ACLRO: AN ONTOLOGY FOR THE BEST PRACTICE IN ACLR REHABILITATION

Kanitha Phalakornkule

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Doctor of Philosophy
in the School of Informatics and Computing,
Indiana University

October 2020

Accepted by the Graduate Faculty of Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Doctoral Committee

____
Josette F. Jones, RN, Ph.D., Chair

____
Ben Boukai, Ph.D.

April 6, 2020

____
Xiaowen Liu, Ph.D.

____
Saptarshi Purkayatha, Ph.D.

____
William D. Duncan, MS, Ph.D.

DEDICATION

This dissertation is dedicated to my Dad, who loved me unconditionally. He is the source of my inspiration and motivation when I doubt myself during difficult times. He will always be in my memory. Love you and Thank you, Dad.

To P'Ja, my dear sister, who wants the best for me and is someone I can always lean on.

To Jim McDaniel, my adored partner, I value your consistent understanding and endless encouragement.

To Bentley and Neko, my best study partners, thank you for being perfect dogs and always beside me.

ACKNOWLEDGEMENT

First, I want to express my most profound appreciation to my advisor Professor Josette Jones. During this long road of the Ph.D. journey, you guided, advised, and encouraged me through these many years of academic life.

A debt of deepest gratitude is owed to Professor William D. Duncan. You went above and beyond to teach, explain, and work with me on ontology. I remembered evening and weekend WebEx meetings as well as email exchanges. Without your guidance, the completion of my dissertation would not have been possible.

My sincere gratefulness goes to Professor Ben Boukai. You are an excellent teacher. Our one-on-one sections not only helped me understand machine learning but also motivated me to continue learning statistics.

Likewise, my special appreciation is to Professor Xiaowen Liu and Professor Saptarshi Purkayatha for their time and valuable comments.

Furthermore, my heartfelt thank you is to Dr. Donald Shelbourne for allowing me to attend the Informatics program while working for his practice. I have gained many skills and experiences through my works.

Last but not least, thank you to my family in Thailand and Indiana for encouragement and caring during this journey.

Kanitha Phalakornkule

ACLRO: AN ONTOLOGY FOR THE BEST PRACTICE IN ACLR REHABILITATION

With the rise of big data and the demands for leveraging artificial intelligence (AI), healthcare requires more knowledge sharing that offers machine-readable semantic formalization. Even though some applications allow shared data interoperability, they still lack formal machine-readable semantics in ICD9/10 and LOINC. With ontology, the further ability to represent the shared conceptualizations is possible, similar to SNOMED-CT. Nevertheless, SNOMED-CT mainly focuses on electronic health record (EHR) documenting and evidence-based practice. Moreover, due to its independence on data quality, the ontology enhances advanced AI technologies, such as machine learning (ML), by providing a reusable knowledge framework. Developing a machine-readable and sharable semantic knowledge model incorporating external evidence and individual practice's values will create a new revolution for best practice medicine.

The purpose of this research is to implement a sharable ontology for the best practice in healthcare, with anterior cruciate ligament reconstruction (ACLR) as a case study. The ontology represents knowledge derived from both evidence-based practice (EBP) and practice-based evidence (PBE). First, the study presents how the domain-specific knowledge model is built using a combination of Toronto Virtual Enterprise (TOVE) and a bottom-up approach. Then, I propose a top-down approach using Open Biological and Biomedical Ontology (OBO) Foundry ontologies that adheres to the Basic Formal Ontology (BFO)'s framework. In this step, the EBP, PBE, and statistic ontologies are developed independently. Next, the study integrates these individual

ontologies into the final ACLR Ontology (ACLRO) as a more meaningful model that endorses the reusability and the ease of the model-expansion process since the classes can grow independently from one another. Finally, the study employs a use case and DL queries for model validation.

The study's innovation is to present the ontology implementation for best-practice medicine and demonstrate how it can be applied to a real-world setup with semantic information. The ACLRO simultaneously emphasizes knowledge representation in health-intervention, statistics, research design, and external research evidence, while constructing the classes of data-driven and patient-focus processes that allow knowledge sharing explicit of technology. Additionally, the model synthesizes multiple related ontologies, which leads to the successful application of best-practice medicine.

Josette F. Jones, RN, Ph.D., Chair

TABLE OF CONTENTS

List of Tables ... xi
List of Figures ... xii
List of Abbreviations ... xiv
Chapter One Introduction ... 1
1.1 Introduction ... 1
1.2 Problem Statement ... 3
1.3 Significance of the Study ... 4
1.4 Description of the Chapters ... 5
1.5 Study Approval ... 7
Chapter Two Literature Review ... 8
2.1 Ontology and Upper Ontology ... 8
2.2 Ontology in Healthcare ... 12
2.2.1 Knowledge Representation ... 12
2.2.2 Electronic Health Records (EHR) ... 13
2.2.3 Standardization and Guidelines ... 15
2.2.4 Clinical Decision Support (CDS) ... 15
2.2.5 Standard Terminologies ... 17
2.2.6 Qualitative and Predictive Analysis ... 19
2.3 Statistics Ontology ... 21
2.4 Practice-Based Evidence (PBE) Ontology ... 21
2.5 Evidence-Based Practice (EBP) Ontology ... 22
Chapter Three Methodology ... 26
3.1 BFO Foundation Framework ... 26
3.1.1 Ontology for Biomedical Investigations (OBI) ... 28
3.1.2 Information Artifact Ontology (IAO) ... 30
3.2 Methods for Ontology Implementation ... 32
3.3 Aims and Deliverables ... 35
3.3.1 Aim 1 Domain-Specific Ontology ... 35
3.3.1.1 Introduction ... 35
3.3.1.2 Methods ... 35
3.3.1.3 Results ... 36
3.3.1.4 Conclusion ... 36
3.3.2 Aim 2 Statistics Ontology ... 36
3.3.2.1 Introduction ... 36
3.3.2.2 Methods ... 37
3.3.2.3 Results ... 38
3.3.2.4 Conclusion ... 38
3.3.3 Aim 3 Evidence-Based Practice Ontology ... 38
3.3.3.1 Introduction ... 38
3.3.3.2 Methods ... 39
3.3.3.3 Results ... 39
3.3.3.4 Conclusion ... 40
3.3.4 Aim 4 Best-Practice ACLR Ontology (ACLRO) ... 40
3.3.4.1 Introduction ... 40

3.3.4.2 Methods..40
3.3.4.3 Results..41
3.3.4.4 Conclusion 41
3.4 Case Study 41
3.4.1 Dataset..42
3.4.2 Random Forest (RF) Model 44
3.4.3 Modeling in R 44
3.4.4 Formal Representation 47
3.4.5 Validation..50
3.5 Conclusion 53
Chapter Four Aim 1 Domain-Specific Ontology..55
4.1 Introduction 55
4.2 Existing Ontologies 55
4.3 Bottom-Up Approach...58
4.3.1 Determine the Domain and Scope of the Ontology 58
4.3.2 Define the Domain Terminology 59
4.3.3 Define the Informal Definitions of the Terminology..................................60
4.3.4 Define the Formal Definitions and a Concept-Mapping Diagram 63
4.3.5 Define the Domain Class Taxonomy..65
4.3.6 Define the Properties of the Classes 68
4.3.7 Develop Constraints Using Axioms, Rules and Reasoning..........................71
4.4 Top-Down Approach 74
4.4.1 Select the Foundation Framework 74
4.4.2 Review Middle-Level Ontologies and Existing Concepts............................75
4.4.3 Integrate the Middle-Level Ontologies...76
4.5 Integrate the Domain Ontology to the BFO Framework 77
4.6 Validation..80
4.6.1 Create Instances 80
4.6.2 Competency Testing 84
4.7 Conclusion 86
Chapter Five Aim 2 Statistics Ontology 88
5.1 Introduction 88
5.2 Existing Ontologies...89
5.2.1 Statistics Ontology (STATO) 89
5.2.2 Ontology of Biological and Clinical Statistics (OBCS) 92
5.2.3 APA Statistical Cluster (APASTATISTICAL) 96
5.2.4 Ontology of Clinical Research (OCRE) 96
5.2.5 Mathematical Modelling Ontology (MAMO) 96
5.2.6 Semanticscience Integrated Ontology (SIO) 96
5.3 Model Comparison..97
5.4 ACLRO's Statistics-Concept Modeling 101
5.5 Consistency Testing...104
5.6 Validation..104
5.6.1 Statistic Parameters..105
5.6.2 Statistical Test and Parameters 105
5.6.3 Statistical Test and Data Types..106

5.6.4 Statistical Test, Distribution and Hypothesis Test 108
5.6.5 Statistics and Research Design 110
5.7 Conclusion 112
Chapter Six Aim 3 Evidence-Based Practice Ontology 113
6.1 Introduction 113
6.2 Existing Ontologies...115
6.2.1 Evidence and Conclusion Ontology (ECO)...115
6.2.2 Medical Literature Search Agent (MELISA) 116
6.2.3 EMBRACE Data and Methods (EDAM) 117
6.3 Resource Description (RDF) Triples and Concept Mapping Diagram................118
6.3.1 Information Extraction Process..119
6.3.2 Drawing a Conclusion from the Data Process 119
6.3.3 EBP Concept-Mapping Diagram 120
6.4 Validation...124
6.4.1 Instance Creation 124
6.4.2 Competency Testing 126
6.5 Conclusion 128
Chapter Seven Aim 4 Sharable Best-Practice Ontology...129
7.1 Introduction 129
7.2 Practice-Based Evidence (PBE) Ontology..130
7.3 Sharable Best Practice (BP) Ontology 133
7.4 Conclusion 135
Chapter Eight Conclusion and Future Works 136
8.1 Introduction 136
8.2 Impact of the Research...137
8.3 Challenges and Limitations...138
8.4 Future Works 138
Appendices...140
Appendix A...140
References...147
Curriculum Vitae

LIST OF TABLES

Table 1 Mean decrease accuracy of two-year postoperative pain 46
Table 2 Object properties representing the health-event process in the ACLRO..............69
Table 3 Ontology metrics reported by BioPortal on Dec 2019 101

LIST OF FIGURES

Figure 1 The hierarchy of BFO 2 28
Figure 2 OBI’s investigation-related process 30
Figure 3 IAO structure 31
Figure 4 The high-level structure of OBI and IAO under the BFO framework 32
Figure 5 %incMSE of different variables on two-year postoperative pain 46
Figure 6 The visualization of the overall formal ontology structure 49
Figure 7 DL query: the top-five influential factors on ACLR postoperative pain 51
Figure 8 DL query: the top-five increase mean accuracy factors on ACLR postoperative pain 53
Figure 9 Visualization of ACLR health event process 64
Figure 10 Class hierarchy of the health-event concept in the taxonomy structure via Protégé 65
Figure 11 Object property defining the relationship between the patient and health-event concepts via Protégé 70
Figure 12 The implementation of SWRL rule for model constraints, axioms, and relation criteria 72
Figure 13 Concept-mapping diagram of ACLRO under the bottom-up approach 73
Figure 14 Partial high-level conceptual-mapping diagram of OBI and IAO under the BFO framework 77
Figure 15 Conceptual-mapping diagram of the ACLRO with the hybrid approach under the BFO framework 79
Figure 16 DL query for retrieving the female-patient (FemalePatient) concept 82
Figure 17 An instance of the Patient concept with the information of sex and birthdate via Protégé 83
Figure 18 An instance of the diagnosis-encounter concept (encounter_pat1) with the information of primary and secondary diagnosis via Protégé 83
Figure 19 An instance of the surgical-record concept (recordSrg_pat1) with the information of sex and birthdate via Protégé 84
Figure 20 STATO structure of statistics concepts 91
Figure 21 OBCS’s hierarchical structure and key ontology terms 93
Figure 22 OBCS’s semantic representation of statistics-related concepts 95
Figure 23 Sub-classes of ‘investigation’ process in STATO 98
Figure 24 Sub-classes of ‘investigation’ process in OBCS 99
Figure 25 High level of statistics-related hierarchical structure 102
Figure 26 OntoGraph: High level of statistics-related classes 103
Figure 27 Reasoner log of ontology processing by HemiT 104
Figure 28 DL query on probability distribution and parameter criteria 105
Figure 29 DL query on statistic test and parameters 106
Figure 30 DL query on the statistic model for ranking 107
Figure 31 DL query on the statistic model for correlation 107
Figure 32 DL query on the statistic model for the goodness of fit test 108
Figure 33 DL query on the statistic model for group comparisons 109
Figure 34 DL query on the statistic model for group comparison with the F distribution 110

Figure 35 DL query on the statistic model for randomized complete block design studies 111
Figure 36 DL query on the statistic model for cross-over design studies....111
Figure 37 ECO's structure of evidence concepts....116
Figure 38 Overview structure of MELISA 117
Figure 39 The overview concept diagram of EDAM ontology 118
Figure 40 The overview of EBP's concepts and their relationships in ACLRO 121
Figure 41 EBP's taxonomy structure under BFO's continuant concept....122
Figure 42 EBP's taxonomy structure under BFO's occurrent concept....123
Figure 43 Formal structure and attributes of Publication-A 126
Figure 44 DL query on predictive risk analysis of RTS in ACLR 127
Figure 45 DL query for knowledge discovery on RTS in ACLR....128
Figure 46 The concept diagram of PBE ontology 132
Figure 47 The concept diagram of BP ontology with the integration of EBP and PBE 134

LIST OF ABBREVIATIONS

Analysis of Variance (ANOVA)

Anterior Cruciate Ligament Reconstruction (ACLR)

APA Statistical Cluster (APASTATISTICAL)

Artificial Intelligence (AI)

Basic Formal Ontology (BFO)

Clinical Decision Support (CDS)

Clinical Decision Support System (CDSS)

Computer Interpretable Guidelines (CIG)

Descriptive Ontology for Linguistic and Cognitive engineering (DOLCE)

Enterprise Integration (EI)

Evidence and Conclusion Ontology (ECO)

Evidence Based Practice (EBP)

Electronic Health Record (EHR)

Gene Ontology (GO)

Health Level-7 (HL7)

International Classification of Disease (ICD)

Knowledge Interchange Format (KIF)

Knowledge Representation (KR)

Knowledge Sharing (KS)

Information Artifact Ontology (IAO)

Logical Observation Identifiers Names and Codes (LOINC)

Machine Learning (ML)

Malaria Ontology (IDOMAL)

Mathematical Modelling Ontology (MAMO)

Mean Squared Error (MSE)

Mean Decrease Accuracy (incMSE)

Medical Subject Headings (MESH)

Ontology Clinical Research (OCRE)

Ontology for Biomedical Investigation (OBI)

Ontology of Biological and Clinical Statistics (OBCS)

The Oral Health and Disease Ontology (OHD)

Patient Intervention Comparison Outcome (PICO)

Practice Based Evidence (PBE)

Probabilistic Extension to OWL (PR-OWL)

Random Forest (RF)

Range of Motion (ROM)

Resource Description Framework (RDF)

Return to Sport (RTS)

Statistics Ontology (STATO)

Semantic Automated Discovery and Integration (SADI)

Semanticscience Integrated Ontology (SIO)

Service-Oriented Architecture (SOA)

SPARQL Protocol and RDF Query Language (SPARQL)

Suggested Upper Merged Ontology (SUMO)

Systematized Nomenclature of Medicine – Clinical Terms (SNOMED-CT)

Taxonomy for Rehabilitation of Knee Conditions (TRAK)

Toronto Virtual Enterprise (TOVE)

Unified Medical Language System (UMLS)

Web Ontology Language (OWL)

World Wide Web (WWW)

CHAPTER ONE INTRODUCTION

1.1 Introduction

Since the early 1990s, health information technology (HIT) has played an essential role in improving health care delivery globally (Zillner et al., 2014). Electronic Health Record Systems (EHRs) and their clinical decision support (CDS) were essential influencers providing novel functions over traditional documentation in health care and facilitating continuity of care throughout the patient's lifespan across regional and healthcare systems. The CDS system (CDSS) is programmable software that helps clinicians in decision making at the point of care by advising the best evidence and alerting clinicians with analyzed information that is in its knowledge model (Phalakornkule, Jones, & Finnell, 2013). Therefore, the knowledge model in the CDSS is critical to its performance, efficiency, and accuracy (O'Neill, Dluhy, Fortier, & Michel, 2004). Additionally, the demand for information exchange among EHRs is rapidly growing in the company with the efforts to improve the quality of patient care (Adler-Milstein, Bates, & Jha, 2013). In order to succeed in knowledge sharing, individual practices must develop their sharable knowledge framework (Wang & Noe, 2010).

