows the potential application of the elicited theory in the context of collaborative logistics planning (Fitzpatrick, Ratté, et al., 2018d) and collaborative product design for military coalition deployment (Fitzpatrick, Coallier, et al., 2018). This project's holistic design, i.e. of the overall research process, is described in (Fitzpatrick, Ratté, et al., 2018b). The second thesis, in respect to the choice of a dual method qualitative research method, is argued with thick description of three aspects of the research. Firstly, the research protocols and secondly the findings are detailed extensively in the individual SLR and phenomenological research publications (Fitzpatrick, Ratté, et al., 2018a, 2018c). The third aspect, the trustworthiness establishment approach, constitutes the subject of the present paper. As hypothetico-deductive related validation techniques assess quantitative research, trustworthiness four criteria provide the means to the reader to determine how the research deserves to be trusted (Borrego, Douglas, & Amelink, 2009) (Guba & Lincoln, 2001).

Section 2 reviews the four criteria of trustworthiness and also examines data and theoretical saturation. Section 3 presents the findings of both the SLR and the phenomenological research methods. Section 4 assesses the dual method qualitative research approach against the four trustworthiness criteria, first credibility that examines the intrinsic quality of the processes, then dependability that pertains on thick description, thirdly confirmability relating mainly to triangulation and finally transferability, which involves purposeful sampling, used here for the phenomenological study, data and theoretical saturation, and the two use cases. Section 5 outlines a discussion and section 6 concludes the paper with this research project's establishment of trustworthiness.

6.2 State of the art

This project's research approach and strategies consider the trustworthiness criteria as defined in (Guba & Lincoln, 2001) and (Anney, 2014). Added to the trustworthiness transferability criterion, the concept of theoretical and data saturation, first introduced in the grounded theory method, allows to determine at a point during the qualitative research process when no new data or theory are created (Saunders et al., 2017). This is an emerging and elusive concept that is difficult to apply since theoretical sufficiency can only be determined post-mortem (Sim et al., 2018). Since this project intends to serve as a starting point in a series of other research initiatives, the project does not set a saturation goal. The project is set to only measure theoretical (data) saturation for the purpose of planning future work.

Table 6.1 describes the trustworthiness criteria prescribed by (Guba & Lincoln, 2001) and (Anney, 2014) to conduct qualitative research and the key design decisions made to ensure that the research process design satisfies these criteria. First of the trustworthiness criteria is the credibility criterion, which entails that the findings are considered believable by various stakeholders such as publication's editorial boards and the participants (co-researchers) to the research. This is done through thick description and by triangulation, i.e. the relative similarity of the findings using methods with different data sources such as a Systematic Literature Review (SLR) eliciting data from rigorously selected publications and a phenomenological research method extracting data through semi-structured interviews. Secondly, the transferability criterion allows examining how the findings can be used in a specific context through use case scenarios, for example. Thirdly, the dependability criterion involves an audit trail. Finally, the confirmability is established by the capacity of the research design to allow very similar findings to be produced by other researchers.

In the phenomenological research segment of the project, purposeful sampling allowed to select only senior co-researchers with eight years' experience or more. (Suri, 2011) refers to this purposeful sampling approach as criterion sampling. Furthermore, co-researchers are asked to introduce other potential participants on a voluntary basis, which Suri refers to as snowball sampling. Snowballing consists in the co-researchers reaching out referred potential participants and asks permission to be contacted by the researcher or invited to contact the researcher directly. Also part of the phenomenological research method, bracketing allows the researcher to mitigate the risk associated with the researcher's bias on the phenomenon itself. While being a senior practitioner thus establishing investigator authority, a credibility sub-criterion, the researcher may also induce a bias in analyzing and synthesizing the data and producing the findings. The researcher's experience must not influence in any way the findings of this study, complying to (Bevan, 2014) citing (Husserl, 1970) in refraining from using the researcher's personal knowledge in a phenomenological research method. However, the researcher's knowledge of the phenomenon allows determining peripheral issues such as defining the notions of accuracy and quality of a data integration model. The notion of bracketing is covered in more detail in (Fitzpatrick, Ratté, et al., 2018c). Added to the trustworthiness's transferability criterion in (Forero et al., 2018), the concept of theoretical and data saturation, first introduced in grounded theory method that allows to determine the point during the qualitative research process when no new datum or theory is created (Saunders et al., 2017). This is an emerging and elusive concept that is difficult to apply since theoretical sufficiency can only be determined post-mortem (Sim et al., 2018). Since this project intends to serve as a starting point in a series of other research initiatives, the project does not set saturation goals. The project is set to only measure theoretical (data) saturation for the purpose of planning future work and not to establish trustworthiness.

Table 6.1 describes in more detail the trustworthiness criteria and associated measures drawn from (Guba & Lincoln, 2001) (Anney, 2014) (Forero et al., 2018) and (C. Moustakas, 1994) that are used in this project. Firstly, the credibility criterion establishes to what extent the qualitative method(s) and the findings may be trusted and believed. Secondly, the dependability or repeatability criterion intends to demonstrate that the same or at least very similar findings would be obtained using the same data, co-researchers and publications but with a different researcher or researchers. Thirdly, the confirmability criterion attempts to establish to what extent the methods can be used with different co-researchers (participants) in the phenomenological research approach. Also, use cases covering different industry sectors, while addressing the same research question, would contribute similarly in establishing transferability. Fourthly, and finally, the transferability criterion intends to show that the research design can be used in other contexts for different research questions, theses or problems to solve.

Table 6.1 Trustworthiness criteria for a dual method qualitative research

Criteria	**Detailed measures for establishing trustworthiness**
Credibility	• Involving if possible more than one researcher. In Moustakas phenomenological research method, participants may be empowered to become co-researchers and participate in a more active way than in a more traditional setting; • A pilot project allows testing the research protocol; • Researchers possess training and experience in designing questionnaire, conducting interviews and in the research subject matter; • All notes taken during the interviews, the interviews' recordings and all the worksheet representing every stage of the analysis and synthesis activities are kept in safe storage; • The transcripts of the interviews allow the co-researchers to confirm the knowledge transmitted during the interviews.
Dependability	• The research protocol is richly documented; • The findings are richly described; • The establishment of the trustworthiness criteria is richly described as well; • All steps in the protocols with intermediate results are documented and can be audited.
Confirmability	• Investigator triangulation is performed when more than one researchers are involved; • Other researches may constitute data source and investigator triangulations; • Data source triangulation consists in the context of qualitative research as inducing knowledge and know-how from more than one source, e.g. publications vs. participants; • Methodological triangulation originates from the use of more than one research method, either quantitative, qualitative or both;

Table 6.1 Trustworthiness criteria for a dual method qualitative research (continued)

Criteria	Detailed measures for establishing trustworthiness
Transferability	• The use of more than one sampling technique; • A use case provides context to the application of an emerging theory and attempts to demonstrate to what extent the emerging theory can be applied to solve the research problem; • The quantification of data saturation and theoretical saturation may provide a form of assessment on the sample size and the relative state of the theory-building process. However, the data source and theoretical saturation concepts are often confused, lack standards and are still embryonic in the literature (Marshall et al., 2013) (Sim et al., 2018). In this project, data saturation is an assessment of the relative state of theory-building process using the same research question and protocol, e.g. the practical screen definition in an SLR, the questionnaire for semi-structured interviews, etc.; • Theoretical saturation represents a continuation of data saturation by using different research questions and protocols. For example, in the context of this project, new search queries and questionnaires may be designed to allow exploring in greater detail specific modules or subcomponents of the multi-domain ontology. While the concept of saturation of the theory already exists, the distinctive data and theoretical saturations as proposed here represent an innovative addition to qualitative research methodology; • The present project expects that variations of the current SLR practical screen and questionnaire will be needed to ensure completeness of the theory. Other knowledge induction techniques such as focus groups may be needed to reach the sought completeness. At this point, it is difficult to determine a priori when either data or theoretical saturation are reached.

This section examined the four criteria needed to establish the trustworthiness of a dual method qualitative research design executed using SLR and phenomenological research protocols, and, two use cases demonstrating transferability of the set of the elicited agnostic CODPs. The four trustworthiness criteria, credibility, dependability, confirmability and transferability provide the readers the means to assess the proposed research design. Although currently the subject of great scrutiny in treating trustworthiness, data and theoretical saturation represents here means to plan future research activities in using the same protocols as previously executed for the former type of saturation and changing the protocols for the latter type. Data saturation is expected to be reached first with the same practical screen applied to the uncovered years prior to 2009 in the case of SLR study and with the interview of new co-researchers using the same process and questionnaire for the phenomenological study. For theoretical saturation, modified protocols for both research methods, and perhaps new research methods will be used in the attempt to complete the elicitation process. In the next section, the findings for both the SLR and phenomenological studies are represented to support the measures taken for the trustworthiness criteria.

6.3 Protocols and findings from the dual method qualitative research studies

The previous section outlines the approach to establish the trustworthiness of this project's qualitative research process. The quantitative research methods use internal and external validation techniques within a hypothetico-deductive reasoning process (P. Leedy & Ormrod, 2012). While the validation process is clearly the responsibility of the researcher, the burden of establishing trustworthiness is shared with the reader of a qualitative study (Borrego et al., 2009). This project considers that the researcher can ease the burden of trustworthiness in qualitative research on both the researcher and the reader by adopting a rigorous trustworthiness approach as proposed in this paper based on (Guba & Lincoln, 2001), (Anney, 2014) and (Forero et al., 2018). In addition to thick description, an elaborate qualitative theory-building approach clearly establishes and demonstrates to the readers the trustworthiness, i.e. the credibility, dependability, confirmability and transferability of the research approach meant to elicit agnostic CODPs. A qualitative research approach intends to build theory while demonstrating that it and the theory it is building deserve trust. Data saturation occurs when no new theory is added with the same protocol. Theory saturation would consist in no additional theory being added even with various protocols and different research methods. Theory saturation constitutes a state where the theoretical framework is complete. Although undetermined at this point, future phases of the project would see the use of mixed qualitative quantitative and ultimately quantitative hypothetico-deductive research. Since data saturation is not at this point achieved, the next phase of the project will involve the same protocols, including the SLR's same practical screen except for different years and interviews with new co-researchers using the same questionnaire for the phenomenological study.

6.3.1 SLR research protocol and findings

This SLR takes its methodological roots from (Kitchenham, 2004), (Okoli, 2015) and (Okoli & Schabram, 2010). The SLR approach can be performed in either the quantitative or the qualitative research methods. This paper outlines a qualitative SLR based on the need to create theory about agnostic CODPs for a multi-domain ontology for performing data integration (Fitzpatrick, 2012). The following SLR steps are further detailed in (Fitzpatrick, Ratté, et al., 2018a).

6.3.1.1 Previous exploratory literature survey

A previous exploratory literature survey in this project identified conceptualization patterns in semiformal ontologies. Prior to the undertaking of this SLR, a lengthy multiyear conventional literature review was performed. Over 200 articles were found and studied. This conventional literature review supported a qualitative research project conducted using a phenomenology method in an exploratory fashion. Although the guides used in this SLR do not prescribe to start with an SLR research with an exploratory literature survey, this project includes it as a necessary primer step.

6.3.1.2 Formulation of the research objective

This activity indicates the purpose of the research and is reproducible. In the context of a qualitative SLR, as it is the case here, the objective is broad (P. Leedy & Ormrod, 2012).

6.3.1.3 Formulation of a research question

As indicated by (P. Leedy & Ormrod, 2012) and (John W Creswell, 2003), a research question, not hypotheses, guides the remaining activities for a qualitative research.

6.3.1.4 Drafting the protocol

The design of the protocol for this SLR draws from (Okoli, 2015; Okoli & Schabram, 2010) for all steps of the protocol except for the Analysis and Synthesis steps. The Analysis and the Synthesis steps originate from the adapted phenomenology research method outlined in (Fitzpatrick, 2012).

6.3.1.5 Formulating the practical screen

The practical screen establishes the criteria that will allow the researcher of this SLR to select the publications that will be analyzed and synthesized. The criteria ensure the feasibility of completing the SLR by allowing a number of publications that can be read and treated by the authors. The practical screen comprises two subdivisions: metadata level and content level. The metadata level comprises any information available without actually reading the publication. The metadata level part of the practical screen allows only to either entirely reject the publication or allowing it to be further examined at the content level part of the practical screen. The content level provides the criteria that will allow the researcher of this SLR to retain and further process part or all of the content.

6.3.1.6 Search results.

The logical query defined in the previous step is executed in each of the publication databases earmarked in the practical screen. The metadata level criteria allow the retention or the rejection of publications without actually reading the content in first elimination. Once the metadata level part of the screening is completed, the retained publications' content is examined, but not analyzed, to determine if there is any material that can be used in the content of this SLR. Some publications may be rejected if no material of interest is found.

6.3.1.7 Content analysis

Each publication is then read for analysis. This SLR authors' previous publications are the first to be analyzed. The note-taking technique employed here consists in using Nuance Communications's Dragon Naturally Speaking dictation software where speech is converted into text and inserted in a Microsoft Word document. The extracted components are: the main agnostic concept, the subsumed subordinate concepts, the definitions and relationships. The properties, rigid properties and instances are not covered by this SLR. The documentation is segmented by publication and then by main agnostic CODP.

6.3.1.8 Content Synthesis

Agnostic CODPs found in all retained publications are then merged with same concepts that were elicited in the previous step. The documentation for the content synthesis step is segmented by agnostic CODP and represented in a simplified domain diagram where the patterns are represented as classes and not in an axiomatic form. The axes for the synthesis activity are for each CODP: the universal thing concept, the main agnostic concept, the subsumed subordinate concepts, the definitions and relationships. The rules are based on the same rules used in this paper's companion publication that uses a phenomenological research method to also elicit agnostic CODPs for a multi-domain ontology. The ontology elements and structures are considered as meaning units as in the phenomenological approach. And as in the phenomenological research method, the semantic material extracted in this SLR is coalesced using the described rules.

6.3.1.9 SLR findings

The statistics show the total number of publications displayed after executing the search query in all research sources libraries from 2009 through 2017 inclusively. The search query listed a total of 860 publications from the source libraries prescribed in the practical screen over nine years. Figure 6.1 shows 69 papers, or eight percent of the 860 returned publications from the query, retained publications for analysis and synthesis once the filtering criteria are applied. As established in the metadata level criteria of the practical screen, this SLR's authors' publication (Daniel Fitzpatrick et al., 2012) are included in the statistics although being elicited in the query. The small number of publications that were finally retained can be explained mainly by publications that treated the matter regarding data model patterns without actually showing any.

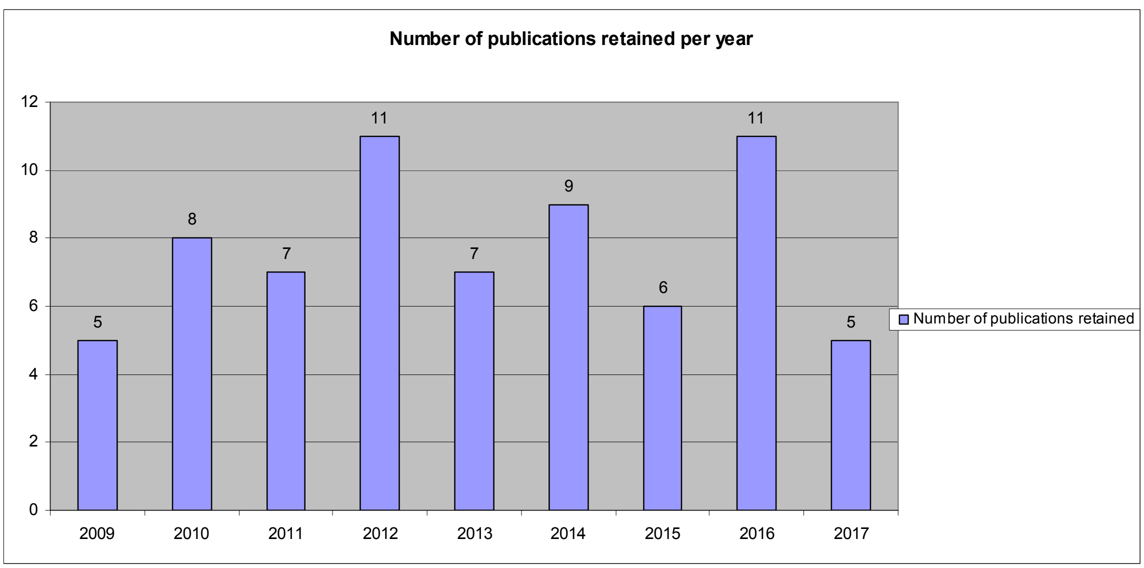


Figure 6.1 Number of publications per year screened and retained
for analysis and synthesis

The first papers analyzed are some of this SLR's authors' previous publications, i.e. (Fitzpatrick, 2012; Daniel Fitzpatrick et al., 2012, 2013; D. Fitzpatrick et al., 2013). These publications cover research performed on the concept of Reference Architecture – Enterprise Knowledge Infrastructure (RA-EKI). RA-EKI defines processes, data structures and ontologies to produce knowledge, actionable information, and know-how, functional knowledge. It proposes an assembly line like epistemological approach to convert data into information, then information into knowledge and know-how. Knowledge and know-how are stored and executed from an ontological structure composed notably of the multi-domain ontology, a contribution of this project. These publications, while describing RA-EKI, also provided the following descriptions of agnostic concepts in Table 6.3. Only concept names and descriptions are provided. This set of agnostic concepts and the multi-domain ontology architecture modules serve as the foundation, the starting point, for the content synthesis process.

6.3.2 Phenomenological research protocol and findings

6.3.2.1 Preparation

This protocol step sees the design of the questionnaire. The first set of questions intends to outline the contextual aspect, i.e. the background, of the co-researcher, notably the number of years the participant had experience in conceptualizing as a data modeler, data architect, software engineer, developer, etc. The question about the years of experience allows the researcher to verify that the potential co-researcher meets the minimal years of experience criterion of eight years. The other background question indicates the various industry sectors the practitioner has performed conceptualization. (Suri, 2011) refers to this purposeful sampling approach as criterion sampling. Co-researchers are asked to introduce other potential participants on a voluntary basis, which Suri refers to as snowball sampling. Snowballing consists in the co-researchers reaching out to the referred potential participants and asked permission to be contacted by the researcher or invited to contact the researcher directly. The preparation step also involves the design of the questionnaire with the following questions outlined in table 6.2.

Table 6.2 Questions used for the semi-structured interview

Question no.	Question formulation
Q01	How many years have you performed conceptualization, e.g. data models, canonical models, domain model, XSD, etc.?
Q02	What are the industry and government sectors have you performed conceptualization?
Q03	Name and describe abstract (agnostic) concepts that you believe may apply to any industry and government sector.
Q04	Indicate relationships between these abstract concepts.
Q05	For a maximum of three industry or government sectors, list domain specific (low abstract) concepts and identify to which abstract concept they relate to (generalization specialization only).
Q06	Do you believe that a data integration function should be designed using abstract (agnostic) concepts as you indicated in question 3? Provide a score from 1 to 10. Please comment.
Q07	Do you believe that a data integration function should be designed using low abstract (domain specific) concepts that would be understandable by business users? Provide a score from 1 to 10. Please comment.
Q08	Do you believe the problem of semantic heterogeneity (see the introduction deck) should be addressed by scientific research?
Q09	Have you participated as a designer, architect, developer or software engineer in the development of a data integration core structure for a data warehouse or of a canonical model? This question does not constitute a precondition for the continuation of the interview.
Q10	Did you ever observe line of business influence on the design of a data integration platform? Please comment.
Q11	How do you or would you define and measure the efficiency of a data integration model?
Q12	How do you or would you define and measure the quality of a data integration model?
Q13	Optional snowballing: If willing, could you please refer one or two persons, with conceptualization experience (8yrs+).

6.3.2.2 Bracketing

This step consists in the researcher to explicitly express own beliefs in answering the questionnaire using text and diagrams. Before the start of the first interview, the researcher answers in writing the questionnaire. The researcher also draws light UML diagrams to represent the agnostic concepts, relationships and associated definitions. Furthermore, the researcher abstains from participating in the phenomenological study. Bracketing and the researcher's non-participation contribute to preserve the integrity of the research process (Bevan, 2014), (C. Moustakas, 1994), (Hays & Wood, 2011).

6.3.2.3 Interview

The researcher provides a preparation document that describes the research along with the questionnaire between three and five days before the scheduled time for the interview. At the scheduled time, the researcher calls the co-researcher as agreed and recapitulates the information previously provided. After obtaining the permission to record the interview, the researcher and co-researcher then cover in order as a very informal conversation. The researcher performs imaginative variation in providing a context or adding detail considerations to a question. For example, the researcher reminds throughout the interview that the co-researcher in answering should discount any constraint that would normally influence the design of a data integration platform in real life, such as politics, funding, etc. The imaginative variation technique, widely recognized as a trademark component of the phenomenological research methodology (C. Moustakas, 1994) (Wertz, 2005).

6.3.2.4 Transcript

While recording the interview, the researcher notes the agnostic concepts, their relationships and the domain specific concepts with generalization-specialization relationships with agnostic concepts, along with a summary of the responses from the other questions from the co-researcher. Once the interview completed, the researcher listens to the recordings and completes the transcripts to be sent to the co-researcher for approval. This approach ensures the accuracy and the richness of the notes taken during the interview and allows eliciting the most difficult data to collect such as comments to questions and the concept and relationship definitions (Bevan, 2014).

6.3.2.5 Content analysis

The analysis process elicits extracts agnostic CODPs along with their definitions, relationships, the "low-abstract" domain specific concepts and the subsumption relationships. The researcher breaks down the elicited material, meaning units, in spreadsheets. The spreadsheets also reflect for each of the 22 interviews which agnostic concepts were provided by the co-researchers. This account are used to contain the meaning units in various forms, such as comparative series of scoring with questions Q06 and Q07, comparing the average and standard deviation of the numeric responses. The domain specific concepts are to be used in future "use case" reports that would comprise a competency question directed to a given industry or government sector.

6.3.2.6 Content synthesis

The researcher aggregates the extracted meaning units and uses the rules listed in table 6.3 to perform the synthesis step. The synthesis step consists in integrating disparate meaning units from the transcripts into a consolidated set of agnostic CODPs. The integration of meaning units extracted from the content analysis step is guided using the RA-EKI multi-domain ontology architecture. The RA-EKI multi-domain ontology architecture provides modules that house the agnostic CODPs.

Table 6.3 Meaning unit coalescence rules

Meaning unit number	**Meaning unit type description**	**Meaning unit coalescence rule description**
1	Years of experience of the co-researcher	Basic aggregating statistical functions such as average and standard deviation.
2	The industry or government sectors that the co-researcher performed conceptualization.	Basic aggregating statistical functions such as average and standard deviation.
3	The agnostic concepts	• Concepts defined in the same manner are retained if it was identified by at least two co-researchers; • In the case of synonyms, only the term with the greatest selection by co-researchers is retained. In case of equal number of selections, the researcher makes the final decision; • In the case of concepts that have been defined in more than one way, the same rule as in the case of synonyms applies.
4	The subsumption and other relationships between the agnostic concepts.	• The relationships need to be selected only once to be retained; • In case of conflicting relationships, only the one with the greatest number of selections is retained.
5	The definition or description of the agnostic concepts.	The texts are integrated by the researcher.

Table 6.3 Meaning unit coalescence rules (continued)

Meaning unit number	**Meaning unit type description**	**Meaning unit coalescence rule description**
6	The de facto agnostic CODPs derived from the above-mentioned meaning units.	The aforementioned meaning units are then integrated in distinct modules using RA-EKI's module structure as a starting point (Daniel Fitzpatrick et al., 2013). The researcher may decide to diverge from the SLR's architecture on a case-by-case basis. The researcher, for example, may opt to rename and redefine the Contract module to Agreement if the phenomenology research reverses the subsumption relationship between Contract and Agreement.

6.3.2.7 Findings from the phenomenological study

The 22 semi-structured interviews by telephone lasted between 60 and 90 minutes. The co-researchers had all previously received preparation material and the questionnaire. The first two questions provided context to the study in terms of years of experience in performing conceptualization, an average of 21 years, and the industry sectors the co-researchers have conceptualized in average 6.3 different industry sectors. The numbers of years of experience in conceptualization of the co-researchers range from eight to 40 years. The three phenomenon questions directly relate to the sought agnostic CODPs, relationships, definitions and associated domain-specific concepts. The findings are outlined in table 6.4 co-located with the findings from the SLR study and the results from the best practice study performed in (Blomqvist, 2010). This table provides an insight that can be used for triangulation, which will be further discussed in section 4.

6.3.3 Findings related to agnostic CODPs from both SLR and phenomenological studies

In the previous sections, the protocols and some of the findings specific to each of the SLR and phenomenological studies have been discussed. The core meaning units that are common to both methods, the agnostic CODPs are shown in table 6.4. This table means to support the assessment of the triangulation trustworthiness criterion, discussed in section 4. The table outlines the CODPs placed in the (architecture) modules identified in RA-EKI (Daniel Fitzpatrick et al., 2013). Both lists in table 6.4 only contain retained agnostic CODPs using the second selection rule, which entails that an agnostic concept is retained when elicited by a paper for a second time in the context of the SLR, or elicited by a co-researcher also for a second time in the context of the phenomenological study. The concept names in bold represent agnostic CODPs those are common to both studies.

Table 6.4 Agnostic CODPs elicited in the dual method SLR and phenomenological studies

RA-EKI modules	**SLR study's agnostic CODPs**	**Phenomenological study's CODPs**
Party	**Organization , Person , Party**, Position, Department, Name, Organization Unit, Company, Government Agency	**organization, person, party**, individual
Product	**Product**, **Service**, **bill-of-material**, **Part, Equipment**, **Facility, Good**, **Inventory**, **Item**, **Unit of Measure**, **Package**, Requirement, Vehicle, Product Type, Service Type, Material, Measure, Order, Order Line, Road, Quantity	**product**, **service**, **bill-of-material**, **part**, **equipment**, **facility**, **good**, **inventory**, **item**, **unit of measure**, **package**, building, cost, market, request
Agreement	**Agreement**, **Contract**, **Term**,	**agreement**, **contract**, **term**, tacit agreement, law
Price	**Rate**, **Price**	**rate, price**

Table 6.4 Agnostic CODPs elicited in the dual method SLR and phenomenological studies (continued)

RA-EKI modules	**SLR study's agnostic CODPs**	**Phenomenological study's CODPs**
Event	**Event , Currency**, **Payment**, **Time**, **Transaction**, Transaction Type, Period of Time	**event, currency**, **payment**, **time**, **transaction**, credit, debit, charge, financial transaction, communication, amount
Document	**Document**	**Document**
Network	Edge, Vertex	network item
Account	**account**	**account**, account receivable, account payable, general ledger, charter of account, invoice, invoice line
Context	**Context**	**Context**
Location	**Address**, **Email address, Telephone**, **Location**, **Country,** State, City	**address**, **email address**, **telephone**, **location**, **country,** web site, place, IP address, URL, continent, grid, area
Role	**Role, Actor**, **Asset**, **Customer**, **Employee**, **Supplier**, **Resource**, Relationship, Relationship Type, Vendor, Role Type, Agent, Contextual Role, Organization Role, Person Role, Contact Mechanism	**role, actor**, **asset, Customer**, **employee, supplier**, **resource**, party role, locator, consumer, contact
Process	**Task**, **action**, **Process**, **Rule**, **Business rule**, Plan, Operation, Sale, Strategy, Task Type, activity, Process Type, Business process, Channel, Goal, Project	**task**, **action**, **process**, **rule**, **business rule,** control, regulation, objective
Concept	**Concept**, Entity, Model	**Concept**

Both studies show common concepts in all modules except Network. In the case of the Network modules, the SLR shows the "edge" and the "vertex" concepts corresponding to network links and nodes respectively. For the phenomenological study, only the "network item" concept is elicited. If the Network modules of both studies were integrated, the "network item" concept, being higher abstract, would subsume the "edge" and "vertex" concepts. The other twelve remaining modules for both studies share the same key concepts, i.e. the concepts that bear the same name as the modules. In the case of the Party module, both studies share, in addition to “party”, two other important concepts i.e. “person” and “organization”. The Product modules share “good” and “service”, the two important subclasses subsumed by “product” and shared by both studies. (Blomqvist, 2010) also shares the same common concepts for the Party and the Product modules as both studies in this project. The cited Blomqvist paper reports a research project that elicited ontology design patterns as best practice cases based on their cross-domain applicability. The Blomqvist study also shares common concepts with the SLR and phenomenological studies not only for Product and Party, but also with the Process, Role and Event modules. It is noteworthy indicating that both studies contain unselected concepts, i.e. elicited only once in a paper or during an interview, that would have been common to both studies. For example, the SLR elicited but did not retain concepts such as “cost”, “building”, “control” and “network item” that were retained in the phenomenological study. It also important to indicate that, in addition to cases of same name concepts, concepts from both studies, or within the studies themselves, may be related on bases of synonymy, antonymy, hyponymy, meronymy and holonymy. For example, “request”, “individual”, “objective” from the phenomenological study and “order”, “person” and “goal” from the SLR may respectively be considered as synonyms. Several concepts within each study and between them can also be related in generalization-specialization relationships. For example, “position”, department”, “organization unit”, “company”, government agency” could be construed as specializations of the “organization” concept. These exemplary ascertainments may be confirmed with the use of the Wordnet (Miller, 1995) thesaurus in conjunction with more focused systematic literature searches and phenomenological semi-structured interviews in future steps of the project. The use of Wordnet may assist in preparing proposed terminological assertions to be submitted in SLR and phenomenological studies (Zong et al., 2015).

The outcome of the elicitation of agnostic CODPs from both SLR and phenomenological protocols, albeit in a project still in its infancy, provides interesting insight considering that either data source or theoretical saturations have been achieved. In the case of data saturation, both protocols in their current design have not reached a point where no additional theory is added. Consequently, the SLR study would expand its scope for publication years before 2009 and after 2017 while being treated with exactly the same protocol, i.e. with the same research question and definition of the practical screen. Similarly, the phenomenological research needs to continue with new co-researchers while using the same questionnaire. Extending the searched publication years may allow to better assess data saturation for the SLR. Figure 6.2 shows a downward sinusoidal trend in the number of agnostic concepts being retained in chronological order of analyzed and synthesized publications.

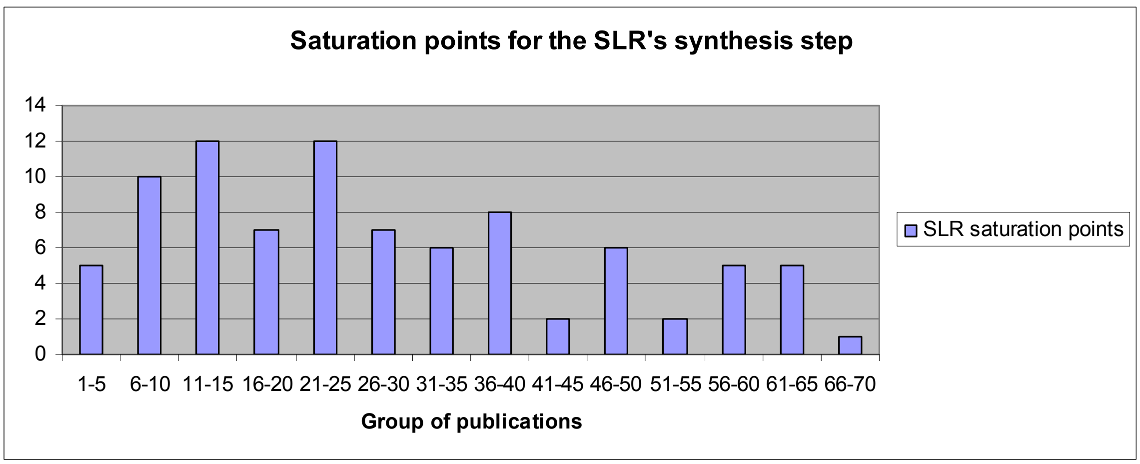


Figure 6.2 Saturation events in the SLR synthesis step

Figure 6.2 appears to indicate that the protocol in its current design is near a point of data saturation. In the next stages of this project, an expansion of the publication years range and processing of other papers will provide a better understanding of the data saturation concept. This expansion is to be performed with the same research question and the same practical screen. The expansion consists in extending the study to cover a number of years before 2009. In figure 6.3, the same data saturation downward sinusoidal graph this time for the phenomenological study provides a relative state of completion of the research process. Data source or theoretical saturations do not represent accurate and reliable a priori indications for when the research is expected to be completed (Sim et al., 2018). The authors in (Guest et al., 2006) mentioned that in the context of their project, the planned 60 interviews were completed before realizing post-mortem that their social study project achieved 92%

saturation on the 12th interview, which would have satisfied the requirements of their research. In the case of this project's phenomenological study, data saturation appears to be less advanced than the SLR's.

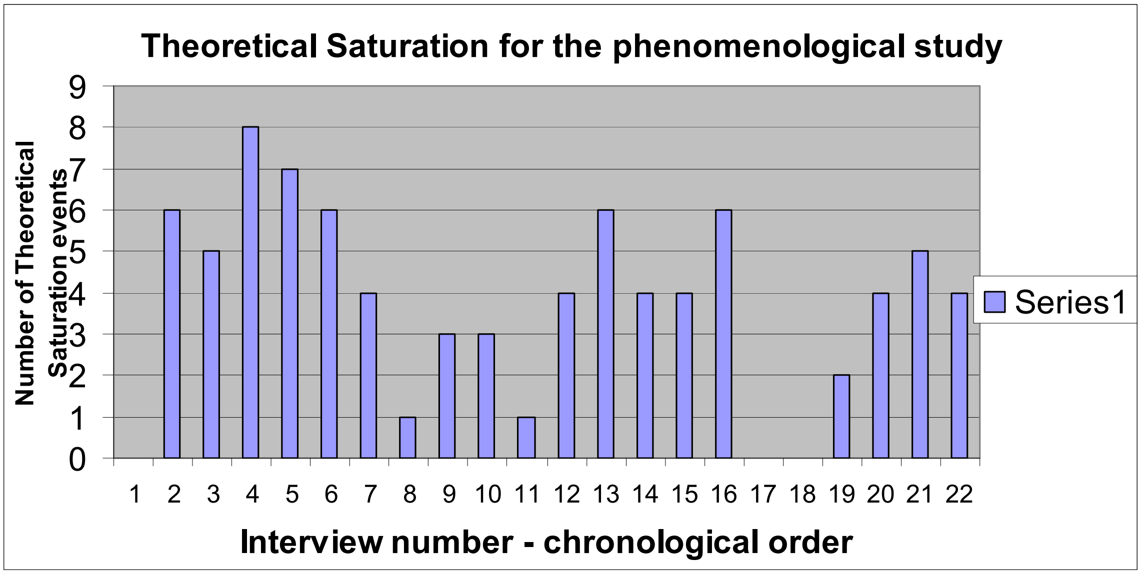


Figure 6.3 Progression of the theoretical saturation events

The next stage of the project will involve expanding the number of publication years prior to 2009 for the SLR, and in due time 2018 and on, and recruiting new co-researchers for the phenomenological protocol. In the next section, the dual method qualitative research approach and the findings relevant to this paper are discussed.