There are two types of knowledge sources for CDSS: (1) Evidence-Based Practice (EBP) and (2) Practice-Based Evidence (PBE). EBP collects knowledge through the best external research evidence, while PBE learns through clinicians' own experiences gained in a local environment (Barkham & Mellor-Clark, 2003). EBP has been in medicine for decades and is more commonly used in CDSS (Purcell, 2005). However, it does not offer a perfect approach in all situations. First, EBP is neither practice nor patient-focused, leading to a lack of information about individual patients, healthcare providers, as well as clinical workflow from its conclusions (Slim, 2005). Therefore, EBP should not be used by itself without requiring additional information about the patient, clinician, and practice criteria(Wilkinson et al., 2000). Next, the knowledge discovered in EBP is gathered from the aggregated data analysis under a controlled setting, such as randomized and controlled trials (Tonelli, 1998). This setting is not proper for day-to-day patient cares. Besides, many research confounder variables and the description of the study environment are not published in articles (Chung, Swanson, Schmitz, Sullivan, & Rohrich, 2009). The influences of this information are essential, but unknown in publications that can cause a misinterpretation of the actual knowledge reported in EBP. On the other hand, Practice-Based Evidence (PBE) utilizes clinical expertise and gathers data from individual and routine practice settings (Horn & Gassaway, 2007). Comparing to EBP, PBE is more patient and process-centered. Hence, PBE interventions can overcome the limitation of EBP's actionable performances. Nonetheless, PBE can be timely and costly since its implementation requires a large amount of data to collect enough evidence for its knowledge model (Loane et al., 2000). Additionally, PBE faces a challenge in sharing its finding and integrating new knowledge to its current knowledge across many practices (Hudson & Collins, 2015)

The semantic mapping between EBP and PBE combines both advantages for the best-practice knowledge model. The reusability and exchangeability of knowledge are the critical success of semantic mapping, which cannot be achieved without using a foundation ontology model. The purpose of this research is to implement a sharable ontology for the best practice in healthcare, with anterior cruciate ligament reconstruction (ACLR) as a study domain, presenting the knowledge from both evidence-based practice

(EBP) and practice-based evidence (PBE). First, the study defines how the domain-specific knowledge model is built using a bottom-up approach with the Toronto Virtual Enterprise (TOVE) guideline. The study then proposes a top-down approach using the Open Biological and Biomedical Ontology (OBO) Foundry ontologies that adheres to the Basic Formal Ontology (BFO) framework. In this step, the EBP, PBE, and statistics ontologies are developed independently. Next, the study integrates these individual ontologies into the final ACLR Ontology (ACLRO) as a more meaningful model that endorses the reusability and the ease of the model-expansion process since the classes can grow independently from one another. Besides, the study employs a use case, and DL queries for model validation.

1.2 Problem Statement

With the rise of big data and the high demands for leveraging artificial intelligence (AI), healthcare requires more knowledge sharing that offers machine-readable semantic formalization. Although some applications allow interoperability of shared data such as in ICD9/10 and LOINC, they still lack formal machine-readable semantics. With ontology, the further ability to represent the shared conceptualizations is possible, similar to SNOMED-CT. Nevertheless, SNOMED-CT mainly focuses on electronic health record (EHR) documenting and evidence-based practice. Additionally, Health Level-7 (HL-7) and ontology are used for information exchange in healthcare. HL-7 successfully provides a method for developing standards for the EHR data exchange and bridging the gap between EHR systems (Hammond, 1993). However, the only goal of HL-7 is to exchange healthcare data or information, not knowledge discovered in individual practices that cannot be extended to other non-healthcare domains.

Moreover, due to its independence on data quality, the ontology enhances advanced AI technologies, such as machine learning (ML), by providing a reusable knowledge framework (Hwang, Park, Lee, Kim, & Lee, 2018). The development of a machine-readable and sharable semantic knowledge model that unites external evidence and individual practice's values will create a new revolution for best practice medicine. As now, there is no technology with a complete set of ontology's functionalities. These functionalities allow the exchangeability of semantic information across various domains and are insensitive to the volume, velocity, and variety of big data.

1.3 Significance of the Study

Opportunely, with an ontology implementation, the development of a knowledge-model framework is possible and can be deployed and dispersed rapidly over healthcare organizations. An ontology offers a formal representation and semantic maps of knowledge across practices as staying independent from healthcare applications that allows individual practices to maintain their workflow and organization requirements while exchanging their knowledge (Dang, Hedayati, Hampel, & Toklu, 2008). Concept mapping diagrams and knowledge editors such as Protégé by Stanford University are practical tools used in semantic networks representing and implementing knowledge models. Besides, the framework of an upper ontology allows us to build a standard model that can be reusable and sharable with the independency of technology (Jarrar, 2005).

For the above reasons, this study implemented the sharable knowledge model combining the strengths of both EBP and PBE using a hybrid approach that is a novel method of ontology implementation, facilitating both bottom-up and top-down approaches (López-Pellicer et al., 2007). The bottom-up approach emphasized capturing domain knowledge in an ontological structure, while the top-down approach adds on the foundation framework increasing the reusability and shareability of the domain ontology. Another innovation of this study is to present the ontology implementation for best-practice medicine and demonstrate how it can be applied to a real-world setup with semantic information. The ACLRO simultaneously emphasizes knowledge representation in health-intervention, statistics, research design, and external research evidence, while constructing the classes of data-driven and patient-focus processes that allow knowledge sharing explicit of technology. Accordingly, the model synthesizes multiple related ontologies, which leads to the successful application of best-practice medicine.

1.4 Description of the Chapters

Chapter 2 includes five literature reviews: (1) Ontology and Upper Ontology. This section reviews the definitions of an ontology and presents the roles of ontology and upper ontology along with the values they bring, focusing on the most common upper ontology used in the biomedical and healthcare domain. (2) Ontology in Healthcare. This review includes the introduction of ontology and presents some successful ontology applications in different areas of healthcare: knowledge representation, Electronic Medical Record, Standardization and Guidelines, Clinical Decision Support, and Qualitative and Predictive Analysis. (3) Statistics Ontology. This section starts from reviewing all known existing ontologies related to statistics models. Then, the comparison for usability and maturities are performed. (4) Practice-Based Evidence (PBE). This sector analyzes the method and roles of PBE along with the review of ontologies in clinical researches for domain-specific and local practices. (5) Evidence-Based Practice (EBP). This section reviews the roles of EBP in the patient cares and discusses its strengths and weakness, along with the analysis of existing ontologies for scientific evidence representation.

Chapter 3 describes the methodology of the study and the review of an upper-level ontology, i.e., the Basic Formal Ontology (BFO) along with the middle-level ontologies, i.e., Ontology for Biomedical Investigations (OBI) and Information Artifact Ontology (IAO), of this dissertation, as well as the methods of Ontology implementations. Additionally, Chapter 3 summarizes the dissertation's aims, implementation methods, and outcomes of five individual models:

- Domain-Specific Ontology
- Statistics Ontology
- Evidence-Based Practice (EBP) Ontology
- Practice-Based Evidence (PBE) Ontology
- The Best-Practice (BP) ACLRO

Also, this chapter presents the Random-Forrest case study, its formal representation, and its validations.

Chapter 4 displays the first aim of the study on the implementation of the domain-specific ontology starting from the bottom-up approach following Toronto Virtual Enterprise (TOVE) methods and the American Nurses Association (ANA) guidelines.

Additionally, Chapter 4 presents the scopes of the ACLRO and the implementation of sharable formal structures with a set of predicates under the BFO framework.

Chapter 5 describes the Aim 2 of this dissertation. i.e., statistics ontologies. The chapter reviews existing statistical ontologies and compares their reusability and structures. Multiple queries perform for validating the structure and relations between statistical models and their criteria.

Chapter 6 reveals the Aim 3 of the dissertation on the implementation of the evidence-based model's formal structure. This chapter records two publications in the return-to-sport possibility after ACLR as two instances of publication classes. Then, the chapter presents the validation and conclusion.

Chapter 7 exhibits Aim 4, the final aim, of the dissertation on the integration of domain-specific, PBE, EBP, and statistics ontology to the final Best Practice ACLRO model. Under the BFO framework, the integration requires no editing in any individual models proving the success of ACLRO’s semantic-information shareability.

Chapter 8 summarizes the dissertation's overall conclusion, the limitation, and future works of the study.

1.5 Study Approval

The Indiana University Institutional Review Boards (IRBs) reviewed and exempted this study.

CHAPTER TWO LITERATURE REVIEW

Due to the dissertation's aim on the sharable best-practice ontology model, this dissertation's background is categorized into five areas:

- Ontology and Upper Ontology
- Ontology in Healthcare
- Statistics Ontology
- Practice-Based Evidence (PBE) Ontology
- Evidence-Based Practice (EBP) Ontology

The Ontology and Upper Ontology literature explain the roles and types of an ontology framework that provides a formal structure of a sharable semantic information model in various fields. The Ontology in Healthcare literature presents the various areas in healthcare that an ontology had an impact on, like knowledge representation, electronic medical record, standardization and guidelines, clinical decision support, qualitative and predictive analysis. The statistic-ontology literature reviews the ontology framework of statistical concepts and its applications in the healthcare domain. The PBE ontology literature presents the ontology applications built under the local environment and for a practice's specific aim. The EBP ontology literature reviews the role of EBP in healthcare, and the ontology framework improves its impact on healthcare researches.

2.1 Ontology and Upper Ontology

The 'Ontology' term was initially considered a branch of philosophy but has received more recognition in computer science and informatics. In modern technology, the 'ontology' term does not have the exact meaning in all research areas. Some refer to ontology as its philosophical meanings, while some think of ontology as a new area of science used in information systems, databases, and computer science. Philosophical ontology is a type of science about the object-theory of existing entities, their properties, and relations in reality. The theory is represented in the form of hierarchical structures called taxonomies connecting entities through the framework of classes and sub-classes. Despite the differences in Ontologies' roles, all sciences accept an ontology as an understanding of how things existing in human's perspectives (Zúñiga, 2001).

The use of contemporary ontology has involved in various branches of sciences, including computer science, medicine, informatics, and life sciences. In computer science, an ontology is a representation of knowledge modeling capturing the semantic information of concepts used in a specific domain and making them interpretatively exchangeable and shareable (Gruber, 2018). In an information system, an ontology is an explicit description of conceptualization and a structure of a formal description of the concepts, attributes, and relations in a specific domain (Guarino, Oberle, & Staab, 2009). A conceptualization is an abstract of individuals' beliefs or perspectives. In the artificial intelligence (AI) field, the term Ontology means the new field of knowledge engineering or knowledge representation derived from its original definition in philosophy. The list of known applications with an ontology includes natural language representation, decision-support systems, standard terminology, and machine learning. A unique example is a centralized application like service-oriented architecture (SOA), which can manage complex and heterogeneity (Mohammadi & Mukhtar, 2013). The most popular ontology applications are the semantic web structuring World Wild Web (WWW)' meta-data to a taxonomy framework, which allows web pages to be sharable or linkable to one another (Berners-Lee, Hendler, & Lassila, 2001).

An upper ontology, referred to as a foundation or top-level ontology, is a high-level ontology offering a universal and domain-independent formal framework that more domain-specific ontologies can be built on (Elmhadhbi, Karray, & Archimède, 2019). The primary purpose of upper ontologies to provide universal knowledge and terminologies across domains to ensure generality and reusability by offering the level of granularity of the model leading to exchangeability (Obrst et al., 2014). A practical upper ontology must be small and generalized. It allows consistency across different domains to extend the reusability and shareability by providing a common framework as a bridge for integrating heterogeneous ontologies in an automatic way (Degen, Heller, Herre, & Smith, 2001). There are several known upper models used in biomedicine and healthcare, such as Suggested Upper Merged Ontology (SUMO), Basic Formal Ontology (BFO), and Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) (Mascardi, Cordì, & Rosso, 2007). Nevertheless, each upper ontology has its strengths and emphasis in different fields. For instance, the BFO is more accepted in biomedical, medicine, and life sciences. One of the widely used BFO is a part of the genomic domain, as in the Gene Ontology (GO) project. The GO project offers controlled vocabularies and standard structure in molecular and cellular biology domains by offering semantic information of the vocabularies and contributed annotations (Consortium, 2004). The BFO is also a big part of cancer researches, such as prostate cancer (Overton, Romagnoli, & Chhem, 2011), cervical cancer (Maramis et al., 2013), and cancer cell research (Rasmussen & Dolan, 2013). Another most known BFO application is Systematized Nomenclature of Medicine – Clinical Terms (SNOMED-CT) that is a comprehensive medical terminology used for standardizing the storage, retrieval, and exchange of electronic health data. The SNOMED-CT ontology (SCTO) is implemented in OWL 2 and has Open Biomedical Ontologies (OBO) Foundry ontologies that are under the BFO framework as middle-ontologies for the integration of various terminologies (El-Sappagh, Franda, Ali, & Kwak, 2018). The BFO-is also widespread in healthcare applications as presented in the patient safety application using BFO as an upper-level ontology to improve the interoperability of patient-safety reporting systems (Cornet, 2015).

SUMO is another standard upper ontology originated by a group of engineers from the IEEE standard Upper Ontology Working Group (Pease & Niles, 2002). The goal was to promote data interoperability, information search, automated inference, and natural language processing. A successful SUMO application was presented in the 2007 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications as a framework of general concepts for a universal ontology for sensor networks data (Álvez, Lucio, & Rigau, 2019). The project successfully enhanced the interoperability between multiple sensor networks by improving both the precision and recall rates. Another success of SUMO was presented in modeling legal terminology (Mitrović, Pease, & Granitzer). The paper utilized SUMO as a connection to the lexical-semantic network WordNet in the legal domain. Similar to SUMO, DOLCE plays a vital role in linguistic and cognitive researches. One of its successful application was the LOIS project on multilingual information retrieval from legal databases (Tiscornia, 2006). Another success of DOLCE involved in a project on language technology for eLearning, using multilingual language technology tools and semantic web techniques for improving the retrieval of eLearning material (Monachesi et al., 2006).

In conclusion, all three upper ontologies provide the same ability of semantic information exchangeability in different applications. An ontology developer should select the upper ontology based on the scope and domain of the study. For this project, the study selected the BFO as its upper-level ontology. More introductions to the BFO and its advantages over other top-level models are addressed in the next chapter.

2.2 Ontology in Healthcare

In healthcare, ontologies have been received much attention from different areas, from knowledge representation to predictive applications (Alan Jovic, Marin Prcela, & Dragan Gamberger, 2007). These areas include, but not limited to, knowledge representation, electronic health records (EHR), standardization and guidelines, clinical decision support (CDS), standard terminologies, qualitative and predictive analysis (Flouris, Plexousakis, & Antoniou, 2006).