6.4 Assessment of the trustworthiness of the dual method approach

In the previous section, the findings relative to both SLR and phenomenological studies were outlined and discussed. In this section, both of the dual method approach and the relevant findings are assessed using the trustworthiness criteria described in section 2. Although the research project is arguably only in its infancy, interesting conclusions of this first leg of the journey may be drawn.

As covered in detail in section 2, trustworthiness of qualitative research encompasses the credibility, dependability, confirmability and transferability criteria. The dual method qualitative research approach used in this project and the specific findings relative to elicit agnostic CODPs is assessed in the next few lines.

6.4.1 Credibility

- In the phenomenological study, the participants are deemed co-researchers (C. Moustakas, 1994).The researcher empowers the co-researchers in providing background material such as a summary of the project, research method, etc. Suggested reading is also provided to the participants;
- A pilot project performed prior to the design and execution of the dual method approach, allowed testing the questionnaire for the semi-structured interview used in the phenomenological research study and the fine-tuning of the search criteria for the SLR;
- The researcher possesses training and experience in designing questionnaires and conducting interviews. Furthermore, the researcher tallies over 30 years experience in conceptualization and data model patterns;
- All notes taken during the interviews, the interviews' recordings and all the worksheet representing every stage of the analysis and synthesis activities are kept in safe storage;
- The transcripts of the interviews allow the co-researchers to confirm the knowledge transmitted during the interviews.

6.4.2 Dependability

- The dual method qualitative research design is thickly documented in (Fitzpatrick, Ratté, et al., 2018b). The individual protocols for the SLR and phenomenological research methods are richly described;
- The findings for both SLR and phenomenological research studies are also richly documented. The findings in the SLR include the agnostic CODPs, their definition and their relationships. The findings in the phenomenological study include the same

meaning units as in the SLR. In addition to the same meaning units as the SLR, the phenomenological study also produced findings, the contextual meaning units, about the years of experience of the co-researchers and the industry sectors they intervened. Finally, the phenomenological study elicited peripheral meaning units such as the co-researchers' appreciation on the usage of agnostic CODPs, forming the multi-domain ontology, in the design of a data integration function;

- The establishment of the trustworthiness criteria on the dual method qualitative research design is thickly described in the present paper;
- All steps in the SLR and phenomenological research protocols are documented with intermediate results in spreadsheets.

6.4.3 Confirmability

- Investigator triangulation is performed in the phenomenological study with 22 co-researchers eliciting agnostic CODPs. Also in other research projects, strong commonality with this project's findings is deemed as investigator triangulation;
- Data source triangulation stems from the dual method protocols where data on agnostic CODPs are elicited from publications and senior experienced practitioners, as well as other research projects;
- Methodological triangulation originates in this project from using two different qualitative research protocols for data collection. Both SLR and phenomenological protocols use the same analysis and synthesis approach, inspired from Moustakas' approach on performing analysis and synthesis (C. Moustakas, 1994). The analysis and synthesis approach is considered very common in qualitative research with some variants and also described as a de-contextualization and re-contextualization cycle (Thomas & Harden, 2008).

6.4.4 Transferability

- A single purposeful sampling criterion for the phenomenological research protocol consists in choosing practitioners with eight years' experience in conceptualization;
- Two use cases are written and executed to demonstrate the applicability of the elicited agnostic CODPs in the context of collaborative (manufacturing) product design and collaborative logistics planning for military coalition force deployment.

6.5 Discussion

Although this project is still at an early stage, both of the SLR and phenomenological studies' set of agnostic CODPs show common concepts in 12 out of 13 modules. This comparison only considers same name concepts. The comparison does not consider synonyms or generalization-specialization relationships, notably, within and between both studies. Using a reliable process that would not involve the researcher's opinion and that data saturation would be achieved in both protocols, it is likely that a much greater number of common concepts would be obtained. This would contribute significantly to establish a much greater trustworthiness and to a greater consensus on resolving semantic heterogeneity.

As discussed in section 3, the same commonality of concepts is also observed with the set of best practice ontology design patterns identified in (Blomqvist, 2010). Furthermore, in one of the publications elicited in the SLR (West, 2011), the author proposes an agnostic cross-industry general-purpose data model, the High Quality Data Model (HQDM). The HQDM model is inspired from ISO 15926 (Leal, 2005), a data integration model standards for the oil industry sector but generic enough to be used in other sectors as well. HQDM uses similar concepts to the ones elicited in this project. Concepts such as "party", "role", "agreement", "person", organization", "price", "process" and several others are common to HQDM and this project dual methods' findings.

This project primarily explores the architecture and design of a multi-domain formal ontology to resolve semantic heterogeneity using agnostic content ontology design patterns that would be usable in any industry sector. During the phenomenological study, the co-researchers provided a relatively close to unanimous response about the usage of agnostic concepts for the design of a data integration platform. Responding with an average score of 8.6 out of 10 with a standard deviation of 1.4, the co-researchers clearly and collectively emphasized the importance of agnostic concepts in data model patterns for data integration. On the other hand, two systems of beliefs emerged regarding the use of domain specific (low-abstract) concepts in the design of a data integration platform. One group opposed at various degrees the use of domain specific concepts in designing a data integration function and considered that only agnostic concepts should be used. The other group considered that domain specific should be used with agnostic concepts to design a data integration platform. More details can be found in respect to the specific findings of the phenomenological study in (Fitzpatrick, Ratté, et al., 2018c). This project expects both system of beliefs will be examined concurrently to explore both designs of a multi-domain ontology: one that only uses agnostic CODPs and a second one that uses both (cross-industry) agnostic and domain specific CODPs. It is important to indicate that the two new research tracks are either completely antagonistic or partially antagonistic to the position taken by the authors in (Diego Calvanese et al., 2009). In their paper, Calvanese and co-authors argue that a data integration function's design would be based on domain specific concepts as viewed by a user. We counter-argue that a domain specific data integration ontology would either partially or totally exacerbate the semantic heterogeneity problem in an enterprise, based on the early evidence elicited using this project's dual method qualitative research protocols. This project considers the publication of the ISO 15926 standards as a significant achievement in terms of the recognition by a whole industry sector of the importance of agnostic conceptualization in the design of a data integration platform. As argued in this project, agnostic concepts may be used as agnostic CODPs for the formal multi-domain ontology, to be eventually used in the development and run time operation of a cognitive data integration application.

6.6 Conclusion

This paper intended to establish the trustworthiness of a research project based on a dual method qualitative design. The project's fundamental purpose is to contribute in solving the semantic heterogeneity problem. The semantic heterogeneity problem hinders all industry sectors' efforts, private and governmental alike, to ensure interoperability between enterprises IT systems. The RA-EKI architecture model uses an ontology layered structure that includes a type of mid-level ontology called a multi-domain ontology, composed of modules, that is designed to play a key role in data integration and other cognitive applications by defining a cross-industry semantic structure. Guarino and co-authors in (Guarino et al., 2009) posit that only a richer set of axioms may enhance an ontology. Such a richer set of axioms can only be obtained through an effective and quality-driven conceptualization, a language-independent concept. Such quality conceptualization can be found in data model patterns as proposed by M. West in (West, 2011) based on ISO 15926, a highly generic data integration model standards. Based on the works of Thomas Erl on service-oriented architecture, conceptualization such as West's HQDM is considered agnostic since the data model's conceptualization can serve to design a data integration platform usable in any industry sector. This project's objective is to elicit agnostic data model patterns here considered as content ontology design patterns. The primary thesis of this project is that such agnostic CODPs do exist and can be used to solve the semantic heterogeneity problem. Due to the theory-building role of this project, a qualitative research approach constitutes the appropriate manner to conduct research. Contrary to theory-testing quantitative methods that rely on well-established validation techniques to determine the reliability of the outcome of a given study, theory-building qualitative methods do not possess standardized techniques to ascertain the reliability of a study. The secondary thesis of this project is that a dual method theory-building approach may demonstrate trustworthiness. The first method, a qualitative SLR approach based mainly on the guide provided in (Okoli, 2015), induces the sought knowledge from publications using a practical screen. The second method, a phenomenological research method based on the works of C. Moustakas, elicits mainly the agnostic concepts from semi-structured interviews involving senior practitioners with eight years or more of experience in conceptualization (C. Moustakas, 1994).

The SLR retains a set of 89 agnostic concepts from 69 publications from 2009 through 2017. The phenomenological study in turn retains 83 agnostic concepts from 22 interviews. During the synthesis stage for both studies, data saturation was calculated for each of the retained concepts at the point, publication or co-researcher sequential number, where the concepts have been selected for a second time. The saturation points are tallied and represented on a diagram for each of the two studies. While this measure constitutes an element of trustworthiness notably by (Forero et al., 2018), this project can only use it for planning purposes since data saturation cannot be used on an a priori basis, i.e. it cannot serve to predict if the planned sample size for interviews or otherwise is sufficient. Although it can be asserted that this effort of establishing the trustworthiness can be construed as extensive and this research track is promising, data saturation for both studies has still not been reached. Further work is required using exactly the same protocols for each of the methods, expand the year range for the SLR and to recruit new co-researchers for the phenomenological protocol. This work will continue until these protocols do not elicit new theory material. At this point, new protocols for both methods will be designed and executed with the intent to measure theoretical saturation. For both the methods, this entails in formulating new research questions that may, for example, focus on agnostic themes such as finances, infrastructure, relationships, classifications, etc. For the SLR, this may translate into designing a practical screen that will search for publications specialized in specific agnostic themes. For the phenomenological study, this may entail designing new questionnaires for semi-structured interviews and possibly employing other knowledge elicitation techniques such as focus groups.

CHAPTER 7

DISCUSSION

We set sail on this new sea because there is new knowledge to be gained, and new rights to be won, and they must be won and used for the progress of all people…We choose to…do… things, not because they are easy, but because they are hard; because that goal will serve to organize and measure the best of our energies and skills.

John F. Kennedy, September 12, 1962, "We choose to go to the Moon" speech.

This chapter aims to provide a better understanding of this project's findings. This chapter also intends to explain consequences and ramifications related to not only the individual research processes described in the chapters but also to the project as a whole. We also relate the findings to key studies and look back to the research questions (Hess, 2004) (Jenicek, 2006).

Firstly, we discuss the selection of the individual research methods and the design of the overall approach. We then cover the specifications and findings of the qualitative SLR. The two use cases on collaborative product design and collaborative military logistics planning are discussed specifically about the intent to transfer the elicited SLR knowledge in the form of agnostic CODPs to industry settings. In a similar fashion to the SLR, the phenomenology study is critically examined in respect to its activities and its findings. Finally, we discuss the consequences and ramifications of the establishment of trustworthiness.

The inherent challenge to theory building is that there is no validation approach in a true hypothetico-deductive sense. Also, there is no unified framework to perform analysis and synthesis in qualitative research. Contrary to quantitative researchers, qualitative investigators must spend time and effort to design a research process that deserves to be trusted. Qualitative researchers in IS, IT and software engineering are sometime confronted with distrust from fellow researchers who did not heed Wanda Orlikowski and Jack

Baroudi's warning that homogeneously using positivist inspired research methods may be detrimental to the IS research domain (Orlikowski & Baroudi, 1991). Furthermore, any attempt to explicitly establish the confidence on a qualitative research may contradict fundamental interpretivist tenets, although not universally shared, against attempting to validate qualitative findings. We greatly inspired ourselves from the Bano team's multi-method approach. Instead of phenomenology, the Bano team utilize case studies along with an SLR (Bano et al., 2017). Bano and her team have not cover or demonstrate the intent to explicitly covered trustworthiness criteria. This project's choice of the phenomenological research method allowed, we believe, a more credible inductive process at least until the multi-domain ontology can submit to experimental trials with consensually agreed upon measurements. As Mulrow indicated in a short but thoughtfully written paper: «*Systematic literature review is a fundamental scientific activity*» (Mulrow, 1994) . We strongly consider that the use of a qualitative SLR method, while not universally recognized as such in the IS, IT and software engineering domains, constitutes an imperative, especially for kick-starting a new research track.

The qualitative SLR method is the more subjective of the two methods used in this project's dual method approach since it relies on a single individual to elicit the data. In the case of the phenomenological method, all 22 co-researchers elicit from their «first person» experience the sought agnostic patterns and other knowledge such the notions of quality and efficiency. The weakest point in terms of rigor remains the reading of publications retained after applying the metadata screen, the filtering logic used to reject or accept publications before being read. The researcher reads the text of each publication after passing the metadata screen in the attempts to detect and collect agnostic concepts and relationships to form agnostic CODPs. The researcher has previously performed bracketing before starting the dual method process by documenting his own belief. Nevertheless, the researcher remains a imperfect data collection instrument. This fact further justifies using a dual method approach. The query selected 860 publications before being filtered by the practical screen. The practical screen retained 69 publications that had agnostic concepts and relationships and yielded 89 agnostic concepts.

For the most part, publications that were not retained did not actually show any agnostic concept or pertained on formal ontologies. The SLR elicited 89 agnostic concepts from 69 retained publications from 2009 through 2017. The SLR's original query selected 860 publications. We determined that the study is nearing (relative) data saturation based on the position of the downward sinusoidal curve's position reaching the abscissa of the SLR's diagram of saturation events. However, this should not be construed as an end of the execution of this particular SLR protocol. It is important to note that there is only one SLR for the ODP science community to date (Hammar & Sandkuhl, 2010). This project produced the first SLR that intended to elicit ODPs.

As indicated earlier in this chapter, the use cases were designed to explore their use for establishing transferability. This has not yet been covered in the contemporary literature. In a quantitative research, we would use an external validation approach to show potential generalizability. Both use cases, after introducing the problem and context, surveyed each two areas of literature: the subject matter literature and the publications pertaining to ontology research applied to the subject matter (business) domain. For collaborative (military) logistics planning, the use case includes a review of military planning papers and publications related to research on the use of ontologies for supporting business processes. For collaborative product design, the use case surveyed papers pertaining to: Set-Based Design (SBD) (Kerga et al., 2016) and the modular approach (Buergin et al., 2018). In both use cases, the transferability attempt to demonstrate is significantly limited to the subjective application of the SLR's elicited agnostic CODPs to a specific business area by the researcher. Although the business concepts were in relatively small number and related mostly to the product and process modules, the future of this specific transferability approach is to be reviewed for methodological enhancements to reduce the level of subjectivity.

The phenomenological research method used in this project gathered one of the most experienced group of participants, the co-researchers, in a subject related study based on (Simsion et al., 2012) with over 20 years experience. The co-researchers contributed 83 agnostic concepts. The co-researchers also provided mainly generalization-specialization relationships and examples of low-abstract domain specific concepts subsumed to the agnostic CODPs. It is important to note that few concept definitions have been provided, which will be taken into account when future phases and projects are planned for continuing this research track. During the phenomenological study, the co-researchers provided a relatively close to unanimous response about the usage of agnostic concepts for the design of a data integration platform. Responding with an average score of 8.6 out of 10 with a standard deviation of 1.4, the co-researchers clearly and collectively emphasized the importance of agnostic concepts in data model patterns for data integration. This in itself, provides an antagonistic position to using a user-centric set of domain specific concepts to design a semantic data integration platform as advocated by (Diego Calvanese et al., 2009).

On the other hand, two systems of beliefs emerged regarding the use of domain specific (low-abstract) concepts in the design of a data integration platform. One group opposed to various degrees the use of domain specific concepts in designing a data integration function and considered that only agnostic concepts should be used. The other group considered that domain specific concepts and associated semantic elements should be used with agnostic concepts to design a data integration platform. This former position closely aligns with the HQDM model, based on the ISO 15926 data integration model, developed by Matthew West in (West, 2011) and contains only highly abstract (agnostic) concepts. As discussed in chapter 2, a commonality of concepts is also observed with the set of best practice ontology design patterns identified in (Blomqvist, 2010), which reinforces confidence in the approach and the findings.

CONCLUSION AND CONTRIBUTIONS

> "What one believes is irrelevant in [science,
> only what can be argued matters…]"
> Stephen Hawking character's in the movie
> Theory of everything (2013)

This section provides a recapitulation of this project's fundamental tenets, the problem to be solved and the research questions that were addressed by the research processes. A closing statement ends this section establishes the direction of future research as prescribed in (Aitchison, 2016).

Twenty-five ago, Orlikowski and Baroudi alerted the IS scientific community to the detrimental effect of homogeneously applying positivists inspired, hypothetico-deductive, methodology to advancing science. A decade later, Gregor and co-authors have proposed a theory on theory, the descriptive, explicative, predictive and prescriptive components of any theoretical framework. They also incited the larger computer science community, which this project includes IS, IT and software engineering communities to conduct projects using interpretativists inspired qualitative methods. In 2012, in a vibrant call to order, Ivar Jacobson, pioneer in software engineering and reputed member of the UML "three amigos", called upon software researchers to get together and formally build a universal software engineering theory, citing Gregor's work. The project deliberately embarked on the contentious road of inductive research to solve the semantic heterogeneity problem, which affects all enterprises' efforts to interoperate their systems. We now conclude the first leg of a journey we hope will definitely solve the "old" problem.