2.2.1 Knowledge Representation

One example of the ontology success in knowledge representations was reported in the "Even Oriented Representation for Collaborative Activities (EORCA)" paper implementing a method covering the observation and the representation of collective activities during patients' management which could be reusable by the team members in order to prepare themselves for accreditation (L Pellegrin et al., 2007). The ontology was used to build as a knowledge representation of standard guidelines for task observation by the ICU team. In the study in the cardiology field, the UMLS-based ontology supported the cardiology procedures for cognitive support in medical decision making allowing different stakeholders and healthcare groups to share knowledge management and communication (Biolchini, 2002). The ontology also allows domain knowledge to be independent of technology so that it can be reused effectively across multiple platforms. One example is in the “Clinical Decision Support System for Point of Care” paper utilizing an ontology model and reasoning rules in machine learnings, and further proving that when the knowledge was changed, only the reasoning was changed without any work on the software system (Farion et al., 2009). As well, Doyle and his team implemented an ontological knowledge base for public health surveillance that could be referred to in various applications. Because of its independence from technology, the study concluded that ontology was beneficial for information exchange in EHRs (Doyle, Ma, Groseclose, & Hopkins, 2005).

2.2.2 Electronic Health Records (EHR)

Nowadays, it is impossible to look at healthcare without involving EHRs. Formalizing a concept is the first and foremost step in establishing a knowledge-based software system of any kind in healthcare. One of the main struggles of EHRs is a lack of standardization. Without standardization, the concept of sharable and combinable EHRs would not be feasible. The goal of applying Ontology in EHRs is to provide a systematic representation of various medical knowledge used for different types of reasoning in healthcare activities (A. Jovic, M. Prcela, & D. Gamberger, 2007). An example system using Ontology to enhance EHRs’ semantic functionality is presented in the XOntoRank system, the tool for Ontology-aware search of EHRs used to solve the problem of facilitation Ontology-aware data extraction for EHR database for XML-based documents (Farfan, Hristidis, Ranganathan, & Burke, 2008). With Ontological definitions in Systematized Nomenclature of Medicine (SNOMED), the system could perform a semantic search on the XML documents. There was no need for a perfect match between keywords in a query and words in documents due to the semantic-term ontology. First, the author listed the potential problematic keywords. Next, he developed the semantic concepts for these keywords. Then, the degree of association between ontological concepts and keywords was assigned. Last, the study calculated and evaluated the algorithms to answer ontology-aware keyword queries in EHRs. As a result, the XOntonRank search-engine found more matches. It found the exact-match keywords and the matched concepts with a better precision rate and a recall rate than the baseline algorithm for the top-k results. The study presented that using conceptual mapping can be more efficient than terminological mapping. Peleg also developed another Ontology-based mapping tool in EHRs presented in the 'Mapping Computerized Clinical Guidelines to Electronic Medical Records: Knowledge-data Ontological Mapper (KDOM)" paper (Peleg M., 2008). The KDOM was an ontology-mapping framework connecting computer-interpretable guidelines (CIGs) to EHRs to create shareability and reusability between various institutions. With the ontology, the solution for an incompatible issue was discovered by separating the medical domain knowledge from the operational knowledge, making domain assumptions explicit, and developing a bottom-up ontology. The classes were divided by properties into two groups: properties conceptualizing the abstract knowledge and properties retrieving fields in the EHR tables. The slots in the mapping ontology were used to refer to the destination fields in EHR databases and to specify how the retrieved value was matched with a constant value. As a result, the ontology framework successfully mapped guidelines to the EHR and bridged the gap between the abstractions in a CIG and EHR, which could be further configured independently.

2.2.3 Standardization and Guidelines

There are many uses of ontology as a reference or guideline for healthcare organizations. Each organization or facility requires a set of their own needs and preferences, which will not be compatible with the others. In the "Ontological Knowledge Framework for Adaptive Medical Workflow" paper, the study presents an ontological knowledge framework that covers tasks from administration to patient care by capturing all-important knowledge for complex personalized events, including patient care and insurance policies and drug prescriptions (Dang et al., 2008). While involving more of a business perspective, ontology defined concepts in business rules and policies along with personalized patient context for machine interpretation to support adaptive workflow composition and execution. The adaptive workflow system had the functionalities for users to -control and monitor the patient-process, to manage patient's medical records, to create new tasks from a medical service repository, and to maintain historical process data for future use without the need of technical support. The study proposed was to develop an adaptive workflow system that could manage without the knowledge of technical parts. Moreover, the software used ontology's meta-data to learn about the domain's environment and rules. It separated the business rules from process rules. This adaptive workflow system was claimed as the first achievement in bridging healthcare needs and technology in any hospital environment.

2.2.4 Clinical Decision Support (CDS)

For clinical decision support, an ontology enhances its performance as reported in the "Even Oriented Representation for Collaborative Activities (EORCA)" paper that implemented a method covering the observation and the representation of collective activities during patients' management which could be reusable by the team members in order to prepare themselves for accreditation (L. Pellegrin et al., 2007). The ontology model was built to represent and combine data of ECG signal and heart image for CDSS (Chiarugi et al., 2008). Correspondingly, Jovic, and his team published the paper, the ontology for knowledge representation in heart-failure-patient managements, was the integration between domain ontology and UMLS terminology ontology (Alan Jovic et al., 2007). The study's outcome showed that the utilization of both OWL and SWRL was a successful tool for reasoning in complex medical systems. The OWL approach with the closed work assumption enabled complete, actionable knowledge into the ontology framework, which led to the development of CDSS. In summary, the study concluded that Ontology was famous for the standardization of medical terms, knowledge sharing, and automatic reasoning.

The additional use of Ontology in CDS involves statistical methodologies to run reasoning rule-base utilities (Montani et al.). For instance, in the "Surgical Models for Computer-Assisted Neurosurgery" paper (P. Jannin, 2007), the ontology was used to develop a decision tree helping the surgical CDSS in patient outcomes and correctly predicted 76.27% in subgroups and 45.28% in the whole group. Kim and Choi also demonstrated that CDSS for heart disease detection could combine multiple domain knowledge (K. Kim, 2007). Their ontology was built with Protégé, SWRL, and JESS for rule creations. Additionally, the CDSS developed in Mago's study was implemented using a multi-agent system for healthcare practitioners for Indian rural childcare (V. Mago, 2007). The Ontology was used in a user-agent side to understand the illness through common vocabulary to diagnose the disease and treatment plans. Likewise, Panzarara presented in her paper that CDSS with Ontology improved knowledge and skills management and coordinated patient care over time (Panzarasa, Madde, Quaglini, Pistarini, & Stefanelli, 2002). All of these studies delivered CDSS's reasoning directly to patients. Equally, the "Era of Patient Safety Implications for Nursing Informatics Curricula" paper concluded that the ontology played the leading role in CDSS, integration, and standard for patient safety in a clinical environment (J. Effken, 2002). In summary, modern EHRs require a framework to support their enhanced functionalities (S. Mersmann, 2004). When the demand for using EHRs expands from local to multiple institutions, a standard framework is in need. The ontology can increase EHR's functionalities in standardizing medical terms, knowledge sharing, and support for automatic reasoning using in CDSS.

2.2.5 Standard Terminologies

One of the most well-known uses of ontology in healthcare is for standard terminology systems such as UMLS (Unified Medical Language System), SNOMED, and LOINC (Pisanelli, Gangemi, & Steve, 1998). These systems offer standardized communication, documentation, and classification of health/medical vocabularies (Cole, 2004). Nonetheless, even these terminology systems are all based on a standard structured framework; the concepts among these systems are not entirely consistent nor compatibility (B. Bolbel, 2006). Additionally, ontology has played an essential role in linking a study's domain knowledge to a standard terminology system. For example, in the "Facilitating Pre-operative Assessment Guidelines Representation Using SNOMED CT" paper, the authors investigated whether SNOMED CT covers the terms using in pre-operative assessment guidelines paper (L. Ahmadian, 2009). They found out that 71% of guidelines were matched with SNOMED, while 69% of 39 non-completely covered concepts violated at least one of SNOMED CT formats. The authors proposed that ontology could be a potential solution for formalizing the guidelines in SNOMED CT. Another illustration is the "Towards Role-Based Filtering of Disease Outbreak Report" paper, using ontology to analyze conceptual classifications of infectious diseases that were not presented in any terminology systems (S. Doan, 2009). Also in the "CardioOP data Class (CDC)" paper, the study reported that none of the terminological systems; i.e., ICD 10, SNOMED, UMLS nor MESH provided enough granularity of contents or domain completeness for metadata in multimedia data in the Cardio domain (Fried, Klas, & Westermann, 2003). In the study of Elkin, Ontology was used to build terminology structure for automated systems providing classification in negation and propagation in clinical notes (Elkin et al., 2005). Accordingly, these existing terminology domains can integrate existing definitions and terminologies across different healthcare level (D. Pappa, 2006), as presented in the "In the Category Structure for a Terminology System in Traditional System in Traditional Medicine, Symptoms, Signs, and their Combination Patterns" paper using ontology to develop a standard terminology for different areas and countries (Park et al., 2009). Another example of using ontology in mapping terminologies across languages was presented in the "Coupling Indigenous Patient-Friendly Cultural Communication with Clinical Care Guidelines for Type 2 Diabetic Mellitus" paper proposing a standard system to local facilities in order to reduce any biases from local medical staffs, possibly influenced by cultures and environments (Forbes, Sidhu, & Singh, 2011). The ontology was a solution offering a mapping model between the local Australian clinical taxonomy guidelines and the Aboriginal-English version, solving the issues of individual ambiguity and misinterpretation found in AE educational literature Aboriginal English guideline to research-based guidance. As well, Nardarni and the team presented in their “Migrating Existing Clinical Content from ICD-9 to SNOMED” paper showed how ontology could be used to map the concepts between two different systems (Prakash M Nadkarni, 2010). Likewise, Saunders used ontology to pool Anatomical Therapeutic Chemical (ATC) and Critical Term to Australian Drug Safety Data, enhancing the association rule methods to explore more rare disease data. In another sample, in the study in “ACR Appropriateness Criteria: Translation to Practice and Research” paper, the study was designed to convert the current text-based contents into a relational database (Sistrom, 2005). With the implementation of ontology, the study successfully represented a formal structure of a relational database storing the master version of the guidelines, which allowed the criteria to be distributed quickly.

2.2.6 Qualitative and Predictive Analysis

Another benefit of ontology is to structure a qualitative work. The traditional medical research design involves more quantitative analysis and not much of qualitative analysis. Qualitative research is performed differently from the quantitative study since it does not measure quantified outcomes nor answer hypothesis research questions as an essential part of the research process (Phillimore & Goodson, 2004). Ontology offers the ability to represent a formal qualitative knowledge model, a challenge in healthcare. As an example, in Shankar's published paper, the ontology was used to understand patient perceptions, which could be transformed into a formal structure (Ramakrishnan & Vijayan, 2014). Likewise, Meghani used ontology to acquire knowledge of patient attitudes toward pain described by cancer patients (S. Meghani, 2007). Correspondingly, an ontology was also presented as a tool to acquire knowledge of the language for spiritual pain through a research study (McGrath, 2002). Another ontology application was presented in the Waiting for A Liver Transplant paper; the research team used an ontology model to understand the nature of the way patients waiting for liver transplant (Jill Brown, 2006). In addition, more advanced use of ontology is in the predictive model and reasoning. With OWL, the utilizations of ontology and semantic web advances to new challenging areas like machining learnings and mapping to a relational database schema. Some examples of machine-learning ontologies include Probabilistic Extension to OWL (PR-OWL) and Fuzzy Ontology. PR-OWL overcomes the limitation of deterministic classical logic in legacy ontologies by providing a principle of uncertainty concept in ontologies (Da Costa, Laskey, & Laskey, 2006). Similarly, fuzzy ontology incorporates fuzzy logic into ontologies to deal with vague, imprecise information, which is a common issue in the real-world setting (Calegari & Ciucci, 2007). Another research by Abidi and his team presented that the ontology presented the knowledge structure allowing the computerization of a specific clinical pathway for prostate cancer disease (S. Abidi, 2009). Their ontology was based on the concept of branching and merging nodes, which were modeled as interclass intersections. Accordingly, the paper was able to merge three different clinical pathways to one. In the "An Intelligent and Integrated Platform for Supporting the Management of Chronic Heart Failure Patient" paper (S Colantonio, 2008), the ontology was structured as a formalization of the chronic heart failure domain for the knowledge sharing of information across stakeholders and facilities.

2.3 Statistics Ontology

Statistics is the science of collecting, analyzing, interpreting, and presentation of data. Biostatistics has played an essential role in medicine and healthcare through research designs and experiments. The results of the studies' outcomes have been reported in publications. Nevertheless, the reproducibility of the study outcomes was not possible in new different environments. Consequently, the reusability of knowledge discovered is unmanageable due to the lack of sharable statistical information. As an example of the solution, an ontology provides a formal metadata structure that describes how fuzzy knowledge were captured (Zheng et al., 2016). There are two accepted statistics ontologies in the OBO foundry library: The Ontology of Biological and Clinical Statistics (OBCS) and Statistics Ontology (STATO). An example of the OBCS application reported in the study of a meta-analysis of host responses to yellow fever vaccines (Zheng, Li, Liu, & He, 2017). The OBCS and the vaccine ontology (VO) structured an ontological model of various components and relations in the study. The study's result reported a statistical model using OBCS successfully conducted a literature meta-analysis to survey yellow fever vaccine response papers and statistical methods. Another example was published in the study in gene ontology with statistics. The study reported the prospering implementation of the GoPipe that was a standalone package integrating DNA sequences files from various sources to the Gene Ontology (GO) annotation with a built-in statistical option (Chen et al., 2005).

2.4 Practice-Based Evidence (PBE) Ontology

Domain knowledge commonly originates from experts' knowledge. The experts extend their knowledge through working experience, education, conferences, and research literature. The practice-based evidence (PBE) records actual patient outcomes specific to local practice and uses the outcome data on all patients to support decision-making on the treatments. The PBE can remove the inefficiency of EBP in the re-productivity of study outcomes. One of the good examples displaying the impact on an ontology on PBE is the study on ontology-driven hypothesis generation to explain anomalous patient responses to treatment (Moss et al., 2010). The study implemented a tool to determine how ICU clinicians identify anomalous patient responses. The high-level reasoning deployed by the clinicians was structured as a formal ontology model of the procedural component. After the evaluation process, the study reported the success of the reproduction of the clinician's hypotheses in the majority of cases. Another ontology development for PBE was a part of the Type II Diabetes Mellitus (DM) Clinical Support System (Chalortham, Buranarach, & Supnithi, 2009). The study's goal was to propose an ontology-based tool allowing non-experts to advise DM patients for improving life quality. The tool applied suitable criteria for the practice to fit the needs. In the study about personalized treatment, an ontology model provides a framework of retrospective and prospective diagnosis and medical knowledge personalization for the care of chronically ill patients in the local project (Romero-Tris, Riaño, & Real, 2010). The study reported the successful experiences of the improvement in missing data, wrong diagnosis, comorbidities, and prediction of patient outcomes.

2.5 Evidence-Based Practice (EBP) Ontology

EBP is an interdisciplinary methodology that has been involved in medicine over many decades. It is derived from a meta-analysis of literature or research studies in randomized controlled trials (O. Nee, 2010). In medicine, EBP is the integration of best external research evidence with clinical expertise and patient values (Robinson, 1997). The most common source of external evidence is published in the medical research literature (Wright, Swiontkowski, & Heckman, 2003). Based on its definition, EBP has three main components, i.e., external academic or scientific evidence, clinical expertise, and individual patient values, providing the best possible individualized recommendation based on the available evidence and patient characteristics (Titler, 2008). The EBP exercise integrates the best available external clinical evidence from systematic research into individual clinical practices to improve patient outcomes and quality of care. Then, the domain experts analyze the EBP evidence. Without new external resources, the practice is at risk for outdated technologies and being behind in information, such as new drug discoveries or surgical tools. Therefore, two core components of EBP and PBE are clinical expertise and external evidence. The traditional EBP aims to apply external evidence to a non-specific local practice but is not intended for direct knowledge reusability or the exchange of knowledge across local practices. In a sense, the traditional EBP has a similar limitation to KIF in terms of knowledge transfer.

Another quality of the EBP is in its performance of information extraction (Zhao et al., 2010). One of the most common challenges in text mining and natural language abstraction is transforming natural human language into a computable format like classification that cannot be implemented effectively when the meanings of the variables are unclear. Again, the ability to assign classification correctly is essential for computerized CDSS because the classification is one of the fundamental keys in knowledge models and to calculate probabilistic inference (Anitha & Rajagopalan, 2011). Furthermore, EBP excludes the ability to adapt its knowledge to new settings. It is known among researchers that publishing an article is a long process. As Morris stated in his paper on time lags in translational research, only fourteen percent of original research could benefit patient care within seventeen years, causing the information reported in the literature to be outdated (Morris, Wooding, & Grant, 2011). Thus, the outcomes and summaries of literature can be inconsistent and changed through time (Cercone, An, Li, Gu, & An, 2011). Without the ability to capture the most recent evidence related to the domain, it cannot deliver the best-updated evidence to the practice (Stiwne & Abrandt Dahlgren, 2004). As a consequence, the practices can be lagged behind knowledge by several years if they only depend on EBP sources (Kazdin, 2008). Besides, the EBP offers an aggregate analysis of patient demographics, study design, and practice setup but is not designed for personalized patient cares (Zhao et al., 2010). Nevertheless, there are unknown or impossible influential confounders in EBP, such as patient life-style, sample size, as well as statistics tools (Lucock et al., 2003b). Thus, to help individual practices maintain patient-focused intervention, EBP needs to integrate with local evidence (Patel, Vichich, Lang, Lin, & Zheng, 2017).

Another use case of EBP ontology is the Evidence and Conclusion Ontology (ECO) that describes scientific-evidence types collected from laboratory experiments, computational methods, and literature within the biological research domain (Chibucos, Siegele, Hu, & Giglio, 2017). Since 2016, ECO implemented development in the Web Ontology Language using Protégé for viewing and editing on a small scale, as well as ROBOT (http://robot.obolibrary.org) on a large scale. The ECO also reused and collaborated with other ontologies, such as Gene Ontology (GO), Ontology for Biomedical Investigations (Robinson), Ontology of Microbial Phenotypes (OMP), and

Synapse Gene Ontology Annotation Initiative (SynGO). The ECO terms were grouped based mainly on the evidence's biological context and the technique used to generate the evidence. Some terms related to both categories. Therefore, ECO developed logical definitions of these terms under technique concepts linked to relevant assay-based OBI terms. As a result, the ECO model reduced the issue of ambiguous classes.

To sum up, when EBP provides knowledge that might be missing or incomplete in practice, PBE can add the practice information to the knowledge base in CDSS (Sim et al., 2001) and allow clinicians to evaluate EBP for the most suitable outcomes (Lucock et al., 2003a). Without the evaluation, the adjustment for the best fit of the EBP use in local practice setup is impossible; all benefits of EBP to be dismissed in reality. Therefore, the synthesis of EBP and PBE will enhance the quality of patient care in EHR, complementing the Meaningful Use’s requirement (McDonald & Viehbeck, 2007).

CHAPTER THREE METHODOLOGY

3.1 BFO Foundation Framework

A foundation ontology, also known as an upper ontology or top-level ontology, is a domain-neutral ontology that comprises of general terms across all domains (Hoehndorf, 2010). There are several top-level ontologies such as Basic Formal Ontology (BFO), Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE), General Formal Ontology (GFO), Suggested Upper Merged Ontology (SUMO) and Cyc Upper Ontology. Their primary purposes are to ensure the ability of semantic integration and reusability of specific ontologies (Gibson, 2010). Therefore, the critical factor of selecting one top-level ontology over the others is on the level of interoperability that a specific domain ontology receives. For instance, an ontology in medicine should select a top-level ontology that is commonly used in the biomedical field.

After thoroughly reviewing the top-level ontologies in biomedicine, this study selects the Basic Formal Ontology (BFO) as the foundation ontology. Barry Smith and his team at the Institute developed BFO for Formal Ontology and Medical Information Science (IFOMIS) at the University of Leipzig (Grenon, Smith, & Goldberg, 2004). The design of BFO is to support information retrieval, analysis, and integration in scientific and biomedical research (Arp & Smith, 2008). Its primary purpose is to endorse semantic information applications' interoperability as a non-domain specific foundation ontology (https://basic-formal-ontology.org/). BFO’s organization is built on entities and their relationships under the single framework of time and space consolidation (Galton, 2018). Entities are classified in a taxonomy format, comparable to the parent-child class.

The relations are categorized based on the level and type of entities to which they connect; for example, instance-level, type-level, and instance-type relations. The entities represent both material and immaterial objects that are categorized into two main classes, i.e., continuant and continuous.

The continuant entities continue their existence through time, while concurrent entities require temporal parts for their existence, such as events and processes. For instance, a human is a continuant entity, but aging is a process that is a concurrent entity (Arp, Smith, & Spear, 2015). The continuant entities are further categorized into dependent and independent. Dependent entities can represent their entities; whereas, independent entities require another entity in order to exist. For example, “Person” is an independent entity, and “Healthy Person” is a subclass of “Person” and is related to the “Healthy”-dependent class. The “Healthy” entity does not have its own identity without requiring another entity to create a unique entity at an instance level. The hierarchy of BFO 2 is shown in Figure 1 (B Smith et al., 2015). The concurrent concept is the top parent class of process, process boundary, temporal region, and spatiotemporal region.

This study selects BFO as the foundation framework of the study, due to its multiple advantages as listed below:

- Small and domain-neutral
- Used by more than 300 ontologies-driven endeavors worldwide
- Has active user groups
- Provides supportive documents and training
- Actively improves

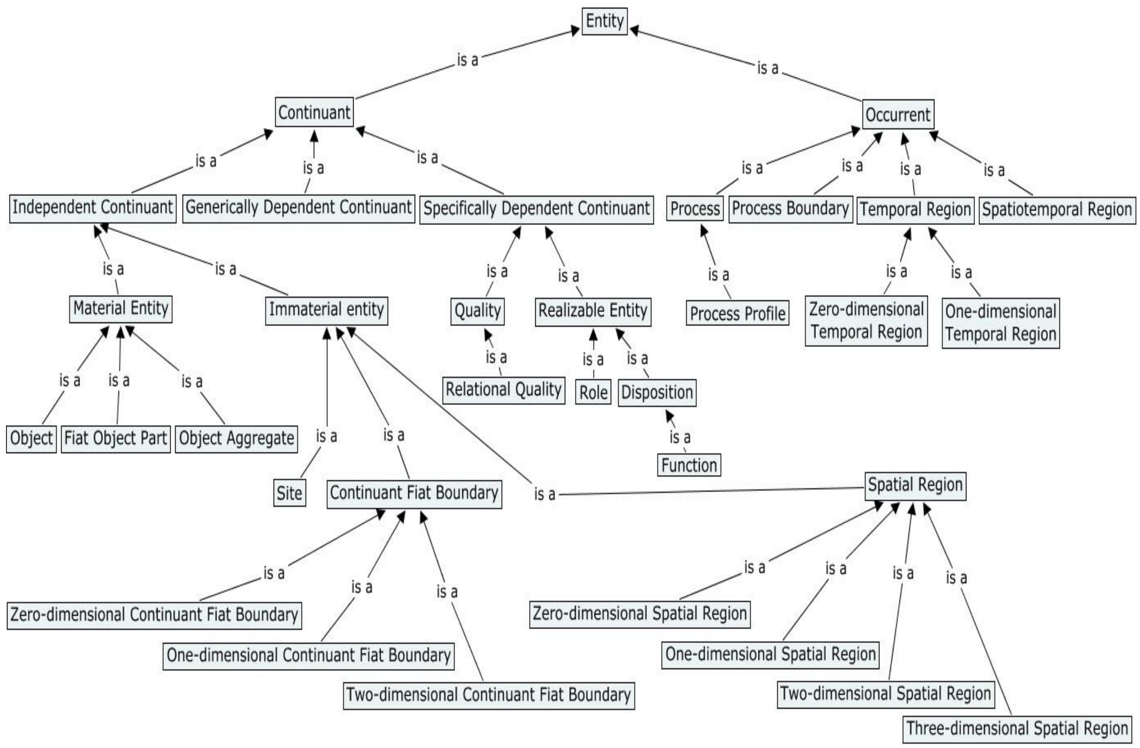


Figure 1 The hierarchy of BFO 2 (B Smith et al., 2015)

3.1.1 Ontology for Biomedical Investigations (OBI)

The OBI is the first ontology that provides a single resource for multiple types of experiments (Bandrowski et al., 2016). It follows the formal structure of the BFO framework with the ability of knowledge capturing representation that is extended from the Open Biological and Biomedical Ontologies (OBO) (Bandrowski et al., 2016). The OBI's primary purpose is to structure all aspects of the investigation process and all phases of scientific-investigation processes, including study design, protocols and instrumentation, data, and analysis in the biological and medical domains (https://www.ebi.ac.uk/ols/ontologies/obi). Besides, the OBI directly imports many formal terms of other prevalent ontologies such as Relations Ontology (RO), Gene Ontology (GO), Phenotype Attribute and Trait Ontology (PATO), and Chemical Entities of Biological Interest (ChEBI). Under the 'Material Entities' class in the BFO framework, many of OBI child classes are reused of external ontologies; for instance, the ‘organism’ class from the MCBI taxonomy, the ‘gross anatomical part’ class from the Uber Anatomy Ontology (UBERON) (Mungall, Torniai, Gkoutos, Lewis, & Haendel, 2012) and the Common Anatomy Reference Ontology (CARO) (Haendel et al., 2008). The OBI provides more than 2,500 terms, 84 individuals, 40 relations used in scientific-investigation areas, such as information, objective, planning, execution, and reporting process. Some examples of the OBI’s new classes are processed material, specimen, and organization. The ‘processed material’ class is a parent class of ‘processes specimen’ and ‘device’ classes. From the ‘Planned Processes’ concept in the BFO, OBO added many time-related concepts like investigation, collection, assay, research enrollment, and material maintenance and processing.

In this dissertation, the reused concepts from the OBI are the ‘Investigation’ and ‘Data Transformation’ planned processes. The “Investigation” process has two primary relations, “has specified input” and “has specified output’ relation connecting to continuous entities (Kong, Liu, & Wang, 2011). Additionally, the “Investigation” process can have a relation with another planned process, such as “Study Design Execution” and “Drawing a Conclusion From Data,” as shown in Figure 2. The ‘data transformation’ process is a process with the raw-data input and the ‘analyzed-data’ output, which participates in both EBP and PBE ontology models.

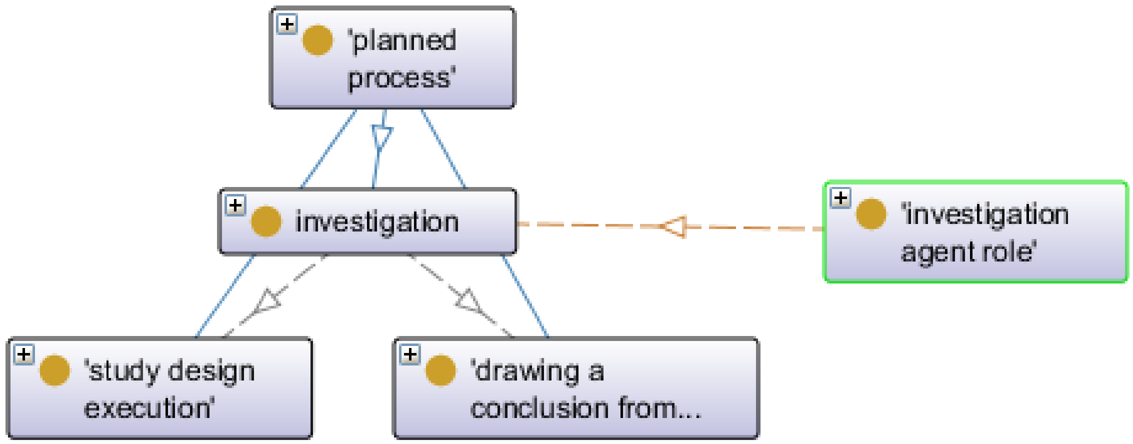


Figure 2 OBI's investigation-related process

3.1.2 Information Artifact Ontology (IAO)

The IAO is an extension of OBI in the efforts to represent the additional aspect of information entities (Barry Smith et al., 2015). The principal concept of the IAO is the information content entity describing all information-related entities in scientific domains, including data, investigation process, and the outcomes of data analysis. The information-content entity focuses on data collections and associated representational artifacts. Its subclasses extend to data-item, directive information entity, and document, as shown in Figure 3 (Ceusters, 2012). The information-content entity is a generically dependent continuant entity that is about another entity. Furthermore, the deeper level of IAO hierarchical classes include directive information, entity, plan specification, algorithm, protocol, and study design. The document class defines the repository of information like publications, articles, and reports.

Although IAO is a domain-neutral ontology, a domain-specific IAO can be further designed for more details for a specific need, such as IAO-Intel. The IAO-Intel is an extended IAO developed to support the needs of the US Army intelligence community by structuring formal controlled vocabulary for metadata about documents used in multiple military registries within the framework of the distributed Common Ground System (Smith et al., 2013).

In this dissertation, the information-content entity plays an essential role in data collection during the health-intervention and data extraction processes. Due to its domain-neural characteristic, the IAO capably involves in both EBP and PBE ontology models.

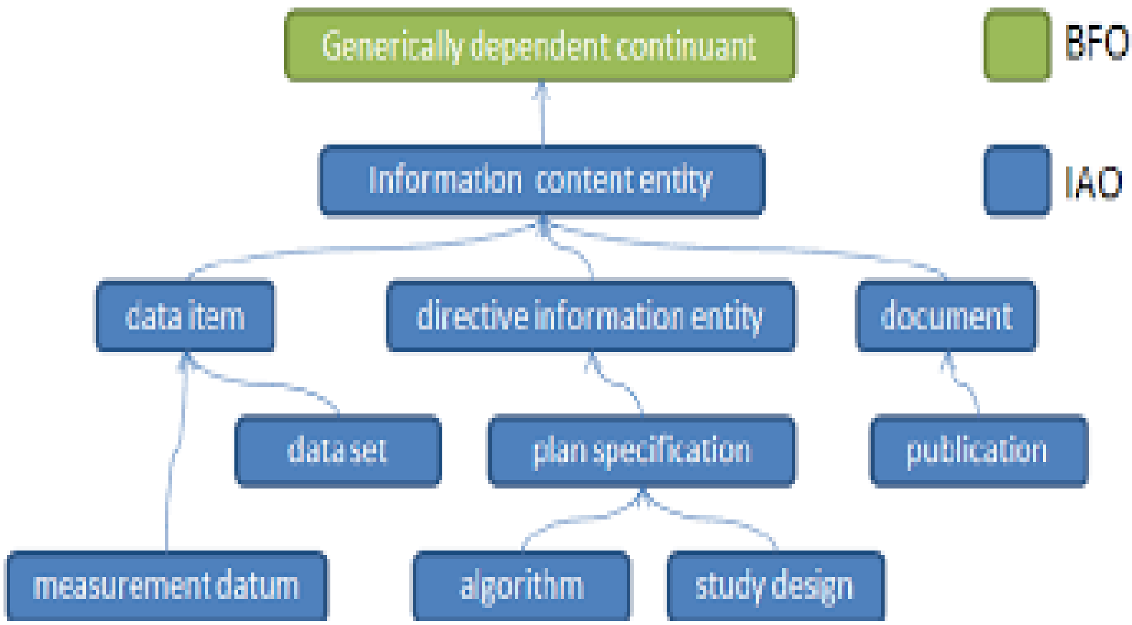


Figure 3 IAO structure (Hastings, Batchelor, Neuhaus, & Steinbeck, 2012)

The high-level structure of data, information and investigation process combining from OBI and IAO under the BFO framework as be presented in Figure 4.