The exploration initiative we have also taken considers the sensitivity of using theory building methods within the engineering field. The project has taken great care by considering developing a trustworthiness establishment approach. The project also made it clear that theory testing deductive methods will also be used at the earliest opportunity. This project initiated the first steps of this research track by asking two questions: what are the agnostic CODPs that may constitute the building block of a cognitive data integration platform, and, what is the appropriate approach to conduct the research. For the second question, we proposed a dual method qualitative research approach based on an IS methodologically similar project conducted by Bano and her team on the relationships between user involvement and the success of a system development project. We then introduce a clear and explicit strategy to establish the trustworthiness of the approach and its findings with the understanding that the project will spawn into subsequent phases and other projects that will likely use inductive methods.

The selected methodology, the SLR and the phenomenological research methods, have elicited 89 and 83 highly abstract concepts respectively in the form of agnostic CODPs. These design patterns will eventually be translated into terminological axioms and compose the multi-domain ontology, centerpiece of the RA-EKI framework and reference model for the purpose of solving the semantic heterogeneity problem. In the course of the phenomenological study, the project also showed in a preliminary fashion, that the use of agnostic concepts in the design of a data integration platform is strongly prescribed almost unanimously by the 22 co-researchers. We also demonstrated that significantly more research is needed to eventually derive quality and efficiency metrics for measuring the data integration function.

Based on the triangulation criterion and other trustworthiness criteria as well, we conclude that the dual method inductive approach has produced, again in a preliminary fashion, interesting insight in identifying candidate agnostic CODPs notably that will be critical to plan and design future protocols. The project has also demonstrated in an adequate manner that this research track deserves to be continued.

Contributions

The project has contributed methodological and architectural elements that may benefit not only ontology engineering but also other IS, IT and software engineering scientific domains as well.

These contributions are:

- The Reference Architecture – Enterprise Knowledge Infrastructure.

  RA-EKI represents one of the first cognitive architecture models to be outlined and to cover the full epistemological spectrum. It was the subject of conference papers (…) and presented at various conferences including the International Conference on Product Lifecycle Management and on the International Conference on Military Computer and Communications Systems. In an earlier form, the research plan of this framework was submitted to the first doctoral workshop of the 2012 International Conference on Product Lifecycle Management in which it was awarded the first prize of the best research plan;

- A Multi-Domain Ontology.

  This type of mid-level formal ontology was the first to be published that entails the conceptualization and representation of a universal set of all business concepts. As agnostic CODPs are elicited, the multi-domain ontology will expand and will be experimentally developed as the terminological component of a cognitive data integration platform;

- A dual method qualitative research approach.

  This is the second dual method qualitative approach to be used in the greater computer science domain. This approach contrasts with the popular quantitative-qualitative mixed method approach in that it only includes inductive processes and techniques;

- Distinct data and theoretical saturation concepts.

  Currently confused as synonyms in all related publications, the two saturation concepts are used distinctively by this project. Data saturation represents the state of completeness of a theory at the protocol level. Theoretical saturation shows a relative state of completeness for the entire research track. Since several protocols and methods will be required to complete building the data integration theoretical framework, data saturation will be measured for each protocol execution. Theoretical saturation should provide an overall assessment of the research track's progression.

RECOMMENDATIONS AND FUTURE WORK

"*As with all aspects of the research design, the theoretical perspective one chooses, whether positivist, interpretivist… is ultimately driven by, and must be consistent with, the research questions of the study [and the problem it is trying to resolve]*."
(Borrego et al., 2009)

Based on the findings of this research, we recommend that use cases in other domains be written to illustrate the role of the SLR's agnostic CODPs for solving competency questions. The competency questions are drawn from two conference papers that previously covered these domains at a more holistic architectural level (Daniel Fitzpatrick et al., 2013) and (D. Fitzpatrick et al., 2013). The new use cases will cover the competency questions at a more detail ontology design level, using this SLR's elicited CODPs.

Following the final formulation of the resulting conceptualization composed of the set of agnostic CODPs elicited in this research project, the multi-domain ontology is to be formulated as a formal ontology using the OWL language with an approach as proposed in (J. Dietrich & Elgar, 2005) and deployed in the form of an Application Programming Interface (API) as prescribed by (Horridge & Bechhofer, 2011).

Finally, in the wake of this project, it is recommended to investigate intends a position in which single domain ontologies would be contraindicated for runtime operation of any cognitive applications. This contraindication would apply for cognitive application capable of knowledge reuse, as described in this SLR at section 2.2.3, for data integration or any other inferential applications. However, single domain ontologies would be used in development time as input to the design of the multi-domain ontology prior to its deployment in run time within a cognitive application.

The author consider that the phenomenological research method has supported quite adequately their needs for eliciting agnostic CODPs and other insights, such as prescriptive directions to eventually study design methods for multi-domain ontology based applications to resolve semantic heterogeneity. While it is expected that qualitative research protocol will predominate in this research project for some time in the future, it is conceivable that, on occasions, when sample size and other conditions are met to perform hypothetico-deductive methods that theory-testing protocols may complement the current approach.

Following this phase project, where an SLR approach and a phenomenological research method were used, a new group of about twenty-five participants will be solicited to become co-researchers. The phenomenological research method will be executed identically as in the present study. Additional semi-structured interview questionnaire, surveys and focus group sessions will be designed to further investigate some questions studied in this paper such as additional agnostic CODPs, additional domain-specific concepts, the influence of lines of business and others. This project intends to increase the size of the co-researcher group from twenty-two to approximately 100.

ANNEX I

A reference architecture for semantic EDW with multi-domain data integration capability

Daniel Fitzpatrick[1]

[1]Department of Software Engineering & Information Technology, École de technologie supérieure,
1100 Notre-Dame West, Montréal, Quebec, Canada H3C 1K3

Research Plan submitted to the IFIP WG 5.1 1st Doctoral Workshop
International Conference, PLM 2012, Montreal, July 9-11, 2012

Abstract

In the context of a broadened product lifecycle management environment, a traditional product information management, also referred to as product master data management (P-MDM) needs to be complemented by other MDM domains. Such MDM domains may include Customers, Financials, Suppliers, Human Resources, Events and other domains. To satisfy such a transversal set of requirements requires a true cross-enterprise semantic integration capability. This capability cannot be met by current off-the-shelf technologies. This paper proposes a research approach that would elicit the definition of a reference architecture and a multi-domain ontology, from research and development work performed notably in ontology engineering, in both academic and industry domains.

Keywords. Product lifecycle management, product master data management, ontology-based data integration, data architecture, qualitative research

I.1 Context

Industry sectors have vested interest in technology that allows sharing data, information and knowledge within the enterprise and with the outside world. Through interoperability, the enterprises are looking to improve the product-centric processes' efficiency and robustness to cut waste and sustain growth. The PLM concept comprises a large array of data domains, e.g. financials, customer, etc., which are traditionally used also by other process paradigms such as customer-centric and supplier-centric, notably.

The pervasiveness of the data used by product-centric processes represents a challenge in providing consistent, coherent and unified data as if provided seamlessly by a single source.Product lifecycle management (PLM) is one of the keystone paradigms that bring value to the stakeholders, notably shareholders and customers. In the aftermath of what is currently called the great recession, PLM processes are focused to sustain growth, to improve products and processes on a continuous basis and eliminate wasteful activities and constraints.

I.2 Problem statement

For cross-enterprise PLM product-centric processes, source database heterogeneity constitutes an important problem. The processes require a single point of truth in acquiring coherent and consistent data in a seamless manner. Especially in large enterprises, data must be extracted from a great number of systems, each possessing its own syntactic and semantic structures. Shortcomings in the methodology and technology increase the complexity of the work of designing a multi-domain data integration capability to be not only challenging but also failure prone.

I.3 Hypotheses

1. There exist data architecture patterns that allow efficient (through reusability) multi-domain semantic integration in the enterprise. A pattern here is a generic solution to a recurring problem in the form of a conceptual data model or any other types of ontology;

2. The primary concepts of these data architecture patterns are rich axioms that can constitute the core structure of a multi-domain ontology. In other words, semantic efficiency and cross-enterprise capability in semi-formal ontologies, as obtained in certain best-of-breed EDW projects, will be obtained with the same types of concepts in formal ontologies.

I.4 Research objective

This doctoral thesis intends to propose a reference architecture comprising a multi-domain ontology-based data integration capability, as a corner stone, to fulfill the inherent interoperability requirements for the PLM product-centric processes.

I.5 Theoretical background

I.5.1 Product Lifecycle Management

The Open Group Architecture Framework, or TOGAF, provides the theoretical foundation that can assist an organization to implement an enterprise architecture practice. TOGAF comprises notably an Architecture Development Methodology, a documentation management approach, high level specifications to ensure system interoperability with the use of an enterprise ontology and of the Integrated Information Infrastructure Reference Model (III-RM). The III-RM represents a high level architecture pattern to implement system interoperability through integrated information brokerage between the organization's systems. The reference architecture proposed by the research project is a more specific instance of the III-RM pattern and uses the semantic enterprise data warehouse concept to deliver the information brokerage capability.(Group, 2009)

I.5.2 An epistemological perspective

Information technologies draw in part from philosophy. Logic and epistemology have inspired, per example, the creation of the relational model (Codd, 1970) and the new emerging research initiatives such as IBM’s Hyper project that introduces the use of epistemic logic at the heart of the peer to peer data integration concept (D. Calvanese, Damaggio, De Giacomo, Lenzerini, & Rosati, 2004).

Data integration is a term that may even be questioned in the course of this study as a suitable title for the research project. Furthermore, the primary purpose of data integration is to supply ultimately knowledge and intelligence for research, decision making, predicting and other knowledge based activities.

As indicated in (Liew, 2007) and (Bouthillier & Shearer, 2002), the notions of data, information and knowledge remains elusive. In order to clearly elaborate an architectural approach for data integration, a theoretical stance must be taken where these fundamentals are described with as much rigor as possible. Figure I.1 illustrates the building blocks behind semantic integration.

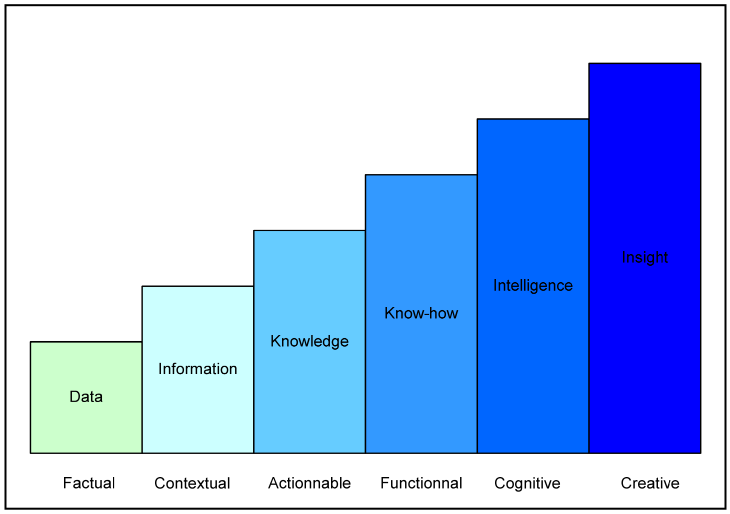


Figure I.1 Building blocks behind data integration.

The building blocks are represented here as concepts used by the human mind to, per example, decide on a course of action. The concepts here, from (Liew, 2007), (Michaels, Goucher, & McCarthy, 2006), (Sajja, 2008) and (McInerney, 2002) are:

- Data: factual elements represented by symbols that can used for analysis or computer processing;
- Information: data that are assembled thru a context, contextual data
- Knowledge: a set of information elements that can lead to taking action, actionable information;
- Know-how: a structure composed of knowledge and propositional predicate, forming a functional construct;
- Intelligence: a super set of know-how allowing self-learning capability, a cognitive construct;
- Insight: or wisdom gained by cumulative form of intelligence resulting in advanced reasoning and creativity.

I.5.3 Product Lifecycle Management

This business paradigm covers human, material and data assets, along with processes to manage and execute the various activities involved for each product from the early stages of R&D and design, or beginning-of-life (BOL), thru the commercial stage of the product life, or middle-of-life (MOL), and terminating at its retirement, or end-of-life (EOL) (Terzi et al., 2010).

PLM evolved as a more complex set of processes, a value-chain, used for creating value for shareholders and customers alike. It involves using information, knowledge and know-how to continuously perfect on product efficiency, performance and quality. Some of its processes have the capacity to trace manufacturing errors and other quality and performance issues, to monitor product through logistics store and transport, material recycling and energy saving. Finally, PLM also consists in optimal decision-making through product lifecycle stages, from BOL to EOL. A data integration capacity makes it possible to properly deliver timely information and knowledge for PLM processes and also for collaborative activities with other business paradigms, such as the customer-centric CRM. Table 1 illustrates various types of data needed for the PLM product life stages (Matsokis & Kiritsis, 2010; Terzi et al., 2010). This is only a minimal list of types of data. This research is likely to unearth a much greater list.

Table I.5 Types of data needed at the PLM product lifecycle stages

PLM Product life stages	**Types of data**
Beginning-of-life	Product, equipment, material, plant, employees, tools, techniques, methodologies, document, suppliers,
Middle-of-life	Product, customer, employees, services, service providers, events, geography, financials, document,
End-of-life	Product, customer, service, service providers.

I.5.4 Master data management (MDM)

Master data are the data that allow the organization to reach its objectives. Master data is used to produce valuable contextualized information and knowledge in to support PLM. (Panetto, Dassisti, & Tursi, 2012) This research considers all data as master data in the context of a specific enterprise's PLM environment. Data subjects, such as parties, products and others constitute a more reliable data taxonomy system.

(Dreibelbis et al., 2008) and (Dyché & Levy, 2006) propose the Coexistence implementation style. The Coexistence style (see figure I.2) integrates data from heterogeneous sources in a batch mode in the context of an enterprise data warehouse environment. It integrates master data and delivers back to its sources, but usually also in a batch mode. Although it produces a golden record that can be used to alter master data located in source system, it does not constitute a system of record since change is not instantaneous. Great care must be taken in the correcting master data in operational systems using the MDM's golden record. It uses a physical database instance in a read-only mode approach. In some cases, a direct Enterprise Application Integration (EAI) feed allows some near-real time or even real-time events or other data to be loaded for intraday event processing. The coexistence implementation style serves as the basis for the architecture of an enterprise data warehouse for the PLM pardigm.(Loser, Legner, & Gizanis, 2004)

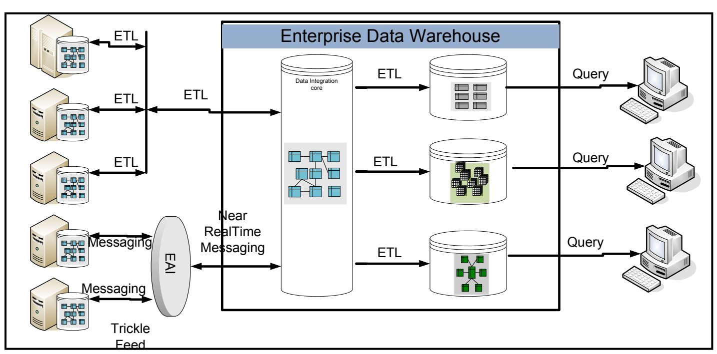


Figure I.2 Coexistence implementation style
with trickle feed

I.5.5 Ontology

An ontology is defined as an «explicit representation of a shared conceptualization». (T. R. Gruber, 1993) The basic purpose of the ontology is to produce a shareable and reusable set of information elements to be used by people and computer systems. Also, the ontology must distinguish between domain knowledge that may be extra organizational versus localized application level knowledge. The criterion of orthogonality is defined as the requirement of basing a newly created ontology on one or more existing ontologies. This practice, if generalized, would help reduce the silo effect in the development of ontologies. It would therefore favor the trend toward a greater universal interoperability across all industries. (Smith, 2008) The preliminary results outlined in this paper illustrate how the criterion of orthogonality is applied.(D. Fitzpatrick, F. Coallier, & S. Ratté, 2012)

A conceptualization is independent of the notional language. However, an ontology's specification, or representation, is dependent on the language. An ontology is a logical theory that describes the intended meaning to its defined vocabulary, in other words, using the committed concepts to a particular conceptualization of the real world. It is important to remember that ontologies only approximate a conceptualization. The only way to enhance the representation is to develop a richer set of axioms.(Gruber, 1995) The search for a richer set of axioms explains this research project's interest for data architecture patterns for multi-domain data integration developed in the industry for acquiring the sought semantic richness. An ontology is also defined as a formal, referenceable and consensual representation of a set of shared concepts to a domain with classes, properties, and relationships amongst them. (Salguero, Araque, & Delgado, 2008) The use of a formal ontology implies treating it through a semantic Reasoner.

A domain comprises objects and properties verbs and paraphrases that identify activities, processes and primitive concepts constituting the theoretical basis. A task ontology provides a specification of strategies designed to solve problems, for example fuzzy logic, neural network, constraint solver, etc.

Guarino classifies all ontologies in four types:

- Top level or foundational ontologies, such as Cyc, SUMO and Proton describe some of the basic objects of reality such as time, matter, action etc. These concepts are independent of a particular problem or domain. This type of ontology supplies the fundamental concepts serving as the basis to define the other type of ontologies;
- Domain ontologies, where domain ontology represents semantically the vocabulary of a generic domain that may exist in several organizations;
- Task ontologies describe a generic process structure that can be used to solve a certain type of problem;
- Application ontologies, which describe semantic entities that stem from a domain and task ontology or ontologies, both providing a specific function context (N. Guarino, 1998).