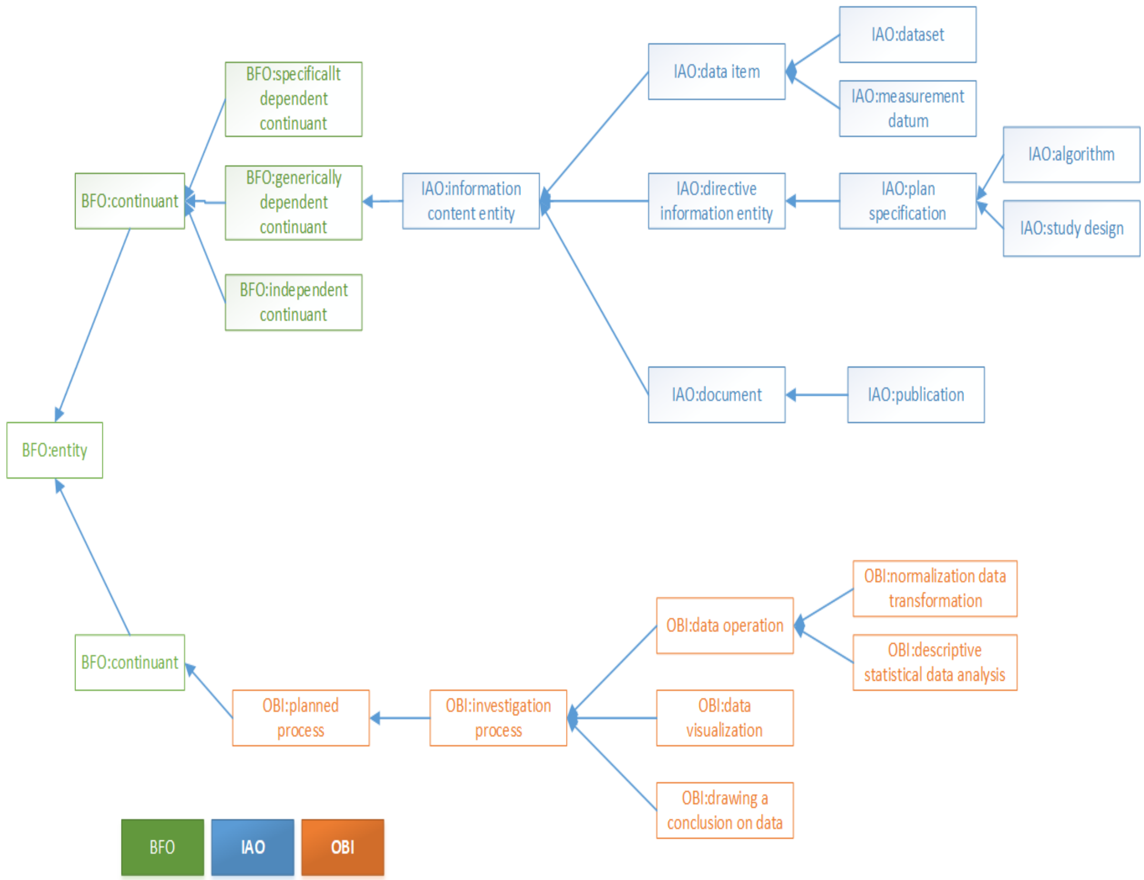


Figure 4 The high-level structure of OBI and IAO under the BFO framework

3.2 Methods for Ontology Implementation

The design of ontology depends on the need and purpose of individual application and implementers. This section presents the guideline steps of the implementation of an ontology model. There are various existing ontology models implemented in healthcare using either top-down or bottom-up approaches. This study implements the domain-specific ontology model with both approaches in order to combine their strengths. The bottom-up approach allows the domain experts to work closely with the implementation. As a result, the ontology explicitly defines the domain terms and their relations. On the other hand, the top-down approach offers the standard framework of the shareability to individual domain ontologies. The implementation steps of both approaches are the same in the beginning stages, such as determining domain scopes and terminologies. Their difference is primarily due to the structure framework: class and sub-class. At the end of this section, the study summarizes how the framework can influence the efficiency of ontology models, and what the effective structure the study purposes for the best practice setup.

In this study, the ontology development is a combination of the Toronto Virtual Enterprise (TOVE) methodology (Grüninger, Atefi, & Fox, 2000) and the simple knowledge-engineering methodologies (Noy & McGuinness, 2001). The TOVE project is a framework of enterprise integration (EI) for high levels of productivity, flexibility, and quality (Fox, Chionglo, & Fadel, 1993) which emphasizes on Ontology development and can be embedded with the clinical workflow (Jones, Phalakornkule, Fitzpatrick, Iyer, & Ombac, 2011). With TOVE guideline, the knowledge model can (1) define precise and unambiguous terms, (2) provide a shared terminology, (3) be able to automate answers of underlying questions and (4) define a graph visualization of terms (Fadel, Fox, & Gruninger, 1994). As well, Noy and McGuinness suggested an iterative methodology starting with a foundation concept and then be more details and more complex through the repetitive processes. The ontology editor used in the project is Portege 5.5 by Stanford University. This version offers reasoning tools such as DL queries, SWRL rule, and HermiT reasoning. The complete steps of the ontological-model can be summarized into ten steps, as listed below:

1. Determine the domain and scope of an ontology model
2. Define the terminology of the domain
3. Define the informal definitions of the terminology

4. Define the ontological definitions and constraints of the terminology
5. Consider reusing existing ontologies
6. Define the class hierarchy
7. Define the properties of the classes
   - Define the data properties or relations between classes
   - Define the object properties or slots of the classes
8. Create instances
9. Develop rules and run reasoning
10. Test the competency of the ontology through queries

All four individual ontology models of this dissertation follow the ten-step implementation process, even though each model emphasizes each step in different weights. For instance, the first model, the domain-specific ontology in the knee-treatment specialty, emphasizes on step1 to step 3, but less on step 4 and step 5 because it aims to capture and represent the focusing domain. On the other hand, for Aim 2, the implementation of statistics ontology does not involve step 1 to step 4 because the model reuses two existing ontologies, i.e., OBCS and STATO. The reusing of existing ontologies belongs to step 5. Nevertheless, all models in the dissertation apply step 10 to test and validate the individual model's competency and consistency. The details of the dissertation's aims are explained in the following section.

3.3 Aims and Deliverables

3.3.1 Aim1 Domain-Specific Ontology

3.3.1.1 Introduction

The dissertation selects the ACLR area as its specific domain due to the involvement in both evidence-based practice and practice-based evidence in ACLR-related research. The primary purpose of ontology implementation for a domain-specific ontology is to present its formal structure of domain knowledge and define the domain's terminologies, without attention to a broader framework (Shaw, Detwiler, Brinkley, & Suciu, 2008). However, a more significant advantage of ontology is about semantic-information sharing. For that reason, the domain ontology for a specific aim can increase productivity, reusability, and shareability by using a foundation framework that provides general-purpose definitions (Faber, Mairal, & Magaña, 2011). In Aim 1, both bottom-up and top-down approaches are used for the final domain-specific model. Both approaches provide a unique set of advantages that can be merged into a single framework. The last two steps in Aim 1 involves the validating tool of the model's competency using reasoning and DL queries.

3.3.1.2 Methods

The methodology is a combination of Toronto Virtual Enterprise (TOVE) and a simple knowledge engineering methodology. The ontology model's editing tool is Portege 5.5, which was developed by Stanford Center for Biomedical Informatics Research, including the Semantic Web Rule Language (SWRL), DL queries, and reasoning. After the domain's scope is well defined, the terminologies are represented in both informal and formal definitions. The formal definitions are defined in the predicate format. The first approach of the ACLRO domain-specific ontology is bottom-up. The second design applies the top-down approach to the top of the first one to transform the model into a sharable sematic-information model under the BFO framework.

3.3.1.3 Results

The efficiency of the domain-specific ontology is validated through two DL queries. Without the top-level framework, the first design with the bottom-up approach can merely represent the domain in the formal structure but lacks the shareability and the representation of semantic information. With the BFO framework, the study successfully implements a sharable model that is automatically classifying patient groups and terminology mapping between the descriptive terms to both ICD9 and ICD10.

3.3.1.4 Conclusion

The domain-specific ACLRO is a proof of concept that demonstrates how the ontology can be applied to a real-world setup. The ACLRO constructs the class of health-intervention processes that allow the ontology to share knowledge explicitly from technology. The model successfully defines the knowledge of the specific domain in a formal structure with the semantic definition. With the BFO framework, the ALCRO domain ontology is sharable and compatible with other ontologies that make the ACLRO more meaningful.

3.3.2 Aim2 Statistic Ontology

3.3.2.1 Introduction

Statistics is the branch of science that uses quantitative or mathematical methods to analyze data. The roles of statistics involve collecting, summarizing, presenting, and drawing a conclusion. With the fast progressing in technologies, the need for statistics is rising along with Big Data. Due to its growing demand, the implementation of various statistics applications and software are offered to statisticians as well as non-statistician researchers (Ocaña-Riola, 2016). However, the description and explanation of statistical methods and research processes are not clearly documented, which prevents the validation and reproducibility for statistical analysis in clinical research (Strasak, Zaman, Pfeiffer, Göbel, & Ulmer, 2007). Therefore, the majority of publications and external evidence are challenging to reapply their methods in a different environment due to the lack of information on study design, statistical criteria, and algorithms (Zheng et al., 2016). The goal of Aim 2 is to construct a standard statistical ontology for scientific research. After reviewing the existing ontologies, the structure of the statistic ontology begins with consolidating existing BFO-based statistical ontologies. The dissertation's statistics ontology is designed to be universal and sharable in both PBE and EBP studies while offering a deeper connection to the statistical theory and research designs.

3.3.2.2 Methods

The method in Aim 2 significantly involves in step 5 of the TOVE methodology, i.e., reusing exiting ontologies. The initial step is to search on all existing statistical ontologies; then perform a comparison on the popularity, structured framework, and maturity of the models. After reviewing, the two appropriate models are STATO and OBCS under the BFO framework and a part of OBO Foundry. The next step is a gap analysis of these two models structure for the similarities and differences. The consolidation of both STATO and OBCS provides a solid structure of the statistics model in this study. The additional statistical terms are added to extend the scope of statistical models and mathematics theories, including statistics parameters and assumption criteria.

Lastly, the model competency is validated through multiple DL queries across various components of statistical concepts.

3.3.2.3 Results

The study's statistical vocabularies and concepts are imported from two statistic-related ontologies, i.e., OBCS and STATO. Each ontology suits the study needs in different aspects. For instance, STATO offers the ‘hypothesis’ class, which is essential to ACLRO, while OBCS only contains the ‘null hypothesis’ subclass. STATO includes more based on research aims and statistics models, such as ‘goodness of fit hypothesis’, ‘presence of association hypothesis’, and ‘absence of difference hypothesis’. This study reuses multiple terms from OBCS that are not available in STATO. Some examples are terms under ‘statistical model’ and ‘statistical variable’ classes. In addition to OBCS and STATO, ACLRO also develops more specific terms to meet the study’s needs. For instance, the ‘healthcare variable’ subclass is added to the ‘variable’ class.

3.3.2.4 Conclusion

Under Aim 2, the statistics ontology model consolidates existing concepts from two OBO Foundry ontologies; OBCS and STATO. Furthermore, the study adds new statistical terms to the model and expands to more advanced mathematics like machine learning to be compatible with the growth of healthcare AI.

3.3.3 Aim 3 Evidence-Based Practice Ontology

3.3.3.1 Introduction

Evidence-Based Practice (EBP) is an approach that aims to integrate the best external scientific evidence into local patient care (Howland, 2007). The most apparent external evidence is derived from research data. The ultimate objective of EBP is to improve the effectiveness and quality of patient outcomes through new guidelines or decision-making (Melnyk, Gallagher-Ford, Long, & Fineout-Overholt, 2014). Nevertheless, such achievement can be challenging due to a lack of understanding and differences in clinical settings between EBP's source and the local practice.

Aim 3 purposes the formal structure of EBP ontology under the BFO framework providing the semantic-information shareability. The EBP model's scope includes statistical models, variable characteristics, study hypothesis, research methodology, and publication information. Each concept in the EBP model can be further connecting to other concepts, including in other ontology models such as the Practice-Based Evidence ontology.

3.3.3.2 Methods

Aim 3 also follows the TOVE methodology, which is mentioned in the Methodology section. After defining the scope and reviewing existing ontology models, the formal definitions are defined in the predicate format; then, the concept mapping diagram is created before transforming to OWL language though the Protégé tool. Later, two instances of publications related to return-to-sport after ACLR surgery are added to the EBP model for the validation process. Lastly, the validation process utilizes DL queries to retrieve information related to statistical models and publications in the ACLR domain.

3.3.3.3 Results

The DL queries of the EBP models in Aim 3 successfully review the structure and relations of the statistical model and the study outcomes in publications. The knowledge gained in both publications can be combined and listed together for the query correctly.

3.3.3.4 Conclusion

The study presents the implementation of an evidence-based evidence ontology in the ACLR rehabilitation domain while complying with the shareable semantic ontological framework. The two publications are recorded at the instance level for validation purposes. The DL queries effectively retrieve information presented in the two publications along with the surplus information from the statistics model.

3.3.4 Aim 4 Best-Practice ACLR Ontology (ACLRO)

3.3.4.1 Introduction

Both EBP and PBE approaches aim to improve patient outcomes focusing on different sources of knowledge. The EBP approach applies scientific evidence as to its guidance and decision support to interpret the best suitable evidence from systematic research into clinical expertise and the environment (Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996). Alternatively, the PBE approach is a relatively new procedure for gathering good-quality data from routine practices in real-world settings with trial and error, initiating innovation and knowledge discovered in healthcare with objective support based on community values (Evans, 2000). The combination of EBP and PBE approaches enhances the quality of patient care. For this reason, the best practice (BP) model in Aim 4 integrates the domain-specific knowledge in Aim 1 and the EBP model in Aim 3 into one single that represents the reusable and sharable ontology framework, obtaining external and internal evidence from both approaches.

3.3.4.2 Methods

With the BFO framework, the best-practice ontology model does not have an implementation of a new concept. Consequently, Aim 4 does not follow all steps in the

TOVE methodology. Two main steps in Aim 4 are the merging of the EBP and PBE model and the validation of the final BP mode. Then, the concept diagram of the best practice model shows the successful integration between the EBP and PBE models.

3.3.4.3 Results

All sub-models in Aim 1, Aim 2, and Aim 3 are successfully merged into a single framework of the best practice ACLRO. The DL queries in Aim 4 successfully retrieve information existing across multiple ontology models, including the domain-specific, statistics, EPB, and the case study in the PBE models.