There are essentially three types of ontology applications:

- To support the mediation between people and ontology representing a vocabulary for the exchanges between people and organizations;
- Domain interoperability, support to develop (development time application) or to operate (run time application) systems of the same or different domains;
- Knowledge reuse requires the highest level of rigor, in addition to axioms, other concepts and their properties, ontologies for knowledge reuse will rely heavily on constraints and other type of restrictions. Problem solving methods or PSM have the capacity to support shared knowledge. They often include generic algorithms to perform various functions within the domain.

Figure I.3 illustrates a summarized definition of an ontology. One type of application that is growing in popularity in the research domain is ontology-based information extraction through natural language processing (NLP). (Navigli & Velardi, 2008; Völker et al., 2008; Wimalasuriya & Dou, 2010) In (Ratté et al., 2007), NLP processes are proposed to extract information from the organization's internal documents. These aspects constitute key elements behind the proposed reference architecture in this research project.

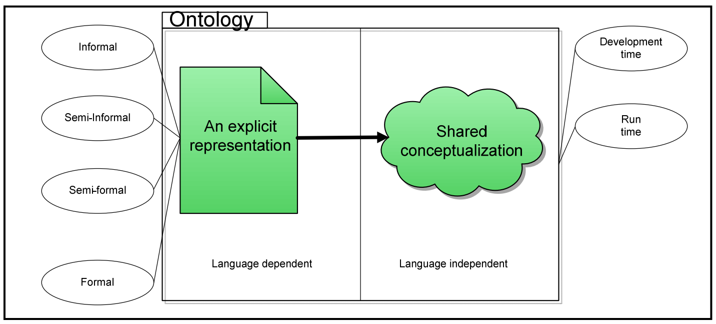


Figure I.3 Summarized definition of an ontology

An ontology does not impose the application of properties to a given instance of a class or concept. The finality here should be to build libraries of reusable knowledge and knowledge services available on networks. Ontological commitments or agreements pertaining to classes and relationships of an ontology are discussed among software agents and knowledge bases. (T. R. Gruber, 1993). A concept definition is a human readable text that in itself provides significance, meaning therefore semantically whole. (Gruber et al., 2009), (Noy & McGuinness, 2001)

An effective equilibrium must be achieved in defining ontology constrains rules in order to avoid affecting the concept abstraction level in the ontology even if it supports interoperability in a more effective manner. Affecting the ontology's abstraction level may lower the robustness and flexibility of the vocabulary. (Spyns et al., 2002)

Semantic relationships are categorized as synonymy, antonymy, hyponymy, meronymy and holonymy relations. Synonymy relationships relate two similar concepts. An antonymy relation indicates opposing or disjoint concepts. The Hyponymy category pertains to a generic to specific relationship between concepts. The meronymy and holonymy relationships support the build of material structure between concepts, the former indicates that a concept is included in another one, while the latter indicates that a concept includes the object of the relationship. Figure I.4 illustrates the conceptualization aspect of an ontology that is language independent (Lacy, 2005).

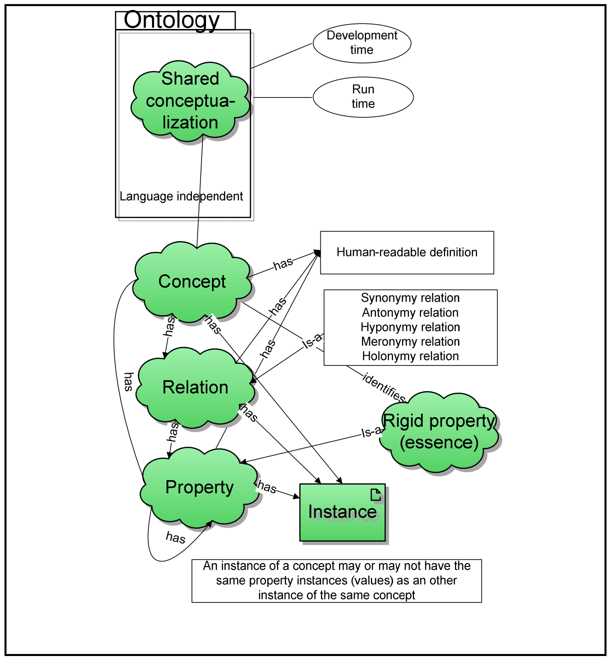


Figure I.4 Language independent aspects of an ontology : the conceptualization

Ontologies can be used to solve syntactic and semantic problems, and to automate data integration. However, some of the ontologies written in specialized languages such as OWL, RDF, RDFS, PLIB have grown to be voluminous and are becoming difficult to execute in main memory. A hybrid solution has been proposed by both academic and industrial organizations to address to address the in memory loading of voluminous ontologies (Khouri & Bellatreche, 2010).

Figure I.5 illustrates the language dependent aspects of ontologies. In terms of their level of formalism, there are: highly informal, semi-informal, semi-formal and highly formal ontologies. The first level of formalism is the highly informal level. It refers to a natural language text. In the case of semi-informal, an ontology is represented as a restricted and structured form of natural language, such as a concept map. In a case of a semi-formal ontology, the vocabulary would be expressed in an artificial language such as pseudocode or an entity relationship diagram. Finally at the highly formal level, ontologies possess "meticulously defined terms with formal semantics, theorems and proofs of such properties as soundness and completeness, i.e. classes including property information, value restrictions, more expressivity, arbitrary logical statements, first order logic constraints between terms and more detailed relationships such as disjoint classes, disjoint coverings, inverse relationships, part and whole relationships, etc.(Xie & Shen, 2006).

Formal ontologies can be based on first-order logic, frame-based constructs or both. (A. Gómez-Pérez et al., 2004; Lacy, 2005) The concept of multi-domain ontologies has been researched to facilitate the exchange data, information and knowledge between domains (Jinxin et al., 2002).

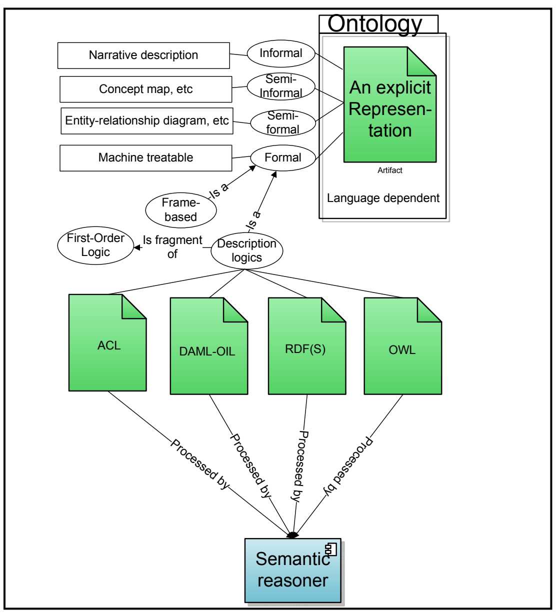


Figure I.5 The language dependent aspects of ontologies

I.5.6 Data integration

Taken holistically, data integration represents the computerized capability to address the problem of providing data thru a single perspective from heterogeneous sources located within an organization (Lenzerini, 2002). Along with data quality, data profiling and other

MDM functions, data integration attempts to service the organizations and the community at large with the widest perspective possible. Data is usually located in specialized systems. These silos are difficult to link together to provide transversal views of the data. There is a growing need to deliver cross-domain data, a usually highly difficult task considering that there are rarely any common semantic convention that may allow interoperability amongst systems (Ullman, 1997).

(Ullman, 1997) proposes a common data integration architecture composed of wrappers and mediators. In this architecture, source databases or systems are wrapped by specialized software components that convert the source's local semantics into a global set of shared concepts. The wrappers allow the source to which it is attached to interact with the rest of the world. Mediators are components that issue queries or sub-queries to wrappers or other mediators to gather data. Mediators are views that are designed to satisfy queries issued by humans and systems. Persistent forms of mediators are also designed in the form, notably, of enterprise data warehouses.

A research track covers the design of semantic enterprise data warehouses. The use of ontologies is central to this concept. Ontologies are not only used to design and execute data integration functions but to design multidimensional databases, design and implement data transfer processes more rapidly and to allow data queries in natural language.(Jiang, Cai, & Xu, 2010; Marrakchi et al., 2010; Nazri, Noah, & Hamid, 2010; Vaisman & Zimányi, 2012; Villanueva Chavez & Li, 2011)

I.5.7 Research gap

This project would address the scarcity of work on enterprise ontologies dealing with multiple domains in the PLM paradigm. The reference architecture will comprise a multi-domain ontology that is neither a foundational ontology, although may be based on some, and neither a domain ontology. A multi-domain ontology approach could significantly contribute to the research on data integration ontology engineering for semantic data warehouses.

I.5.8 Research questions

1. What are the main axioms of an enterprise multi-domain ontology for data integration that can support contemporary product-centric processes?
2. What generic architecture, or reference architecture, can cover the design of a semantic enterprise data warehouse (EDW) that can support PLM?

I.6 Methodology and data

The qualitative research protocol in this project involves a series of semi structured interviews to collect data architecture patterns and other related knowledge and know-how from seasoned and experienced practitioners. A pilot project phase will be conducted to test the questionnaire prior to the actual field research phase. Purposeful sampling will be done for both the pilot project and field research phases. Both phases will focus on the conceptualization aspect of the design of a multi-domain data integration capability. In addition to allow the extraction of more and richer pattern-like information throughout the field research part of the project, this approach provides two other important benefits: it assists the researcher to better select the interviewees («first-persons») and allows the researcher to submit himself or herself to a very rigorous and effective preparation to better conduct interviews. The data collection processes are executed in the context of the field research phase of the project in which a minimum of 15 participants are interviewed individually (C. E. Moustakas, 1994).

The current IT theoretical frameworks do not adequately support the industry in terms of knowledge and know-how.(Shirley Gregor, 2009) A qualitative research project to achieve the research objective is therefore warranted. For this purpose, a theory building qualitative research approach is considered here to tackle this research project problem (Halevy, Rajaraman, & Ordille, 2006).

Through the analysis processes, conceptual data modeling patterns would be identified along with valuable methodological heuristics such as how to ensure the reusability and robustness of the underlying conceptualization, used for the specific purpose of data integration. These findings will be used to formulate the intended reference architecture and multi-domain ontology. The final results of this project will be subjected to a validation process with the contribution of a 20-member committee composed of subject matter experts from the scientific and industry realms.

Following the data collection phase, data analysis is performed as illustrated in Figure I.7 and consists of the following steps (C. E. Moustakas, 1994) (J.W. Creswell, 2007; Patton, 2002; Tesch, 1990):

1. The Bracketing or Epoche step: the researcher, using the transcripts, identifies the preconceived opinions that he possesses on the subject matter, the research problem and the phenomenon itself, i.e. a successful ontology-based data integration capability. The researcher only retains what is essential, unbiased toward the phenomenon, by using a multiple perspective 'peeling-off' approach while going through the transcripts;
2. The Reduction step: the researcher then associates elements of text between them on the basis common characteristics that are existential, perceptual, etc. The reduction step does not entail shortening of the text;
3. The Imaginative variation step: the researcher generates and textural and structural meaning units using various angles, theories, domains, perspectives, that may be diverging, converging, etc. He uses his own experience and the literature pertaining on the phenomenon;

4. The Synthesis step: the researcher finalizes the data analysis activity by consolidating the textural and structural text fragments (or constituents) into data architecture patterns.

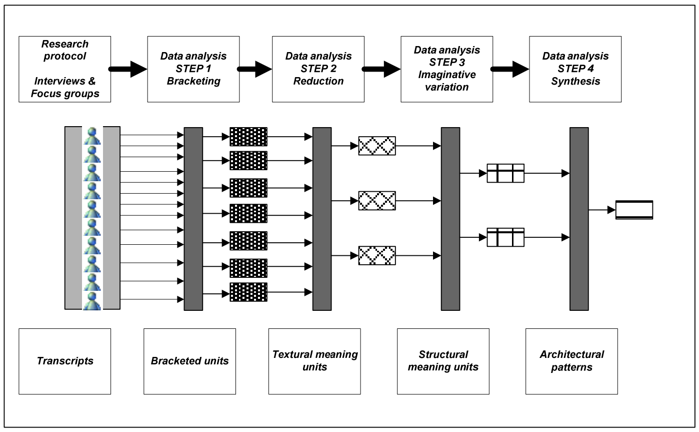


Figure I.7 Data analysis process

I.7 Expected results

About a dozen axioms that would serve as the fundamental set of concepts for the reference architecture's multi-domain ontology have been identified. Although the research is not yet completed, some of these axioms can be found in some of the widely used data modelling patterns used in the industry and successfully implemented in conventional enterprise data warehouse solutions. The reference architecture would also deal with the integration of semi-structured and unstructured data for PLM. It would also cover a dual channel data transfer concept with the ETL and EAI approaches.

Data collected from some of the participating practitioners were used to provide preliminary results for the research project. Inspired by the MDM coexistence implementation style with the trickle-feed function, discussed earlier, a reference architecture of a semantic enterprise data warehouse, as illustrated in figure I.8, is proposed to provide a multi-domain data integration capability to support contemporary PLM. Although, some of the illustrated functions, such as data profiling and archiving, are not detailed in this paper, the multi-domain ontology approach will impact these functions. Per example, data profiling results can constitute factual assertions to allow the ontologies to evolve with little or possibly no supervision.

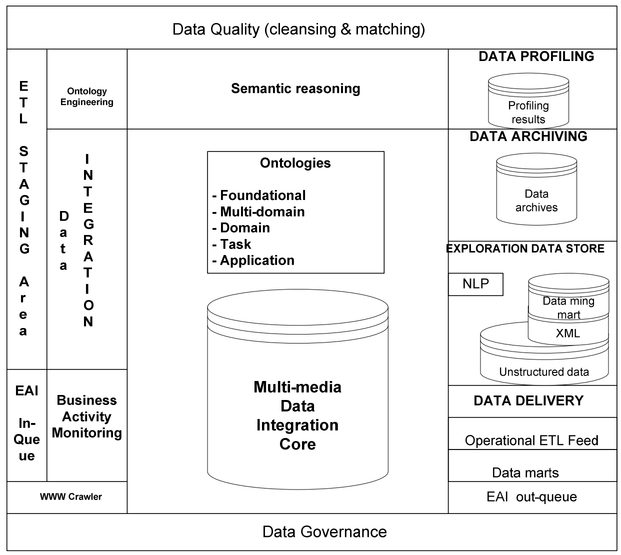


Figure I.8 Reference architecture of a semantic enterprise data warehouse

The proposed reference architecture of the semantic enterprise data warehouse could be used to design a multi-domain data integration capability, notably, to support PLM processes as defined by (Terzi et al., 2010). It would also include other MDM functions such as data quality, data profiling and data archiving, which are essential in insuring effective cross-enterprise data integration for operational and business intelligence applications. Semi-structured and unstructured data can also be extracted internally in the enterprise and externally on the web, and, be annotated with tokens allowing linking with structured data. In light of the criterion of orthogonality, figure I.9 subsumes the proposed multi-domain data integration ontology in respect with the foundational ontologies such as SUMO, Cyc, Proton and others. Domain specific ontologies such as Onto-PDM proposed by (Panetto et al., 2012) which incorporates product technical data standards STEP and IEC62264 are subsumed to the multi-domain ontology proposed in this paper. Then, the ontology structure comprises generic task ontologies, such as for natural language processing (NLP), for dealing with semi-structured and unstructured data, and for mapping heterogeneous sources to the Data Integration Core. Finally, the structure is completed with application ontologies to support domain specific tasks such as processing unstructured text from social media regarding PLM.

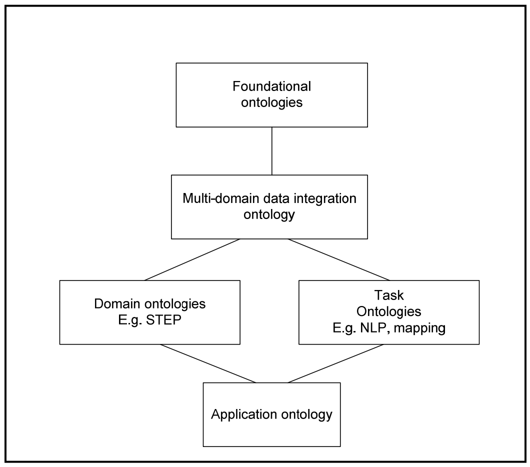


Figure I.9 Reference architecture
ontology structure

Figure I.10 identifies data domains that would compose the multi-domain data integration ontology. In its final formal form, each of these data domains, and others, would include one or more axioms that would serve as the core concepts allowing cross-enterprise interoperability to fully support PLM. Some of these data domains are already well known in the data modelling community. The Party concept was first published by (Hay, 1996) and successfully used in several enterprises and industry data models to represent customers, vendors, employees, partners, organizational structures and more. Then, data architecture patterns were also developed for the Product concept, a key concept for PLM. Through the remaining part of the research project, these artefacts will be detailed while validated by a committee of experts from the scientific and industry realms. The completion of these artefacts will be done through knowledge extraction performed using the research method described in the following section.

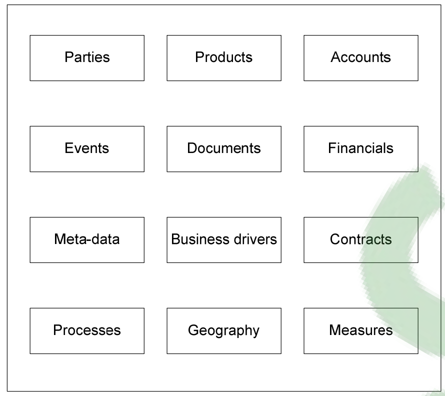


Figure I.10 Data domains for the multi-domain ontology

I.8 Contribution to theory and practice

A significant number of publications have addressed the data integration problem in the context of the semantic web, much less for the semantic enterprise. This project proposes a reference architecture and a multi-domain ontology that specifically addresses the data integration problem in the confine of an enterprise in support of its PLM.