3.3.4.4 Conclusion

Combining the PBE and EBP processes allows an individual practice to incorporate their scientific learnings and best external evidence for the most optimal outcomes as the best practice approach. The best-practice ACLRO model proves that with the BFO framework, the individual models of the specific domain, statistics, practice-based evidence, and evidence-based practice are reusable and sharable with sematic information. Likewise, the study enriches the ontology model's functionalities, allowing knowledge queries across multi-disciplinary areas, such as medical data, statistical models, study design, and scientific knowledge.

3.4 Case Study

This section presents a case study on transforming a process of statistical data analysis process into knowledge discovery that can be queried in ACLRO. The objective of statistical analysis is to measure the relative feature importance of each variable in the dataset on the prediction of two-year postoperative knee pain after ACLR. The selected statistics algorithm used in this case study is one of the most popular and influential machine learning models, random forest, which is explained more later in this chapter. First, the study introduces the dataset and then explain the random forest algorithm. Next, the study presents the formal ontology model representing the complete flow from study design to knowledge capture using real patient-care data.

3.4.1 Dataset

The dataset contains 927 observations of patients who received ACL reconstruction between 1982 and 2018 performed by the same surgeon. The study would exclude patients if they had bilateral ACL tears, undergone revision surgery, or received additional surgical treatments on the knee. The dataset is composed of eight variables as described below:

Pain

ACLR is an elective procedure. ACL tear is not life-threatening. However, it affects the quality of life due to pain and limitation in activities. Therefore, the patient's subjective evaluations are critical and serve as an indicator for surgery success. In this study, the subjective pain score at the 2-year postoperative period serves as the primary outcomes and the study's dependent variable. The score is a Likert scale ranging from 0 to 10, which is considered as a continuous variable since it belongs to the interval scale (Allen & Seaman, 2007).

Inj_typeInj_type represents the type of knee injuries causing an ACL damage. The injuries were classified as acute if patients had not experienced any giving-way episodes after the index injury and before surgery. The injuries were classified as chronic if patients had any additional giving-way episodes after the index injury and before surgery.

Injsrg_mth

Injsrg_mth refers to the period between injury and surgery in months. This is a continuous variable defining the number of months between injury and surgery time.

Srgage

Srgage denotes the age of patients at the time of surgery. It is a continuous variable.

Sex

Sex defines patients' biological characteristics as female or male.

Cartilage

According to the articular cartilage status at the time of surgery, the dataset was categorized into these four primary groups. The patients were grouped as "normal" if they had grade 2 or less articular cartilage damage in all compartments; they were grouped as "damaged" if they had grade 3 or 4 chondromalacia in any compartment.

Med_rem

Med_rem indicates the procedure performed on medial meniscus. The patients were categorized into two medial-meniscus groups based on the status at the end of the ACLR surgical procedure. If medial meniscus partially or entirely removed, a patient was placed into the removal group; otherwise, he or she was added to the intact group.

Lat_rem

Lat_rem represents the procedure performed on lateral meniscus. The patients were categorized into two lateral-meniscus groups based on the status at the end of the ACLR surgical procedure. If lateral meniscus partially or entirely removed, a patient was placed into the removal group; otherwise, he or she was added to the intact group.

3.4.2 Random Forest (RF) Model

RF is one of the most popular ML methods due to its accuracy, robustness, and ease of use. RF is an ensemble-learning algorithm consisting of many decision trees and trained with the bagging technique and a supervised learning algorithm that takes a known set of input and output datasets to learn the model. The bagging or bootstrap method aggregates performances from multiple machine learning algorithms to the final model to improve accuracy. The model first splits the dataset into train and test datasets. The training dataset allows the model to learn the mapping, while the test is used to validate the model (Andy Liaw & Wiener, 2002). The RF can be used for both classification and regression problems. Since the 'Pain' score in this study is considered as a continuous variable, the regression RF is utilized.

3.4.3 Modeling in R

R is a language and environment for statistical computing and graphics developed at Bell Laboratories by John Chambers and colleagues (www.r-project.org). One of the great benefits of R is that its functionalities can be extended via packages. The two main packages applied in this case study were caret and randomForest. The caret package allows ML to split data into the train randomly (70%) and test randomly (30%) datasets

(Kuhn, 2015). The randomForest package provides the random forest algorithm and visualizations for the predictive model (A Liaw & Wiener, 2018). The complete R-code can be seen in Appendix A.

Big data's characteristics are large volumes, velocity, and variety, which make big data complex and hard to analyze. Healthcare AI’s primary aim is to reveal patterns, discover trends, as well as to identify unknown data behavior and interactions. RF’s feather-importance method is one of the simplified methods for feather selection in big data. In this case study, ACLRO only built a formal representation of feature-importance calculations. The Mean Decrease Accuracy ($incMSE) calculated in the randomForest package was used to determine the variable importance. First, the $incMSE starts with calculating the mean squared error (MSE) of the whole model. Then, the ML permutes each variable in the model to calculate the new model MSE according to variable permutation. Next, the difference between the original model MSE and the new model MSE is measured and compared. Last, the ML ranks the variable’s importance in agreement with the value of the %incMSE. The higher value of %incMSE indicates a more prominent feature.

The %incMSE in this case study showed that the number of months between injury and surgery (injsrg_mth) turned to be the most important feature contributing to the two-year postoperative pain, following by the status medial meniscus removal (med_rem), patient sex (SEX), cartilage damage (cartilage), injury type (inj_type), lateral meniscus removal (lat_rem) and age at surgery (srgage) as presented in Figure 5. The calculated value of the mean decrease accuracy is listed in Table 1.

Feature	Mean Decrease Accuracy (%incMSE)
injsrg_mth	13.84614260
med_rem	12.07007574
SEX	8.24495552
cartilage	6.67335309
inj_type	4.31484798
lat_rem	-0.07232279
srgage	-0.48011320

Table 1 Mean decrease accuracy of two-year postoperative pain

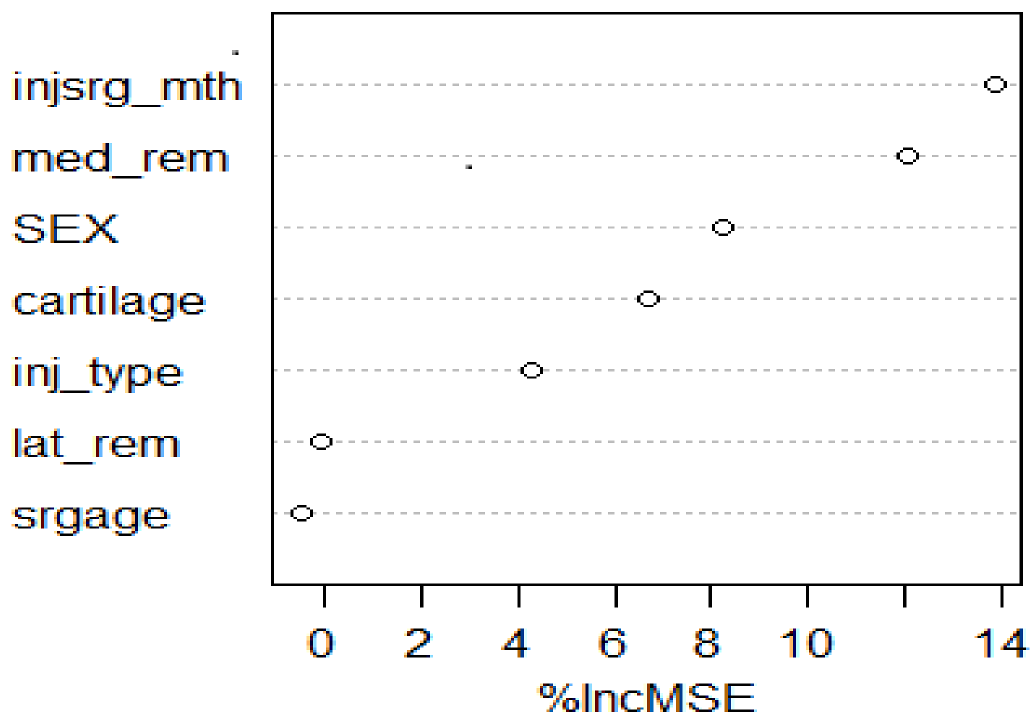


Figure 5 %incMSE of different variables on two-year postoperative pain

3.4.4 Formal Representation

The formal representation of the research process starts at the ‘investigation’ planned The formal representation of the research process starts at the ‘investigation’ planning process, which is defined as a planned process that consists of parts: planning, study design execution, documentation, and which produce conclusion(s) (http://purl.obolibrary.org/obo/OBI_0000066). Here, ACLRO only focuses on ‘study design execution’ and ‘drawing conclusion’ processes. The other two processes were omitted since it is not a focus of this case study. The ‘study design execution’ process carries out a study design. The two key components of the study-design process are to select a suitable statistics algorithm and to produce statistic outcomes. The algorithm requires a set of various variables. Each variable has a unique set of attributes. For instance, the attributes can define dependent vs. independent or continuous vs. categorical variable types. The ‘study design execution’ process outputs are the derived data from statistical analysis that belongs to the ‘data item’ class. Here, this statistic outcome is the calculated value of increase mean accuracy (%incMSE) reported in RF. The researchers use this information to draw the study conclusion. The conclusion can be validated and denotes the knowledge discovered from the study. As in the study’s case study, the values of %incMSE are injsrg_mth (13.84), med_rem (12.07), SEX (8.25), cartilage (6.67), inj_type (4.314), lat_rem (-0.072), and srgage (-0.480). The conclusion drawing from the analysis outcome indicates that out of seven variables, the waiting period for the surgery and the removal of medial meniscus variables are the most two factors contributing to the 2-year postoperative knee pain. In contrast, patient age and the removal of lateral meniscus do not impact the postoperative knee pain. Finally, the conclusions denote the knowledge about postoperative knee pain, i.e., 1) the postoperative knee pain is impacted by surgery waiting time; 2) the postoperative knee is impacted by medial-meniscus removal procedure.

The visualization of the overall formal ontology structure is in Figure 6.

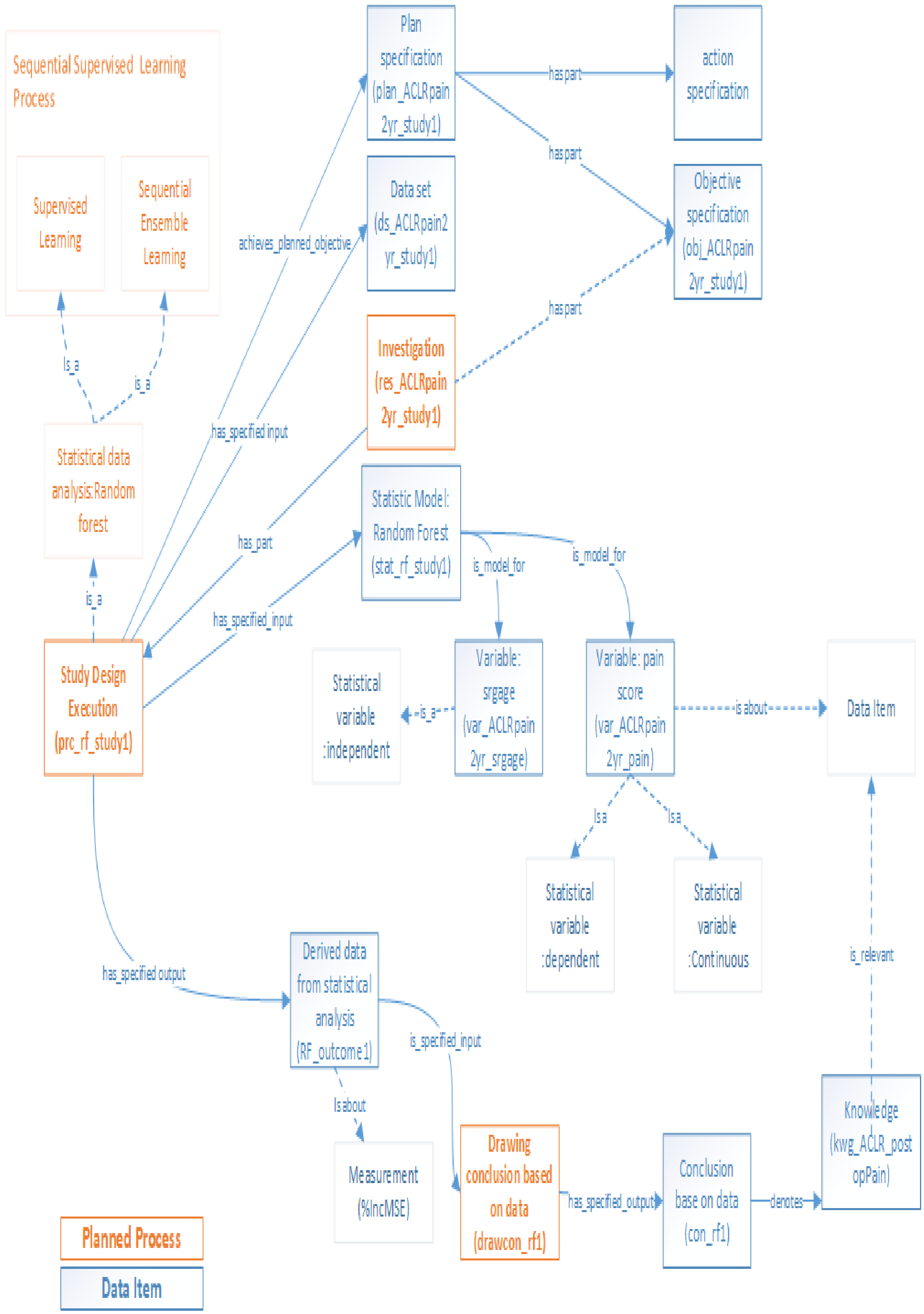


Figure 6 The visualization of the overall formal ontology structure

3.4.5 Validation

Scenario-test techniques check the model's consistency and check if the model satisfies the study's purpose. In detail, the techniques use multiple scenarios that share commonalities and differences. The results should show consistency as it relates to common scenarios while differentiating the others. The study will sample some DL queries on the relationship between statistics and research designs to discover and represent the relationship, knowledge, and publication. The DL query allows ACLRO to confirm its model structure and class's definition (Van Der Straeten, Mens, Simmonds, & Jonckers, 2003). Some examples are:

1. Which response variables have been studies in ACLR using the Random Forest model?

   DL query: 'dependent variable' and 'is about' some ACLR_surgery and 'is modeled by' some 'random forest.'

   Result: All response variables in the random forest for ACLR-surgery related studies are selected. Based on the use case, the study's pain variable was selected.

2. Which variables were discovered as the top-five influential factors on post- ACLR operative pain?

   DL query: 'independent variable' and (is_ranked some xsd:integer[<= 5]) and ('is modeled by' some ('random forest' and 'is model for' some ('dependent variable' and 'is about' only 'pain score' and 'is about' some ACLR_surgery)))

Result: The model can select all predictors as a part of random forest models in any ACLR-surgery related studies. In the use case, the query selected all top-5 influential factors: injsrg_mth, med_rem, sex, cartilage, and inj_type, as in Figure 7.

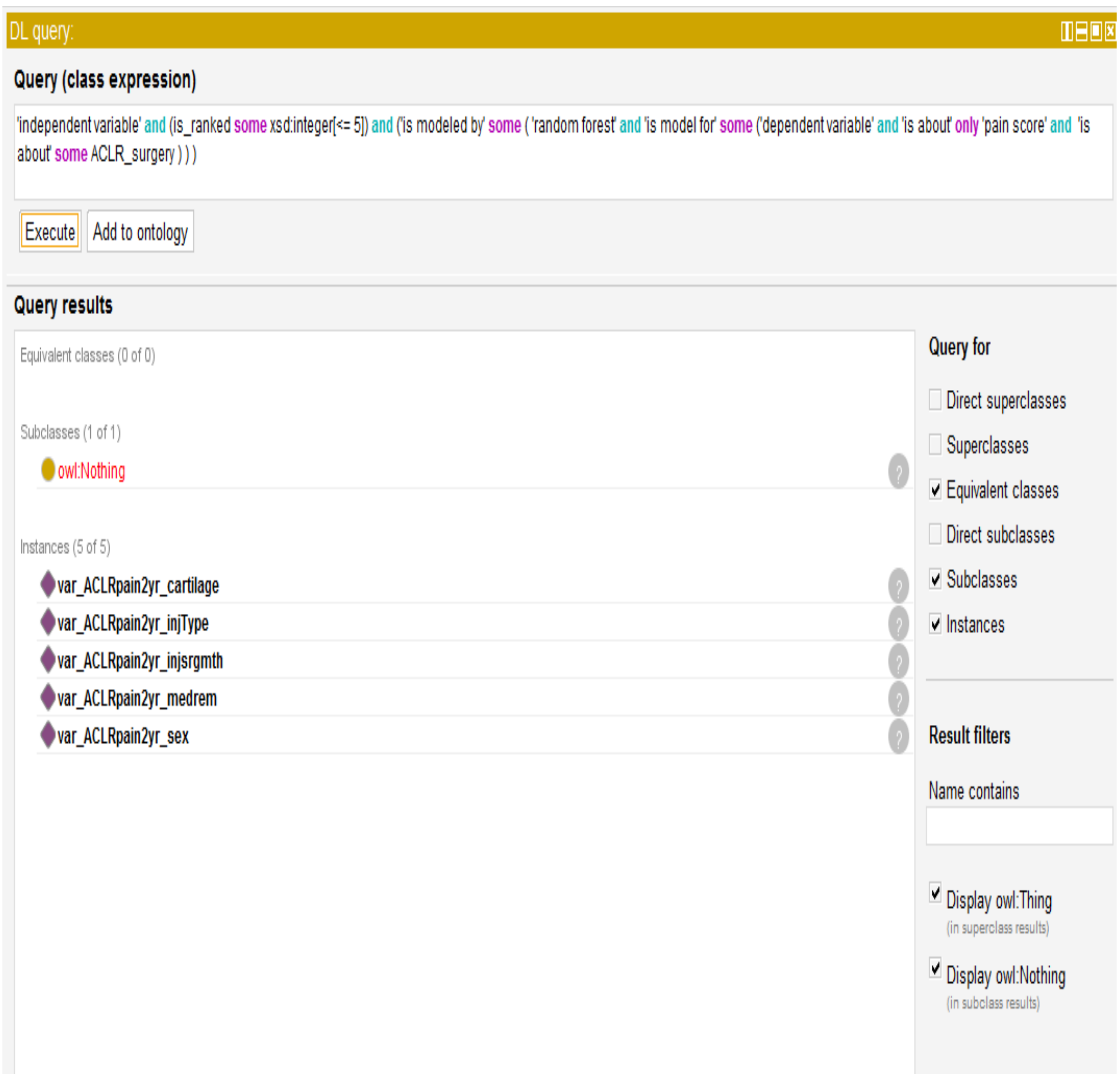


Figure 7 DL query: the top-five influential factors on ACLR postoperative pain

3. Which predictors were discovered at least as the top-five influential to post- ACLR operative pain base on the increase mean accuracy calculation?

   DL query: 'independent variable' and ('is part of' some('derived data from statistical analysis' and ('is about' some incNodePurity))) and (is_ranked some xsd:integer[<= 5]) and ('is modeled by' some ('statistical model' and 'is model for' some ('dependent variable' and 'is about' only 'pain score' and 'is about' some ACLR_surgery)))

   Result: The model can select the list of all predictors as a part of the mean increase accuracy in any ACLR-surgery related studies. In the study’s use case, the query selected all top-5 influential factors: injsrg_mth, med_rem, sex, cartilage, and inj_type, as in Figure 8.

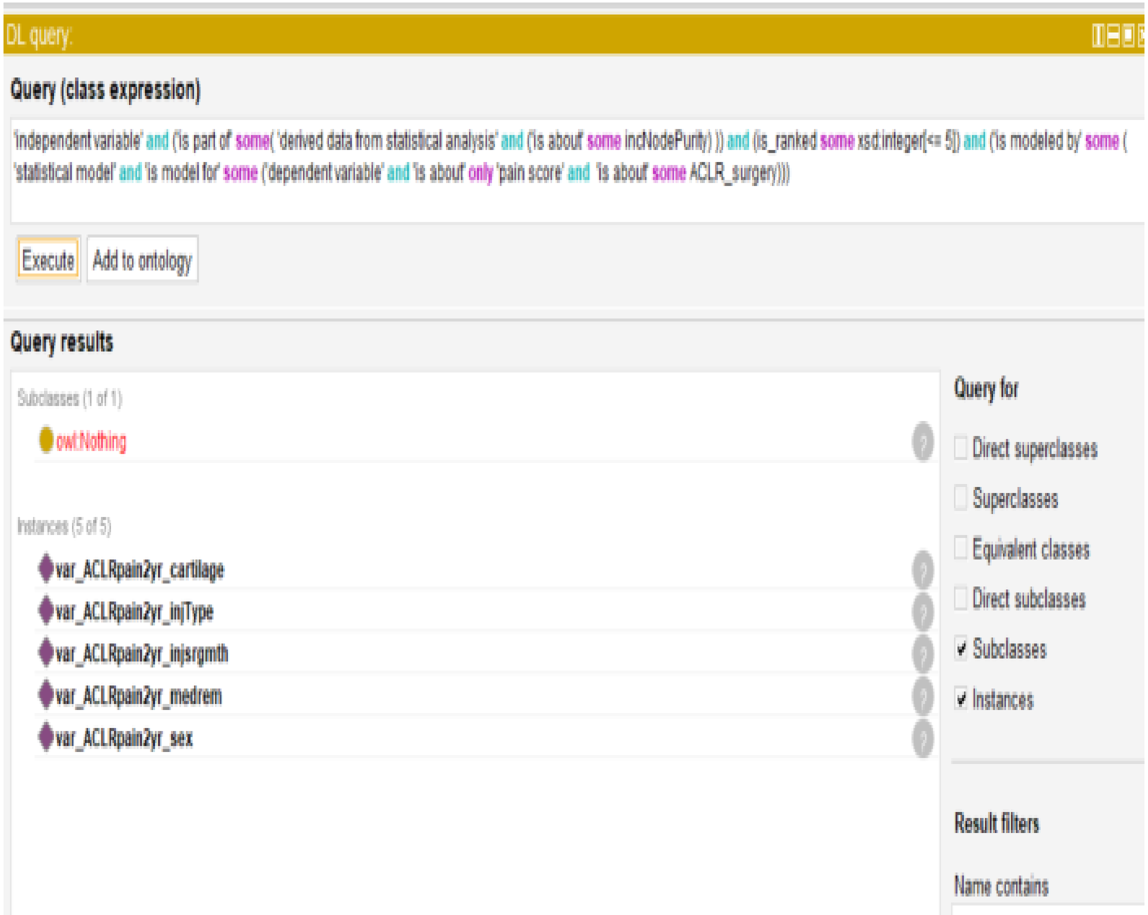


Figure 8 DL query: the top-five increase mean accuracy factors on ACLR postoperative pain

3.5 Conclusion

The study proposes the hybrid approach using the bottom-up and top-down approach with the foundation ontology to improve the shareability and reusability of the ontology model. The bottom-up approach is easier to use a starter for domain-knowledge abstracting, while the top-down approach requires a new understanding of the foundation ontologies. The final best-practice model revealed a unique benefit of ontological knowledge engineering that permitted EBP, PBE, domain, and statistic models to be built separately and then integrated into the final model without re-work on any existing models. This work proved that the ontology model with the foundation structure could be reusable and shareable.

CHAPTER FOUR AIM 1 DOMAIN-SPECIFIC ONTOLOGY

4.1 Introduction

The ACLRO serves as proof of concepts, demonstrates how the ontology can be applied to a real-world setup while bringing benefits as claimed. The study proposes a practical technique to represent the patient-focus process that utilizes both bottom-up and top-down approaches. The purpose of the hybrid approach is to apply the standard framework of the top-down approach with the specific domain knowledge obtained in the bottom-up approach. The foundation framework helps the domain ontology to define a more explicit structure with fewer adding relationships. As a result, the domain ontology's maintenance is more straightforward in expanding and integrating more concepts. The outcome of this work shows the hybrid approach's success that enhances the representation of domain knowledge while improving the reusability and semantic interoperability of the domain ontology.

4.2 Existing Ontologies

Reusable ontologies can be categorized into two main areas: general ontologies and domain-specific ontologies. The examples of general ontologies are DMOZ (www.dmoz.org) and WordNet (www.cogsci.princeton.edi). DMOZ was a multilingual open-content directory of World Wide Web links based on a hierarchical ontology scheme. Its purpose was to organize site-directory listings into categories (https://en.wikipedia.org/wiki/DMOZ). WordNet is a lexical database with Ontology knowledge structure for the English language (https://wordnet.princeton.edu/). It acts as a combination of dictionary and thesaurus and categorizes nouns, verbs, adjectives, along with adverbs using IS-A relationship. One of the most recognized systems for domain-specific ontologies is UMLS, which serves as a centralized resource for medicine, health, and biomedical vocabularies and standards for interoperability between different EHR systems. Another well-known domain-specific ontology is Gene Ontology (GO: http://www.geneontology.org). The GO offers the construction of concepts in the biology domain to describe gene function and their relationships. All of these above ontologies also offer tools or functions for downloading their terms and annotations.

There are multiple ways to search smaller existing domain-specific ontologies such as Protégé Ontology Library (https://protegewiki.stanford.edu/wiki/Protege_Ontology_Library), the semantic-web search engine called Swoogle (http://swoogle.umbc.edu), and Ontobee (www.ontobee.org).

Swoogle, developed by the University of Maryland, Baltimore County (UMBC), is a search engine for Semantic web ontologies, documents, and terms published on the web. It is a crawler-based indexing and retrieval system for the Semantic Web that extracts metadata and analyzes relations between its discovered documents (Ding et al., 2004). Swoogle offers the default search string to find relevant semantic web documents and metadata of RDF, document, and the basic semantic web. With the search “knee” and “def:knee” terms, Swoogle showed no results.

The Protégé Ontology Library is a part of the Protégé Wiki site that is organized into three groups: OWL ontologies, frame-based ontologies, and ontologies in other formats. The submitted ontologies are listed in alphabetical order. The search team of “knee” was performed with a return of zero results.

Ontobee is the default linked data server for biomedical ontologies in the Open Biological Ontology (OBO) Foundry library ontologies with the aims to facilitate ontology term dereferencing, data sharing, visualization, query, integration, and analysis using the RDF triple store technology (Ong et al., 2017). The OBO Foundry library represents a family of ontologies meeting its requirement on formal structure and interoperability (Smith et al., 2007). With the search term as “knee”, the result showed 19-matched term IRI from twelve ontologies. With the exclusion of six ontologies that focused on non-human species domains such as a mushroom, mouse, and Hymenoptera, the other six ontologies are listed below:

- BRENDA tissue/enzyme source (BTO)
- Human Developmental Anatomy, abstracted version (EHDAA)
- Human Developmental Anatomy, timed version (EHDA)
- Foundational Model of Anatomy Ontology (FMA)
- NCI Thesaurus OBO Edition (NCIT)
- Ontology for MIRNA target (OMIT)

These six ontologies offer a taxonomy of body anatomy but not the relations nor properties of the terms. After reviewing each ontology’s structures and aims, the study imported the ACL-related terms from FMA into the body-anatomy concept of the ACLRO, i.e., BodyAnatomy.

The search for more focused terms like ACL reconstruction, knee surgery, and knee rehabilitation did not show any existing ontology. Consequently, the study designs the structure of the domain from the ground up.

4.3 Bottom-Up Approach

The most advantage of the bottom-up approach is the ability to structure concepts, as understood in the domain expert's perspective. The bottom-up approach makes the implementation flexible and enriches the domain-specific knowledge representations (Van Der Vet & Mars, 1998). The processes also follow the TOVE guideline, starting from determining the domain and scope of the model. There are seven steps listed as a part of the bottom-up approach in order to capture and represent the concepts and their relations in the ACLR domain, as follows.

4.3.1 Determine the Domain and Scope of the Ontology

The ground success of ontology implementation is an ability to represent domain knowledge and satisfy the model's aims. The goal of our ontology implementation is to represent a knowledge structure of the best practice of rehabilitation after an acute cruciate ligament reconstruction (ACLR) procedure. ACLR is a surgical procedure on graft replacement of torn ACL in the knee joint. In this domain, the choice of graft replacement is autograft patellar tendon. Here, patient-care or health intervention events are data-driven, influenced by electronic health record systems (EHRs). The study follows the health-intervention process proposed by the American Nurses Association (ANA), which includes Assessment, Diagnosis, Outcomes/Planning, Implementation, and Evaluation.

After learning about the domain, the next step is to define the scope of ontology. From the experience of this ontology implementation, the most challenging task is in maintaining the scope of the domain. Remember that an ontology model is not capable of defining everything in the domain. A good ontology model is to include only matter factors while excluding the rest. Stanford recommends using a list of questions to limit the scope (Noy & McGuinness). This method proved successful in developing a government budgetary ontology (Brusa, Laura Caliusco, & Chiotti, 2008). The study also found that the technique is helpful because it is simple and can capture the crucial factors base on reality and domain experts. The list will make the model development be on track and realize what matters to the model. The followings are the samples of general questions for the project:

- What is the type of ACL surgery?

  The domain is the patient-focus process of rehabilitation of ipsilateral patella-tendon graft ACLR patients.
- What is the purpose of implementation? The purpose of implementation is to construct a knowledge framework for complete care
- How will the ontology be used?

  The model can be used to share a predictive model's outcome, such as influential factors at the time to return to sport.
- Who will get benefit from ontology, and how?

  Care providers, patients, and researchers, who want to use the predictive model to track patient outcome over time and help patients to reduce or avoid risks

4.3.2 Define the Domain Terminology

Uschold and Gruninger suggested another strategy to capture domain knowledge by brainstorming and grouping terminologies (Uschold & Gruninger, 1996). Also, the terminology can be abstracted from the complete set of questions, which are used to form a collection of formal terminologies in this study. For instance, the terms are patient, knee, knee motion, normal knee motion, subjective score, gender, female, male, ACL tear, ACLR failure, sport, activity level, ROM, and postoperative time. Besides, the study ranks the importance of the terms by the frequency of term occurrence in these questions. The rank of terms' frequency determines their involvement and the granularity level of entities in the model. Another technique in the study is to list ACLR-related terms of patient records and research studies. The list has later complied with their frequency rank in the excel spreadsheet. Then the list is categorized into groups of concepts, in which the study defines their informal and formal definitions in the next two steps.

4.3.3 Define the Informal Definitions of the Terminology

The terminology is defined for domain-specific interpretation by domain experts. For instance,

- ‘Patient’ is a person who has a health issue and is under medical care or treatment
- ‘Patient Age’ is a period of patient life from the date of birth to the recorded time measured in years.
- ‘Knee Injury’ is an injury to either or both knees

The informal definition of the health-intervention process are described below:

- History Health-Event

  In this step, a patient is interviewed for health history, demographics, and injury-related information.
- Symptom Health-Event

The patient is observed for physical signs and symptoms such as swelling and gait. The health event can be related to multiple body parts, which is not a case in the history of health-event.