LIST OF BIBLIOGRAPHICAL REFERENCES

Abadi, A., Ben-Azza, H., & Sekkat, S. (2016). *An ontology-based framework for virtual enterprise integration and interoperability.* Paper presented at the Electrical and Information Technologies (ICEIT), 2016 International Conference on.

Abadi, A., Ben-Azza, H., & Sekkat, S. (2017). An ontology-based support for knowledge modeling and Decision-Making in Collaborative Product Design. *International Journal of Applied Engineering Research, 12*(16), 5739-5759.

Abran, A. (2010). *Software metrics and software metrology* (A. Clements Ed.). Los Alamitos, CA USA: John Wiley & Sons.

Ahmed, S., Hacker, P., & Wallace, K. (2005). *The role of knowledge and experience in engineering design.* Paper presented at the DS 35: Proceedings ICED 05, the 15th International Conference on Engineering Design, Melbourne, Australia, 15.-18.08. 2005.

Ahmed, Z., Arif, M., Ullah, M. S., Ahmed, A., & Jabbar, M. (2016). *A Comparative Study for Ontology and Software Design Patterns.* Paper presented at the International Workshop Soft Computing Applications.

Aibdaiwi, B., Noack, R., & Thalheim, B. (2014). *Pattern-Based Conceptual Data Modelling.* Paper presented at the EJC.

Aitchison, C. (2016). How to make a great Conclusion. Retrieved from https://doctoralwriting.wordpress.com/2016/07/11/how-to-make-a-great-conclusion/

Akman, V., & Surav, M. (1997). The use of situation theory in context modeling. *Computational intelligence, 13*(3), 427-438.

Alemu, G., Stevens, B., & Ross, P. (2011). *Semantic metadata interoperability in digital libraries: a constructivist grounded theory approach*. Paper presented at the ACM/IEEE Joint Conference on Digital Libraries, Ottawa (Canada). http://eprints.rclis.org/15829/

Alexander, C. (1977). *A pattern language: towns, buildings, construction*: Oxford University Press.

Alexander, C. (1979). *The timeless way of building* (Vol. 1): New York: Oxford University Press.

Anglim, B., Milton, S. K., Rajapakse, J., & Weber, R. (2009). *Current trends and future directions in the practice of high-level data modeling: An empirical study.* Paper presented at the ECIS.

Anney, V. N. (2014). Ensuring the quality of the findings of qualitative research: Looking at trustworthiness criteria.

Anonymous. (2009). Version 9, The Open Group Architecture Framework (TOGAF) *The Open Group* (Vol. 1).

Anonymous. (2016). *Output by major industry sector*. Bureau of Labor Statistics Retrieved from https://www.bls.gov/emp/ep_table_202.htm.

Anonymous. (2018). Thesis Statements. *Handouts.* Retrieved May 6, 2018, 2018, from https://writingcenter.unc.edu/tips-and-tools/thesis-statements/

Antkiewicz, R., Chmielewski, M., Drozdowski, T., Najgebauer, A., Rulka, J., Tarapata, Z., . . . Pierzchała, D. (2012). Knowledge-Based Approach for Military Mission Planning and Simulation. In D. C. Ramirez (Ed.), *Advances in Knowledge Representation*: InTech.

Aranda-Corral, G., Borrego-Díaz, J., & Jiménez-Mavillard, A. (2010). Social Ontology Documentation for Knowledge Externalization. In S. Sánchez-Alonso & I. Athanasiadis (Eds.), *Metadata and Semantic Research* (Vol. 108, pp. 137-148): Springer Berlin Heidelberg.

Athenikos, S. J., & Song, I. Y. (2013). CAM: A conceptual modeling framework based on the analysis of entity classes and association types. *Journal of Database Management, 24*(4), 51-80.

Azizah, F. N., Bakema, G. P., Sitohang, B., & Santoso, O. S. (2009). *Generic Data Model Patterns using Fully Communication Oriented Information Modeling (FCO-IM).* Paper presented at the Electrical Engineering and Informatics, 2009. ICEEI'09. International Conference on.

Bae, I.-H. (2014). An ontology-based approach to ADL recognition in smart homes. *Future Generation Computer Systems, 33*, 32-41.

Bagheri, M., & Jahromi, M. J. G. (2016). Globalization and extraterritorial application of economic regulation: crisis in international law and balancing interests. *European Journal of Law and Economics, 41*(2), 393-429.

Bano, M., & Zowghi, D. (2013). *Users' involvement in requirements engineering and system success: a systematic literature review.* Paper presented at the Empirical Requirements Engineering (EmpiRE), 2013 IEEE Third International Workshop on.

Bano, M., Zowghi, D., & da Rimini, F. (2017). User satisfaction and system success: an empirical exploration of user involvement in software development. *Empirical Software Engineering, 22*(5), 2339-2372.

Basu, A. (2018). Semantic Web, Ontology, and Linked Data. *Information Retrieval and Management: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications*, 24.

Belay, A. M., Welo, T., & Helo, P. (2014). Approaching lean product development using system dynamics: investigating front-load effects. *Advances in Manufacturing, 2*(2), 130-140.

Bennett, T. A., & Bayrak, C. (2011). Bridging the data integration gap: from theory to implementation. *ACM SIGSOFT Software Engineering Notes, 36*(4), 1-8.

Benson, J. K. (1983). Paradigm and praxis in organizational analysis. *Research in organizational behavior*(5), 33-56.

Bergamaschi, S., Beneventano, D., Mandreoli, F., Martoglia, R., Guerra, F., Orsini, M., . . . Zhu, S. (2018). From Data Integration to Big Data Integration *A Comprehensive Guide Through the Italian Database Research Over the Last 25 Years* (pp. 43-59): Springer.

Bergholtz, M., Andersson, B., & Johannesson, P. (2010). *Abstraction, restriction, and co-creation: three perspectives on services.* Paper presented at the International Conference on Conceptual Modeling.

Bevan, M. T. (2014). A method of phenomenological interviewing. *Qualitative health research, 24*(1), 136-144.

Bharadwaj, A. (2000). Integrating positivist and interpretive approaches to information systems research: a Lakatosian model. *Foundations of Information Systems*.

Blaha, M. (2010a). Data Modeling Is Important for SOA. In J. Trujillo, G. Dobbie, H. Kangassalo, S. Hartmann, M. Kirchberg, M. Rossi, I. Reinhartz-Berger, E. Zimányi & F. Frasincar (Eds.), *Advances in Conceptual Modeling – Applications and Challenges* (Vol. 6413, pp. 255-264): Springer Berlin Heidelberg.

Blaha, M. (2010b). *Patterns of Data Modeling* (Vol. 1): CRC Press.

Blaha, M. (2013). *UML Database Modeling Workbook*: Technics Publications.

Blanco, C., Lasheras, J., Fernández-Medina, E., Valencia-García, R., & Toval, A. (2011). Basis for an integrated security ontology according to a systematic review of existing proposals. *Computer Standards & Interfaces, 33*(4), 372-388.

Blomqvist, E. (2009a). OntoCase-Automatic Ontology Enrichment Based on Ontology Design Patterns. In A. Bernstein, D. Karger, T. Heath, L. Feigenbaum, D. Maynard, E. Motta & K. Thirunarayan (Eds.), *The Semantic Web - ISWC 2009* (Vol. 5823, pp. 65-80): Springer Berlin Heidelberg.

Blomqvist, E. (2009b). *Semi-automatic ontology construction based on patterns.* Linköping University Electronic Press.

Blomqvist, E. (2010). *Ontology patterns: Typology and experiences from design pattern development.* Paper presented at the Linköping Electronic Conference Proceedings.

Borrego, M., Douglas, E. P., & Amelink, C. T. (2009). Quantitative, qualitative, and mixed research methods in engineering education. *Journal of Engineering education, 98*(1), 53-66.

Borst, W. N. (1997). Construction of engineering ontologies for knowledge sharing and reuse.

Bouten, N., Claeys, M., Mijumbi, R., Famaey, J., Latré, S., & Serrat, J. (2016). *Semantic validation of affinity constrained service function chain requests.* Paper presented at the 2016 IEEE NetSoft Conference and Workshops (NetSoft).

Bouthillier, F., & Shearer, K. (2002). Understanding knowledge management and information management: the need for an empirical perspective. *Information research, 8*(1), 8-1.

Brodie, M. L. (2010). *Data integration at scale: From relational data integration to information ecosystems.* Paper presented at the Advanced Information Networking and Applications (AINA), 2010 24th IEEE International Conference on.

Buergin, J., Belkadi, F., Hupays, C., Gupta, R. K., Bitte, F., Lanza, G., & Bernard, A. (2018). A modular-based approach for Just-In-Time Specification of customer orders in the aircraft manufacturing industry. *CIRP Journal of Manufacturing Science and Technology*.

Calhau, R. F., & de Almeida Falbo, R. (2012). *A configuration management task ontology for semantic integration.* Paper presented at the Proceedings of the 27th Annual ACM Symposium on Applied Computing.

Calvanese, D., Damaggio, E., De Giacomo, G., Lenzerini, M., & Rosati, R. (2004). Semantic data integration in P2P systems. *Databases, Information Systems, and Peer-to-Peer Computing*, 77-90.

Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., & Rosati, R. (2009). Conceptual modeling for data integration. In Springer (Ed.), *Conceptual Modeling: Foundations and Applications* (pp. 173-197): Springer.

Camossi, E., Villa, P., & Mazzola, L. (2013). Semantic-based anomalous pattern discovery in moving object trajectories. *arXiv preprint arXiv:1305.1946*.

Carlsson, C. (2018). Fuzzy Ontology Support for Knowledge Mobilisation *Frontiers in Computational Intelligence* (pp. 121-143): Springer.

Chmielewski, M. (2009). *Ontology Applications for Achieving Situation Awareness in Military Decision Support Systems.* Paper presented at the First International Conference, ICCCI 2009, Wrocław, Poland,.

Chmielewski, M., Gałka, A., Jarema, P., Krasowski, K., & Kosiński, A. (2009). *Semantic Knowledge Representation in Terrorist Threat Analysis for Crisis Management Systems.* Paper presented at the International Conference on Computational Collective Intelligence.

Codd, E. F. (1970). RELATIONAL MODEL OF DATA FOR LARGE SHARED DATA BANKS. *Communications of the ACM, 13*(Compendex), 377-387.

Collins, G., Hogan, M., Shibley, M., Williams, C., & Jovanovich, V. (2014). Data Vault and HQDM Principles. *Proceedings of the Southern Association for Information Systems, Paper, 3*.

Corry, E. J., Coakley, D., O'Donnell, J., Pauwels, P., & Keane, M. M. (2013). *The Role of Linked Data and Semantic Web in Building Operation*. Paper presented at the ICEBO - International Conference for Enhanced Building Operations, Texas, USA. http://hdl.handle.net/1969.1/151454

Creswell, J. W. (2003). Chapter 6 Research Questions and Hypotheses *Research design* (pp. 120-135): Thousand Oaks, CA: Sage.

Creswell, J. W. (2007). *Qualitative inquiry & research design: Choosing among five approaches*: Sage Publications, Inc.

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches* (Fifth ed.). Los Angeles, CA USA: Sage publications.

Currim, F., & Ram, S. (2010). *When entities are types: Effectively modeling type-instantiation relationships.* Paper presented at the International Conference on Conceptual Modeling.

Cypress, B. S. (2017). Rigor or reliability and validity in qualitative research: Perspectives, strategies, reconceptualization, and recommendations. *Dimensions of Critical Care Nursing, 36*(4), 253-263.

De Bruyn, P., Van Nuffel, D., Verelst, J., & Mannaert, H. (2012). Towards Applying Normalized Systems Theory Implications to Enterprise Process Reference Models. In A. Albani, D. Aveiro & J. Barjis (Eds.), *Advances in Enterprise Engineering VI* (Vol. 110, pp. 31-45): Springer Berlin Heidelberg.

de Farias, T. M., Roxin, A., & Nicolle, C. (2016). SWRL rule-selection methodology for ontology interoperability. *Data & Knowledge Engineering, 105*, 53-72.

De Giacomo, G., Lembo, D., Lenzerini, M., Poggi, A., & Rosati, R. (2018). Using ontologies for semantic data integration *A Comprehensive Guide Through the Italian Database Research Over the Last 25 Years* (pp. 187-202): Springer.

De Leenheer, P., Christiaens, S., & Meersman, R. (2010). Business semantics management: A case study for competency-centric HRM. *Computers in Industry, 61*(8), 760-775. doi: http://dx.doi.org/10.1016/j.compind.2010.05.005

De Toni, A. F. (2016). Ford Case Study: The Network Evolution from Extended Enterprise to Virtual Enterprise *International Operations Management* (pp. 74-95): Routledge.

Debruyne, C., & De Leenheer, P. (2013). Business Semantics as an Interface between Enterprise Information Management and the Web of Data: A Case Study in the Flemish Public Administration. In M.-A. Aufaure & E. Zimányi (Eds.), *Business Intelligence* (Vol. 138, pp. 208-233): Springer Berlin Heidelberg.

Deen, S. M., Amin, R., & Taylor, M. C. (1987). Data integration in distributed databases. *IEEE Transactions on Software Engineering*(7), 860-864.

Delfmann, P., Breuker, D., Matzner, M., & Becker, J. (2015). Supporting Information Systems Analysis Through Conceptual Model Query–The Diagramed Model Query Language (DMQL). *Communications of the Association for Information Systems, 37*.

Diaz, M. A. C., Antonelli, L., & Sanchez, L. E. (2017). Health Ontology and Information Systems: A Systematic Review. *IEEE Latin America Transactions, 15*(1), 103-120.

Dietrich, J., & Elgar, C. (2005). *A formal description of design patterns using OWL*. Paper presented at the Software Engineering Conference, 2005. Proceedings. 2005 Australian.

Dietrich, M., Lemcke, J., & Stuhec, G. (2013). Iterative Effort Reduction in B2B Schema Integration via a Canonical Data Model. *International Journal of Strategic Information Technology and Applications (IJSITA), 4*(4), 19-43.

Doan, A., Halevy, A., & Ives, Z. (2012). *Principles of data integration*: Elsevier.

Dorneich, M. C., Mott, D., Bahrami, A., Patel, J., & Giammanco, C. (2011). *Evaluation of a Shared Representation to Support Collaborative, Distributed, Coalition, Multilevel Planning*. Paper presented at the Proceedings of the Fifth Annual Conference of the International Technology Alliance.

Dreibelbis, A., Hechler, E., Milman, I., Oberhofer, M., van Run, P., & Wolfson, D. (2008). *Enterprise Master Data Management: An SOA Approach to Managing Core Information*: IBM Press.

Duygan-Bump, B., Levkov, A., & Montoriol-Garriga, J. (2015). Financing constraints and unemployment: Evidence from the Great Recession. *Journal of Monetary Economics, 75*, 89-105.

Dyché, J., & Levy, E. (2006). *Customer data integration: reaching a single version of the truth*: Wiley.

Elsby, M. W., Hobijn, B., & Sahin, A. (2010). The labor market in the Great Recession. In T. B. Institution (Ed.), *Brookings Papers on Economic Activity, Economic Studies Program* (Vol. 41, pp. 1-69): National Bureau of Economic Research.

Erl, T. (2008). *SOA: principles of service design* (Vol. 1): Prentice Hall Upper Saddle River.

Erl, T., Merson, P., & Stoffers, R. (2017). *Service-oriented Architecture: Analysis and Design for Services and Microservices*: Prentice Hall PTR.

Estublier, J., Cunin, P., Belkhatir, N., Amiour, M., & Dami, S. (1998). *Architectures for process support system interoperability.* Paper presented at the Prooceedings of the Fifth International Conference on the Software Process,(Lisle, IL).

Evans, J. H. (1959). Basic design concepts. *Naval Engineers Journal, 71*(4), 671-678.

Fitzpatrick, D. (2012). *A reference architecture for semantic EDW with multi-domain data integration capability, The IFIP WG 5.1 First Doctoral Workshop*. The IFIP WG 5.1 First Doctoral Workshop. PhD Research Plan. Montreal, Canada.

Fitzpatrick, D., Coallier, F., & Ratté, S. (2012). A Holistic Approach for the Architecture and Design of an Ontology-Based Data Integration Capability in Product Master Data Management. In L. Rivest, A. Bouras & B. Louhichi (Eds.), *Product Lifecycle Management. Towards Knowledge-Rich Enterprises* (Vol. 388, pp. 559-568): Springer Berlin Heidelberg.

Fitzpatrick, D., Coallier, F., & Ratté, S. (2012). *A holistic approach for the architecture and design of an ontology-based data integration capability in product master data management*. Paper presented at the 9th International Conference on Product Lifecycle Management, Montreal, QC, Canada.

Fitzpatrick, D., Coallier, F., & Ratté, S. (2013). *A Reference Architecture for an Enterprise Knowledge Infrastructure.* Paper presented at the PLM.

Fitzpatrick, D., Coallier, F., & Ratté, S. (2018). A use case of a multi-domain ontology for collaborative logistics planning in coalition force deployment. *Manuscript submitted for publication*.

Fitzpatrick, D., Ratté, S., & Coallier, F. (2013, 7-9 Oct. 2013). *RA-EKI: A use case for collaborative logistics planning in coalition force deployment.* Paper presented at the Military Communications and Information Systems Conference (MCC), 2013.

Fitzpatrick, D., Ratté, S., & Coallier, F. (2018a). Agnostic content ontology design patterns for enterprise semantic interoperability: a Systematic Literature Review. *Manuscript submitted for publication*.

Fitzpatrick, D., Ratté, S., & Coallier, F. (2018b). A dual method qualitative research design for eliciting agnostic content ontology design patterns for a multi-domain ontology. *Manuscript submitted for publication*.

Fitzpatrick, D., Ratté, S., & Coallier, F. (2018c). Eliciting agnostic content ontology design patterns for enterprise semantic interoperability using a phenomenological research method. *Manuscript submitted for publication*.

Fitzpatrick, D., Ratté, S., & Coallier, F. (2018d). A use case of a multi-domain ontology for collaborative product design. *Manuscript submitted for publication*.

Flynn, S. V., & Korcuska, J. S. (2018). Credible Phenomenological Research: A Mixed-Methods Study. *Counselor Education and Supervision, 57*(1), 34-50.

Fokoue, A., Srivatsa, M., Rohatgi, P., Wrobel, P., & Yesberg, J. (2009). *A decision support system for secure information sharing.* Paper presented at the Proceedings of the 14th ACM symposium on Access control models and technologies.

Ford, R., Martin, D., Elenius, D., & Johnson, M. (2011). Ontologies and tools for analysing and composing simulation confederations for the training and testing domains. *Journal of Simulation (2011), 5*, 230–245.

Forero, R., Nahidi, S., De Costa, J., Mohsin, M., Fitzgerald, G., Gibson, N., . . . Aboagye-Sarfo, P. (2018). Application of four-dimension criteria to assess rigour of qualitative research in emergency medicine. *BMC health services research, 18*(1), 120.

Fortineau, V. (2013). *Contribution à une modélisation ontologique des informations tout au long du cycle de vie du produit.* Paris, ENSAM.

Frosch-Wilke, D., & Scheffler, L. (2015). *Integrating Crime Data by the Use of Generic Data Models*. Paper presented at the ICCGI 2015 : The Tenth International Multi-Conference on Computing in the Global Information Technology.

Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1993). *Design patterns: Abstraction and reuse of object-oriented design.* Paper presented at the European Conference on Object-Oriented Programming.

Gangemi, A., Gómez-Pérez, A., Presutti, V., & Suárez-Figueroa, M. C. (2007). Towards a catalog of owl-based ontology design patterns.

Gangemi, A., & Presutti, V. (2009). Ontology Design Patterns. In S. Staab & R. Studer (Eds.), *Handbook on Ontologies* (pp. 221-243): Springer Berlin Heidelberg.

Gharib, M., Giorgini, P., & Mylopoulos, J. (2017). *Towards an Ontology for Privacy Requirements via a Systematic Literature Review.* Paper presented at the International Conference on Conceptual Modeling.

Giaretta, P., & Guarino, N. (1995). Ontologies and knowledge bases towards a terminological clarification. *Towards very large knowledge bases: knowledge building & knowledge sharing, 25*, 32.

Giraldo, F. D., España, S., Pineda, M. A., Giraldo, W. J., & Pastor, O. (2014). *Conciliating model-driven engineering with technical debt using a quality framework.* Paper presented at the Forum at the Conference on Advanced Information Systems Engineering (CAiSE).

Glöckner, M., & Ludwig, A. (2017, August 23–26, 2017). *Ontological structuring of logistics services.* Paper presented at the Proceedings of the International Conference on Web Intelligence 2017, Leipzig, Germany,.

Gómez-Pérez, A., Fernández-López, M., & Corcho, O. (2004). *Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*: Springer Verlag.

Gómez-Pérez, A., Fernández-López, M., & Corcho, O. (2006). *Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*: Springer Science & Business Media.

González, J., de Castro, P., & Güemes, C. (2011, 6-9 June 2011). *New Information and Communication Technology cutting edge solutions for marine conditions prediction and logistics processes management in offshore renewable energy infrastructures installation and maintenance.* Paper presented at the OCEANS 2011, Santander, Spain.

González, L., Echevarría, A., Morales, D., & Ruggia, R. (2016). An E-government Interoperability Platform Supporting Personal Data Protection Regulations. *CLEI Electronic Journal, 19*(2), 8-8.

Grant, T., & van den Heuvel, G. (2010). *Modelling the information sharing process in military coalitions: A work in progress.* Paper presented at the Proceedings of the 7th International ISCRAM Conference.

Gregor, S. (2006). The nature of theory in information systems. *Management Information Systems Quarterly, 30*(3), 611.

Gregor, S. (2009). *Building theory in the sciences of the artificial.* Paper presented at the 4th International Conference on Design Science Research in Information Systems and Technology, DESRIST '09, May 7, 2009 - May 8, 2009, Philadelphia, CA, United states.

Gregor, S. (2017). On theory *The Routledge Companion to Management Information Systems* (pp. 77-92): Routledge.

Group, T. O. (2009). TOGAF Version 9 The Open Group Architecture Framework (pp. 744): The Open Group.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition, 5*(2), 199-220. doi: https://doi.org/10.1006/knac.1993.1008

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition, 5*(Copyright 1993, IEE), 199-220.

Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies, 43*(Copyright 1996, IEE), 907-928.

Gruber, T. R., Liu, L., & Ozsu, M. T. (2009). Ontology *Encyclopedia of database systems* (1st edition ed., pp. 3752): Springer Publishing Company, Incorporated.

Guarino, N. (1998, 6-8 June 1998). *Formal ontology and information systems.* Paper presented at the Proceedings of Formal Ontology in Information Systems, Amsterdam, Netherlands.

Guarino, N. (1998). *Formal ontology and information systems.* Paper presented at the Proceedings of Formal Ontology in Information Systems, 6-8 June 1998, Amsterdam, Netherlands.

Guarino, N., Oberle, D., & Staab, S. (2009). What is an ontology? *Handbook on ontologies* (pp. 1-17): Springer.

Guba, E., & Lincoln, Y. (2001). Guidelines and Checklist for Constructivist. *AKA Fourth Generation), evaluation paper available, Kalamazoo, MI: Evaluation Centre*.

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field methods, 18*(1), 59-82.

Halevy, A., Rajaraman, A., & Ordille, J. (2006). *Data integration: The teenage years*.

Hall, S. (2016). How Do You Know IT Costs Too Much? *CFO*. http://ww2.cfo.com/it-value/2016/02/know-costs-much/

Hammar, K., & Sandkuhl, K. (2010). *The state of ontology pattern research a systematic review of ISWC, ESWC and ASWC 2005-2009.* Paper presented at the CEUR Workshop Proceedings.

Hay, D. (1996). *Data model patterns: conventions of thought*: Addison-Wesley.

Hays, D. G., & Wood, C. (2011). Infusing qualitative traditions in counseling research designs. *Journal of Counseling & Development, 89*(3), 288-295.

Haziti, M., Qadi, A., Bazzi, M., & Elhassouni, J. (2018). Applying ontologies to data integration systems for bank credit risk management. *Journal of Data Mining & Digital Humanities*.

Henderson-Sellers, B., Low, G., & Gonzalez-Perez, C. (2012). Semiotic Considerations for the Design of an Agent-Oriented Modelling Language. In I. Bider, T. Halpin, J. Krogstie, S. Nurcan, E. Proper, R. Schmidt, P. Soffer & S. Wrycza (Eds.), *Enterprise, Business-Process and Information Systems Modeling* (Vol. 113, pp. 422-434): Springer Berlin Heidelberg.

Héon, M. (2010). *OntoCASE: méthodologie et assistant logiciel pour une ingénierie ontologique fondée sur la transformation d'un modèle semi-formel.* Télé-université du Québec à Montréal. Retrieved from http://r-libre.teluq.ca/616/1/Heon.pdf

Hess, D. R. (2004). How to write an effective discussion. *Respiratory care, 49*(10), 1238-1241.

Hitzler, P., & Shimizu, C. (2018). *Modular Ontologies as a Bridge Between Human Conceptualization and Data.* Paper presented at the International Conference on Conceptual Structures.

Hofman, W., & Rajagopal, M. (2015). *Interoperability in self-organizing systems of multiple enterprises–a case on improving turnaround time prediction at logistics hubs.* Paper presented at the Zelm M., 6th Workshops of the IWEI 2015 Conference, IWEI-WS 2015-co-located with the 6th International IFIP Working Conference on Enterprise Interoperability IWEI 2015; 27 May 2015, Nimes, France.

Hofreiter, B., Huemer, C., Kappel, G., Mayrhofer, D., & vom Brocke, J. (2012). Inter-organizational Reference Models – May Inter-organizational Systems Profit from Reference Modeling? In C. Ardagna, E. Damiani, L. Maciaszek, M. Missikoff & M. Parkin (Eds.), *Business System Management and Engineering* (Vol. 7350, pp. 32-47): Springer Berlin Heidelberg.

Horridge, M., & Bechhofer, S. (2011). The owl api: A java api for owl ontologies. *Semantic Web, 2*(1), 11-21.

Hsu, I. C., & Cheng, F. Q. (2015). SAaaS: a cloud computing service model using semantic-based agent. *Expert Systems, 32*(1), 77-93.

Husserl, E. (1970). *The crisis of European sciences and transcendental phenomenology: An introduction to phenomenological philosophy*: Northwestern University Press.

Hycner, R. H. (1985). Some guidelines for the phenomenological analysis of interview data. *Human studies, 8*(3), 279-303.

Introna, L. (2005). Phenomenological approaches to ethics and information technology. *Stanford Encyclopedia of Philosophy*.

Jenicek, M. (2006). How to read, understand, and write'Discussion'sections in medical articles. An exercise in critical thinking. *Medical Science Monitor, 12*(6), SR28-SR36.

Jhingran, A., Mattos, N., & Pirahesh, H. (2002). Information integration: A research agenda. *IBM Systems Journal, 41*(4), 555-562.

Jiang, L., Cai, H., & Xu, B. (2010). *A domain ontology approach in the ETL process of data warehousing.* Paper presented at the IEEE International Conference on E-Business Engineering, ICEBE 2010, November 10, 2010 - November 12, 2010, Shanghai, China.

Jinxin, S., Cungen, C., Haitao, W., Fang, G., Qiangze, F., Chunxia, Z., . . . Yufei, Z. (2002). *An environment for multi-domain ontology development and knowledge acquisition.* Paper presented at the Engineering and Deployment of Cooperative Information Systems. First International Conference, EDCIS 2002. Proceedings, 17-20 Sept. 2002, Berlin, Germany.

Jirkovský, V., Obitko, M., & Mařík, V. (2017). Understanding data heterogeneity in the context of cyber-physical systems integration. *IEEE Transactions on Industrial Informatics, 13*(2), 660-667.

Johnson, P., Ekstedt, M., & Jacobson, I. (2012). Where's the theory for software engineering? *IEEE software, 29*(5), 96-96.

Jovanovic, V., & Bojicic, I. (2012). Conceptual Data Vault Model.

Jovanovic, V., & Pavlic, M. (2011). *Data modeling patterns—Taxonomy.* Paper presented at the MIPRO, 2011 Proceedings of the 34th International Convention.

Jovanovic, V., Subotic, D., & Mrdalj, S. (2014). *Data modeling styles in data warehousing.* Paper presented at the Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014 37th International Convention on.

Kastner, P., & Saia, R. (2006). The composite applications benchmark report *Dec-2006*: Aberdeen.

Katsumi, M., & Fox, M. (2018). Ontologies for transportation research: A survey. *Transportation Research Part C: Emerging Technologies, 89*, 53-82.

Kelly, J. E. (2015). Computing, cognition and the future of knowing. *Whitepaper, IBM Reseach*, 2.

Kerga, E., Schmid, R., Rebentisch, E., & Terzi, S. (2016). *Modeling the benefits of frontloading and knowledge reuse in lean product development.* Paper presented at the Management of Engineering and Technology (PICMET), 2016 Portland International Conference on.

Khedher, A., Henry, S., & Bouras, A. (2012, July 2-4, 2012). *Quality improvement of product data exchanged between engineering and production through the integration of dedicated information systems.* Paper presented at the 11th Biennial Conference On Engineering Systems Design And Analysis, Nantes, France.

Khouri, S., & Bellatreche, L. (2010). *A methodology and tool for conceptual designing a data warehouse from ontology-based sources.* Paper presented at the 13th ACM International Workshop on Data Warehousing and OLAP, DOLAP'10, Co-located with 19th International Conference on Information and Knowledge Management, CIKM'10, October 26, 2010 - October 30, 2010, Toronto, ON, Canada.

Khouri, S., Bellatreche, L., & Marcel, P. (2011). *Embedding user's requirements in data warehouse repositories.* Paper presented at the OTM Confederated International Conferences" On the Move to Meaningful Internet Systems".

Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University, 33*(2004), 1-26.

Knowles, C., & Jovanovic, V. (2013). Extensible markup language (xml) schemas for data vault models. *Journal of Computer Information Systems, 53*(4), 12-21.

Kozmina, N., Syundyukov, E., & Kozmins, A. (2017). *Data Modelling for Dynamic Monitoring of Vital Signs: Challenges and Perspectives.* Paper presented at the International Conference on Conceptual Modeling.

Kuster, E. (2007). *Coalition interoperability architecture.* Paper presented at the Integration of Knowledge Intensive Multi-Agent Systems, 2007. KIMAS 2007. International Conference on.

Kuster, E. (2007). *Coalition Interoperability Architecture.* Paper presented at the KIMAS 2007, Waltham, MA, USA.

Lacy, L. W. (2005). *OWL: Representing information using the web ontology language*. Victoria, BC Canada: Trafford Publishing.

Laínez, J. M., Schaefer, E., & Reklaitis, G. V. (2012). Challenges and opportunities in enterprise-wide optimization in the pharmaceutical industry. *Computers & Chemical Engineering, 47*, 19-28.

Lankhorst, M. M., Proper, H. A., & Jonkers, H. (2009). The architecture of the archimate language *Enterprise, Business-Process and Information Systems Modeling* (pp. 367-380): Springer.

Laurier, W., & Poels, G. (2012). Ontology-based structuring of conceptual data modeling patterns. *Journal of Database Management, 23*(3), 50-64.

Leal, D. (2005). ISO 15926" Life cycle data for process plant": An overview. *Oil & gas science and technology, 60*(4), 629-637.

Lee, M., Matentzoglu, N., Sattler, U., & Parsia, B. (2015). *Verifying reasoner correctness-a justification based method.* Paper presented at the Informal Proceedings of the 4th International Workshop on OWL Reasoner Evaluation (ORE-2015).

Leedy, P., & Ormrod, J. (2012). Practical research, Planning and Design (10th ed., pp. 336). Boston, MA:: Pearson.

Leedy, P. D., & Ormrod, J. E. (2005). *Practical research: Planning and design*: Pearson/Merrill/Prentice Hall, Upper Saddle River, NJ.

Lemcke, J. (2009). Light-weight semantic integration of generic behavioral component descriptions. *Semantic Enterprise Application Integration for Business Processes*, 131-171.

Lenz, R., Peleg, M., & Reichert, M. (2012). Healthcare process support: achievements, challenges, current research. *International Journal of Knowledge-Based Organizations (IJKBO), 2*(4).

Lenzerini, M. (2002). *Data integration: A theoretical perspective*.

Lieto, A., Lebiere, C., & Oltramari, A. (2018). The knowledge level in cognitive architectures: Current limitations and possible developments. *Cognitive Systems Research, 48*, 39-55.

Liew, A. (2007). Understanding data, information, knowledge and their inter-relationships. *Journal of Knowledge Management Practice, 8*(2).

Loser, C., Legner, C., & Gizanis, D. (2004). *Master data management for collaborative service processes*.

Lu, Y., Panetto, H., Ni, Y., & Gu, X. (2013). Ontology Alignment for Networked Enterprises Information Systems Interoperability in Supply Chain Environment. *International Journal of Computer Integrated Manufacturing, 26*(1-2), 140-151.

Lubyansky, A. (2009). Using Data Model Patterns to Build High-Quality Data Models.

Luttighuis, P. O., Stap, R., & Quartel, D. (2011). *Contexts for concepts: Information modeling for semantic interoperability.* Paper presented at the International IFIP Working Conference on Enterprise Interoperability.

Maier, M. W., & Rechtin, E. (2009). *The art of systems architecting* (Third Edition ed.). Boca Raton, FL USA: CRC Press, Taylor and Francis Group.

Malan, R., & Bredemeyer, D. (2002). Less is more with minimalist architecture. *IT professional, 4*(5), 48-47.

Mamayev, R. (2014). *Data Modeling of Financial Derivatives: A Conceptual Approach*: Apress.

Marchetta, M., Mayer, F., & Forradellas, R. (2011). A reference framework following a proactive approach for Product Lifecycle Management. *Computers in Industry, 62*(7), 672–683.

Marrakchi, K., Briache, A., Kerzazi, A., Navas-Delgado, I., Aldana-Montes, J. F., Ettayebi, M., . . . Rossi Hassani, B. D. (2010). *A data warehouse approach to semantic integration of pseudomonas data.* Paper presented at the 7th International Conference on Data Integration in the Life Sciences, DILS 2010, August 25, 2010 - August 27, 2010, Gothenburg, Sweden.

Marshall, B., Cardon, P., Poddar, A., & Fontenot, R. (2013). Does sample size matter in qualitative research?: A review of qualitative interviews in IS research. *Journal of Computer Information Systems, 54*(1), 11-22.

Matsokis, A., & Kiritsis, D. (2010). An ontology-based approach for Product Lifecycle Management. *Computers in Industry, 61*(8), 787-797. doi: 10.1016/j.compind.2010.05.007

McGuinness, D. L., & Da Silva, P. P. (2004). Explaining answers from the semantic web: The inference web approach. *Web Semantics: Science, Services and Agents on the World Wide Web, 1*(4), 397-413.

McInerney, C. (2002). Knowledge management and the dynamic nature of knowledge. *Journal of the American Society for Information Science and Technology, 53*(12), 1009-1018.

Michaels, S., Goucher, N. P., & McCarthy, D. (2006). Considering knowledge uptake within a cycle of transforming data, information, and knowledge. *Review of Policy Research, 23*(1), 267-279.

Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM, 38*(11), 39-41.

Mirhaji, P., Zhu, M., Vagnoni, M., Bernstam, E. V., Zhang, J., & Smith, J. W. (2009). *Ontology driven integration platform for clinical and translational research.* Paper presented at the BMC bioinformatics.

Molnár, B., & Benczúr, A. (2015). Modeling information systems from the viewpoint of active documents. *Vietnam Journal of Computer Science, 2*(4), 229-241.

Morosoff, P., Rudnicki, R., Bryant, J., Farrell, R., & Smith, B. (2015). Joint Doctrine Ontology: A Benchmark for Military Information Systems Interoperability. *Semantic Technology for Intelligence, Defense and Security (STIDS), 1325*.

Moustakas, C. (1994). *Phenomenological research methods*. Thousand Oaks, CA USA: Sage Publications.

Moustakas, C. E. (1994). *Phenomenological research methods*: Sage Publications, Inc.

Mulrow, C. D. (1994). Systematic reviews: rationale for systematic reviews. *British Medical Journal, 309*(6954), 597-599.

Navigli, R., & Velardi, P. (2008). *From glossaries to ontologies: Extracting semantic structure from textual definitions*.

Nazri, M. N. M., Noah, S. A., & Hamid, Z. (2010). *Using lexical ontology for semi-automatic logical data warehouse design.* Paper presented at the 5th International Conference on Rough Set and Knowledge Technology, RSKT 2010, October 15, 2010 - October 17, 2010, Beijing, China.

Noy, N. F., & McGuinness, D. L. (2001). Ontology development 101: A guide to creating your first ontology.

Obrst, L., Chase, P., & Markeloff, R. (2012). Developing an Ontology of the Cyber Security Domain. *Semantic Technology for Intelligence, Defense, and Security 2012*, 49-56.

Okoli, C. (2015). A guide to conducting a standalone systematic literature review. *Communications of the Association for Information Systems, 37*, 879-910.

Okoli, C., & Schabram, K. (2010). A guide to conducting a systematic literature review of information systems research.

Olivé, A. (2017). *The Universal Ontology: A Vision for Conceptual Modeling and the Semantic Web.* Paper presented at the International Conference on Conceptual Modeling.

Olivé, A. (2018). A Universal Ontology-based Approach to Data Integration. *Enterprise Modelling and Information Systems Architectures, 13*, 110-119.

Orlikowski, W. J., & Baroudi, J. J. (1991). Studying information technology in organizations: Research approaches and assumptions. *Information systems research, 2*(1), 1-28.

Pai, F.-P., Yang, L.-J., & Chung, Y.-C. (2017). Multi-layer ontology based information fusion for situation awareness. *Applied intelligence, 46*(2), 285-307.

Panetto, H., Dassisti, M., & Tursi, A. (2012). ONTO-PDM: Product-driven ONTOlogy for Product Data Management interoperability within manufacturing process environment. *Advanced Engineering Informatics, 26*(2), 334-348. doi: 10.1016/j.aei.2011.12.002

Patel, J., Dorneich, M., Mott, D., Bahrami, A., & Giammanco, C. (2010). A Conceptual Framework to Support a Multi-level Planning Capability. In A. R. L. A. P. G. MD (Ed.).

Patel, J., Dorneich, M. C., Mott, D., Bahrami, A., & Giammanco, C. (2010). A Conceptual Framework to Support a Multi-level Planning Capability: Army Research Lab Aberdeen Proving Ground, MD.

Patton, M. Q. (2002). *Qualitative research and evaluation methods*: Sage.

Perry, N., Bernard, A., Bosch-Mauchand, M., LeDuigou, J., & Xu, Y. (2011). Eco global evaluation: cross benefits of economic and ecological evaluation *Glocalized Solutions for Sustainability in Manufacturing* (pp. 681-686): Springer.

Pfeiffer, R.-H., & Wąsowski, A. (2011). *Taming the confusion of languages.* Paper presented at the European Conference on Modelling Foundations and Applications.

Piho, G., Roost, M., Perkins, D., & Tepandi, J. (2010). Towards Archetypes-Based Software Development. In T. Sobh & K. Elleithy (Eds.), *Innovations in Computing Sciences and Software Engineering* (pp. 561-566): Springer Netherlands.

Piho, G., & Tepandi, J. (2013) Business domain modelling with business archetypes and archetype patterns. *Vol. 251. Frontiers in Artificial Intelligence and Applications* (pp. 221-240).

Piho, G., Tepandi, J., & Parman, M. (2012). *Towards LIMS (Laboratory Information Management Systems) software in global context.* Paper presented at the MIPRO 2012 - 35th International Convention on Information and Communication Technology, Electronics and Microelectronics - Proceedings.

Piho, G., Tepandi, J., Parman, M., & Perkins, D. (2010). *From archetypes-based domain model of clinical laboratory to LIMS software.* Paper presented at the MIPRO 2010 - 33rd International Convention on Information and Communication Technology, Electronics and Microelectronics, Proceedings.

Pinkel, C., Binnig, C., Jiménez-Ruiz, E., May, W., Ritze, D., Skjæveland, M. G., . . . Kharlamov, E. (2015). *RODI: A benchmark for automatic mapping generation in relational-to-ontology data integration.* Paper presented at the European Semantic Web Conference.

Poels, G., Maes, A., Gailly, F., & Paemeleire, R. (2011). The pragmatic quality of Resources-Events-Agents diagrams: An experimental evaluation. *Information Systems Journal, 21*(1), 63-89.

Pohl, K., & Morosoff, P. (2011, 2 Aug, 2011). *ICODES: A Load-Planning System that Demonstrates the Value of Ontologies in the Realm of Logistical Command and Control (C2).* Paper presented at the InterSymp-2011, Baden-Baden, Germany.

Poveda, M., Suárez-Figueroa, M. C., & Gómez-Pérez, A. (2009). *Common pitfalls in ontology development.* Paper presented at the Conference of the Spanish Association for Artificial Intelligence.

Pratt, M. J. (2005). ISO 10303, the STEP standard for product data exchange, and its PLM capabilities. *International Journal of Product Lifecycle Management, 1*(1), 86-94.

President, E. o. o. t. (2017). *North American Industry Classification System*. Washington, DC USA: Office of Management Budget Retrieved from census.gov/naics.

Ptitsyn, P. S., Radko, D. V., & Lankin, O. V. (2016). Designing architecture of software framework for building security infrastructure of global distributed computing systems. *ARPN Journal of Engineering and Applied Sciences, 11*(19), 11599-11610.

Puonti, M., Raitalaakso, T., Aho, T., & Mikkonen, T. (2016). *Automating Transformations in Data Vault Data Warehouse Loads*. Paper presented at the EJC.

Rattanasawad, T., Buranarach, M., Saikaew, K. R., & Supnithi, T. (2018). A Comparative Study of Rule-Based Inference Engines for the Semantic Web. *IEICE TRANSACTIONS on Information and Systems, 101*(1), 82-89.

Ratté, S., Njomgue, W., & Ménard, P. A. (2007). *Highlighting document's structure.* Paper presented at the World Academy of Science, Engineering and Technology.

Roberts, D., Lock, G., & Verma, D. C. (2007). *Holistan: A futuristic scenario for international coalition operations.* Paper presented at the Integration of Knowledge Intensive Multi-Agent Systems, 2007. KIMAS 2007. International Conference on.

Rosenthal, A., Seligman, L., Renner, S., & Manola, F. (2001). *Data integration needs an industrial revolution.* Paper presented at the International Workshop on Foundations of Models for Information Integration (FMII-2001).

Ruan, T., Xue, L., Wang, H., Hu, F., Zhao, L., & Ding, J. (2016). *Building and Exploring an Enterprise Knowledge Graph for Investment Analysis.* Paper presented at the International Semantic Web Conference.

Ruy, F. B., Reginato, C. C., Santos, V. A., Falbo, R. A., & Guizzardi, G. (2015). *Ontology engineering by combining ontology patterns.* Paper presented at the International Conference on Conceptual Modeling.

Sajja, P. S. (2008). Multi-agent system for knowledge-based access to distributed databases. *Interdisciplinary Journal of Information, Knowledge, and Management, 3*, 1-9.

Salguero, A., Araque, F., & Delgado, C. (2008). Ontology based framework for data integration. *WSEAS Transactions on Information Science and Applications, 5*(6), 953-962.

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., . . . Jinks, C. (2017). Saturation in qualitative research: exploring its conceptualization and operationalization. *Quality & Quantity*, 1-15.

Serbanescu, V., Azadbakht, K., Boer, F., Nagarajagowda, C., & Nobakht, B. (2016). A design pattern for optimizations in data intensive applications using ABS and JAVA 8. *Concurrency and Computation: Practice and Experience, 28*(2), 374-385.

Sesera, L. (2011). *Applying fundamental banking patterns: Stories and pattern sequences.* Paper presented at the ACM International Conference Proceeding Series.

Setiawan, F. A., Budiardjo, E. K., Basaruddin, T., & Aminah, S. (2017). *A Systematic Literature Review on Combining Ontology with Bayesian Network to Support Logical and Probabilistic Reasoning.* Paper presented at the Proceedings of the 2017 International Conference on Software and e-Business.

Silverston, L., & Agnew, P. (2011). *The Data Model Resource Book: Volume 3: Universal Patterns for Data Modeling* (Vol. 3): John Wiley & Sons.

Sim, J., Saunders, B., Waterfield, J., & Kingstone, T. (2018). Can sample size in qualitative research be determined a priori? *International Journal of Social Research Methodology*, 1-16.

Simsion, G., Milton, S. K., & Shanks, G. (2012). Data modeling: Description or design? *Information & Management, 49*(3-4), 151-163.

Singer, D. J., Doerry, N., & Buckley, M. E. (2009). What Is Set-Based Design? *Naval Engineers Journal, 121*(4), 31-43.

Smart, P. R., Mott, D., Gentle, E., Braines, D., Sieck, W., Poltrock, S., . . . Strub, M. (2008). Holistan revisited: Demonstrating agent-and knowledge-based capabilities for future coalition military operations.

Smith, B. (2008). Ontology (science). *Nature Precedings*.

Spyns, P., Meersman, R., & Jarrar, M. (2002). Data modelling versus ontology engineering. *ACM SIGMOD Record, 31*(4), 12-17.

Starks, H., & Brown Trinidad, S. (2007). Choose your method: A comparison of phenomenology, discourse analysis, and grounded theory. *Qualitative health research, 17*(10), 1372-1380.

Stirna, J., & Sandkuhl, K. (2014). *An outlook on patterns as an aid for business and IT alignment with capabilities.* Paper presented at the International Conference on Advanced Information Systems Engineering.

Stol, K.-J., Ralph, P., & Fitzgerald, B. (2016). *Grounded theory in software engineering research: a critical review and guidelines.* Paper presented at the Software Engineering (ICSE), 2016 IEEE/ACM 38th International Conference on.

Su, X., Li, P., Riekki, J., Liu, X., Kiljander, J., Soininen, J.-P., . . . Li, Y. (2018). *Distribution of Semantic Reasoning on the Edge of Internet of Things.* Paper presented at the 2018 IEEE International Conference on Pervasive Computing and Communications (PerCom).

Subbaraj, R., & Venkatraman, N. (2015). *A systematic literature review on ontology based context management system.* Paper presented at the Emerging ICT for Bridging the Future-Proceedings of the 49th Annual Convention of the Computer Society of India CSI Volume 2.

Suri, H. (2011). Purposeful sampling in qualitative research synthesis. *Qualitative Research Journal, 11*(2), 63-75.

Tennis, J. T. (2003). Two axes of domains for domain analysis.

Terkaj, W., Pedrielli, G., & Sacco, M. (2011, July 9-14, 2011). *Virtual Factory Data Model.* Paper presented at the Virtual and Mixed Reality - Systems and Applications, Orlando, FL, USA.

Terkaj, W., Pedrielli, G., & Sacco, M. (2012). *Virtual factory data model.* Paper presented at the CEUR Workshop Proceedings.

Terzi, S., Bouras, A., Dutta, D., Garetti, M., & Kiritsis, D. (2010). Product lifecycle management - from its history to its new role. *International Journal of Product Lifecycle Management, 4*(4), 360-389. doi: 10.1504/ijplm.2010.036489

Tesch, R. (1990). *Qualitative research: Analysis types and software tools*: Routledge.

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC medical research methodology, 8*(1), 45.

Tiwari, V., & Thakur, R. S. (2015). Contextual snowflake modelling for pattern warehouse logical design. *Sadhana, 40*(1), 15-33.

Ullman, J. (1997). Information integration using logical views. *Database Theory—ICDT'97*, 19-40.

Vaisman, A., & Zimányi, E. (2012). Data Warehouses: Next Challenges. *Business Intelligence*, 1-26.

Van Grootel, G., Spyns, P., Christiaens, S., & Jörg, B. (2009). *Business semantics management supports government innovation information portal.* Paper presented at the OTM Confederated International Conferences" On the Move to Meaningful Internet Systems".

Verdonck, M., Gailly, F., de Cesare, S., & Poels, G. (2015). Ontology-driven conceptual modeling: A systematic literature mapping and review. *Applied Ontology, 10*(3-4), 197-227.

Villanueva Chavez, J., & Li, X. (2011). *Ontology based ETL process for creation of ontological data warehouse.* Paper presented at the 2011 8th International Conference on Electrical Engineering, Computing Science and Automatic Control, CCE 2011, October 26, 2011 - October 28, 2011, Merida, Yucatan, Mexico.

Völker, J., Haase, P., & Hitzler, P. (2008). *Learning expressive ontologies*.

Wannous, R. (2014). *Computational inference of conceptual trajectory model: considering domain temporal and spatial dimensions.* Université de La Rochelle.

Ward, A., Liker, J. K., Cristiano, J. J., & Sobek, D. K. (1995). The second Toyota paradox: How delaying decisions can make better cars faster. *Sloan management review, 36*(3), 43.

Welty, C. (2003). Ontology research. *AI magazine, 24*(3), 11.

Wertz, F. J. (2005). Phenomenological research methods for counseling psychology. *Journal of counseling psychology, 52*(2), 167.

West, M. (2009). Ontology Meets Business - Applying Ontology to the Development of Business Information Systems. In A. Tolk & L. Jain (Eds.), *Complex Systems in Knowledge-based Environments: Theory, Models and Applications* (Vol. 168, pp. 229-260): Springer Berlin Heidelberg.

West, M. (2011). *Developing high quality data models*. Burlington, MA USA: Morgan Kaufmann, Elsevier.

Williams, A. J., Harland, L., Groth, P., Pettifer, S., Chichester, C., Willighagen, E. L., . . . Goble, C. (2012). Open PHACTS: semantic interoperability for drug discovery. *Drug discovery today, 17*(21-22), 1188-1198.

Wimalasuriya, D. C., & Dou, D. (2010). Ontology-based information extraction: An introduction and a survey of current approaches. *Journal of Information Science, 36*(3), 306-323.

Wohlin, C., & Aurum, A. (2015). Towards a decision-making structure for selecting a research design in empirical software engineering. *Empirical Software Engineering, 20*(6), 1427-1455.

Womack, J. P., Jones, D. T., & Roos, D. (1990). *Machine that changed the world*: Simon and Schuster.

Wu, Z., Eadon, G., Das, S., Chong, E. I., Kolovski, V., Annamalai, M., & Srinivasan, J. (2008). *Implementing an inference engine for RDFS/OWL constructs and user-defined rules in Oracle.* Paper presented at the Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on.

Xi, X., & Hongfeng, X. (2009). *Developing a framework for business intelligence systems integration based on ontology.* Paper presented at the Networking and Digital Society, 2009. ICNDS'09. International Conference on.

Xie, H., & Shen, W. (2006). *Ontology as a mechanism for application integration and knowledge sharing in collaborative design: A review.* Paper presented at the 2006 10th International Conference on Computer Supported Cooperative Work in Design, CSCWD 2006, May 3, 2006 - May 5, 2006, Nanjing, China.

Zhao, Y., Liu, Q., Xu, W., Wu, X., Jiang, X., Zhou, Z., & Pham, D. T. (2017). Dynamic and unified modelling of sustainable manufacturing capability for industrial robots in cloud manufacturing. *The International Journal of Advanced Manufacturing Technology, 93*(5-8), 2753-2771.

Zong, N., Nam, S., Eom, J.-H., Ahn, J., Joe, H., & Kim, H.-G. (2015). Aligning ontologies with subsumption and equivalence relations in Linked Data. *Knowledge-Based Systems, 76*, 30-41.

Zowghi, D., da Rimini, F., & Bano, M. (2015). *Problems and challenges of user involvement in software development: an empirical study.* Paper presented at the Proceedings of the 19th International Conference on Evaluation and Assessment in Software Engineering.

Zuanelli, E. (2017). *The cybersecurity ontology platform: the POC solution*. Paper presented at the e-AGE2017 The 7th International Platform on Integrating Arab e-Infrastructure in a Global Environment, Cairo, Egypt.