- Evaluation Health-Event

  Multiple assessments are recorded for both subjective assessments, such as International Knee Documentation Committee (IKDC) subjective and Noyes, as well as objective measurements, such as KT-1000 for knee laxity; ROM; Cybex, and single-leg hop tests for knee strength. The data can be recorded, either an EHR system or other patient-related files. The raw data gets reviewed and turned to be information in the step. The evaluation can happen either before or after surgery. When it is a pre-operative evaluation, the values are used to determine the diagnosis. On the other hand, the postoperative evaluation allows the care providers to learn about the treatment outcomes.

- Diagnosis Health-Event

  The combination of the collected information in the Evaluation process and the care provider's knowledge leads to a conclusion of the patient's primary and secondary diagnosis in this step. The diagnosis is first recorded as the local practice's terms and then mapped to ICD-9 and ICD-10. The challenge of ICD-9and ICD-10 mapping is that the ICD-10 is more specific that benefits clinical decision making to the payer, for instance, the additional information on the body side and level of disease. For the ACLR domain, the primary diagnosis is either torn ACL on the left or the right leg: i.e., Diagnosis code ICD-9-CM 844.2 and ICD-10-CM S83.509A, S83.511 and S83.512). In this step, the information starts turning into knowledge.

- Treatment Health-Event

  The Treatment Health-Event is categorized into two types: 1) surgical treatment, 2) physical therapy or rehabilitation. All patients with torn ACL receive ACL reconstruction surgical procedure using patella-tendon autograft (CPT code: 29888). After the surgery, a rehabilitation program assists the patients to recover faster with better evaluations. The treatment can also require more assessments and evaluation afterward, which allows the care providers to learn about the patient's recovery progression. The data in the previous process gets compiled and evaluated for the patient's progress based on the treatments. The process usually is not a one-time process. The treatment can trigger a new set of signs and symptoms, creating a feedback loop to the workflow.

- Outcomes Health-Event

  The outcome here denotes the final incidents indicating if the patient successes the goals or fails the long-term goals. In this study, the authors use the ten-year follow-up as the outcome.

There are three main timeframes of ACLR: Pre-operative, Operative, and Postoperative.

- Pre-operative

After being evaluated with torn ACL, a patient will be enrolled in the pre- and postoperative rehabilitation. The condition of the torn-ACL knee before surgery is essential. The injured knee will be treated for swelling and evaluated to a range of motion (ROM) and leg-muscle strength comparing to the not-injured knee. Extension: less than 2 degrees difference between injured and not-injured knee.

- Operative or Surgery

  During the surgical procedure, a surgeon ensures that the ACL knee has a full ROM after graft placement and fixation. All ACLR surgeries used auto-graft patella-tendon procedures. This study is only based on the first unilateral ACLR surgery.

- Postoperative Surgery

  The postoperative periods are divided into one-week, two-week, one-month, two-month, 6-month, and one-year postoperative times.

4.3.4 Define the Formal Definitions and a Concept-Mapping Diagram

In this step, the previous informal definitions are transformed into the first-logic knowledge representation of concepts, which is also referred to as predicate logic. The format is developed as a sentence with three components: Subject -> Predicate -> Object (Phalakornkule et al., 2013).

A person-> has role -> Role(s)

A person-> has part -> Body Parts

Body Part -> experience -> Histories

History -> indicate -> Signs and Symptoms.

Signs and Symptoms -> assess -> Evaluations

Evaluation -> generate -> Diagnosis

Diagnosis -> identify -> Treatments.

Treatment-> impact -> Signs and Symptoms

Treatment -> produce -> Outcomes

The predicate logic can be further transformed into a concept-mapping diagram to see the flow of information and relations between concepts (Sowa, 2000). The concept-mapping diagram of health-event for ACL reconstruction care base on the ANA recommendation for the nursing care process is shown in Figure 9.

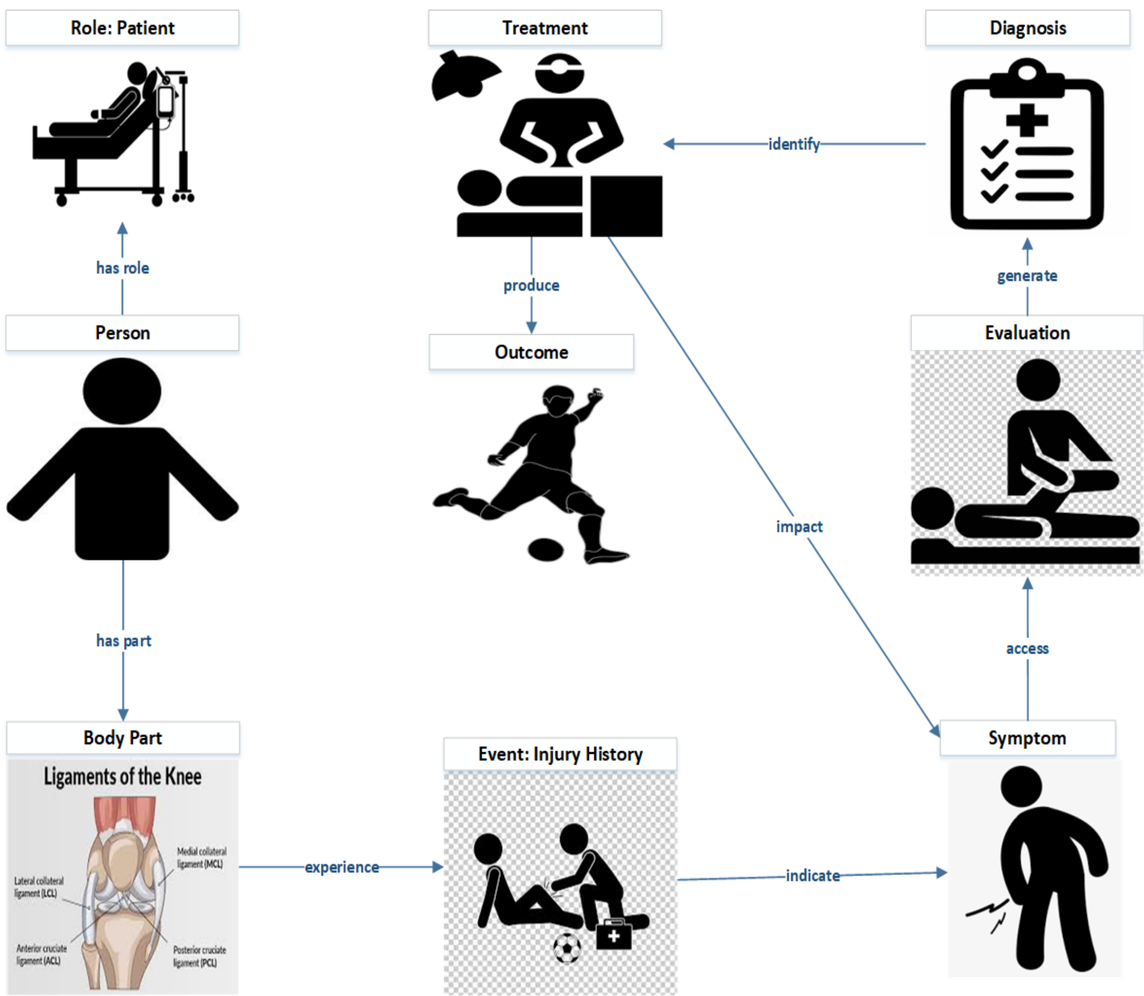


Figure 9 Visualization of ACLR health event process

4.3.5 Define the Domain Class Taxonomy

The ‘Health-Event’ concept mentioned in the domain and workflow above is transformed into the ‘HealthEvent’ class in the Ontology model. In the ACLR domain-specific model, the study categorized the ‘HealthEvent’ class as one main superclass having individual steps as its subclasses. Moreover, other main classes represent concepts related to this domain, such as body anatomy, standard terminology, and time as shown in Figure 10.

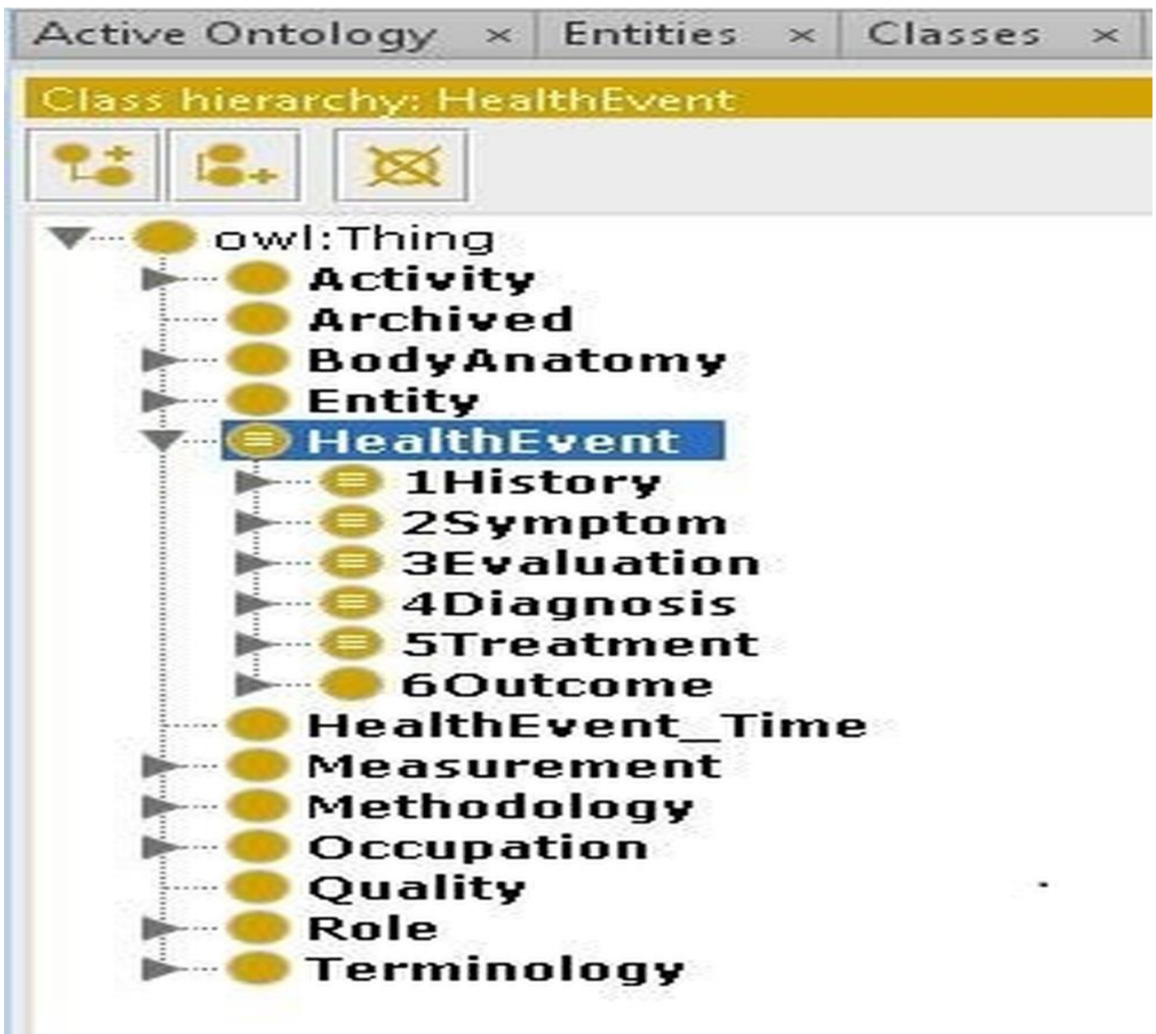


Figure 10 Class hierarchy of the health-event concept in the taxonomy structure via Protégé

In this domain-specific structure, each root class represents an explicit concept. The parent class and its sub-classes have an "is-a" relationship. The vital key here is to define these definite concepts in a way that they could be clustered that is not too broad nor too narrow. Here, the explanation of some crucial concepts that require their classes, is represented in this study.

- Class: Role

  This class represents the healthcare role of individuals for a specific healthcare process. An individual can have multiple healthcare roles, such as a female nurse getting sick and having to visit her family doctor. In general, this nurse has a role as a care provider, but when she visits a family doctor for her health, her role changes to a patient. As a result, an individual's characteristics are not enough for the knowledge model since the role can change depending on the events.

- Class: HeatlhEvent_Time

  Most of the healthcare treatments are a multiple-step process. It starts from the first visit that allows the care provider to observe the patient's sign-and-symptoms, determine diagnosis codes, and plan the treatment plan. The treatment plan can involve multiple visits, especially when it is a part of drug treatment and rehabilitation. In the study, ACLR rehabilitation is patient-centered, allowing individual patients to control their recovering process. The 'HealthEvent_Time' class represents an appointment for treatments and evaluations categorized into two sub-classes: Time_Preop and Time_Postop. Both classes capture the appointment times of evaluating knee conditions in the rehabilitation department, which permits ACLRO to structure its model using the same ‘Evaluation’ sub-class under the ‘HeathEvent’ class for different visit periods. Otherwise, the model would have to contain ‘Evaluation_Preop', ‘Evaluation_Postop_1week', ‘Evaluation_Postop_1month', and so on.

- Class: Terminology

  This class constructs various types of standard terminologies in healthcare in the U.S. For billing, the International Classifications of Diseases, i.e., ICD-9 and ICD-10, are used as the standard diagnosis code sets. In contrast, Current Procedural Terminology (CPT) is a standard procedure code set. Notably, the most benefit of the ‘Terminology’ class is to reserve the domain’s internal knowledge structure from the external change/use of terminologies. For instance, the ICD9 was changed to ICD10 officially in 2015. In ACLRO, both the ICD version are used. With the ‘Terminology’ class, the model can map the list of the domain’s internal diagnosis codes to both ICD 9 and ICD 10, without any works on individual patient records. For instance, patient records in the domain were documented as ACL – Acute tear and Knee pain, which was previously mapped to ICD-9 codes 844.2 and 719.46, respectively.

- Class: Activity

  Since the rehabilitation is patient-centered, individual patients can define recovering satisfaction differently depending on their lifestyles. A young football player would want to gain his knee strength above an average non-injured person, while a 65-year-old person would be happy and satisfied with regular daily activities. Moreover, the level and type of sports and activities can be used as one of the predicting factors of an ACL tear. The sports that involve in pivot knees such as basketball and football increase the risk of ACL tear than running or swimming. For the predictive analysis, both activity types and levels will be evaluated to determine the activity level, which will be used as one score to rate the ACL tear risk.

4.3.6 Define the Properties of the Classes

In Ontology, there are two types of property: object property and data property.

- Object Properties

  Object properties represent the relationships between two classes or objects. The most common relationship between the object is the 'IS-A' relationship between parent and child class. Protégé allows us to assign multiple characteristics to the object properties: functional, inverse functional, transition, symmetric, asymmetric, reflexive, and irreflexive. The object properties are guided by the predicate sentence mentioned in the step (4) Define the ontological definitions and constraints of the terminology. Some of the top-level object properties here are shown in Table 2 and Figure 11. With the object properties, the relationships among multiple classes, which allow us to query the knowledge model, as shown later in the last step of this domain-specific ontology section.

Object Property	**Domain**	**Range**
hasRole	Person	Role
hasBodyPart	Person	Body Part
HealthEventBelongTo	Health Event	Person
historyToSymptom	History	Symptom
symptomToEval	Symptom	Evaluation
evalToFinding	Evaluation	Finding
findingToDiag	Finding	Diagnosis
diagToTreatment	Diagnosis	Treatment
treatmentToOutcome	Treatment	Outcome
measurementToEvaluation	Measurement	Evaluation
healthEventTimePeriod	Health Event	Health-Event Period

Table 2 Object properties representing the health-event process in the ACLRO

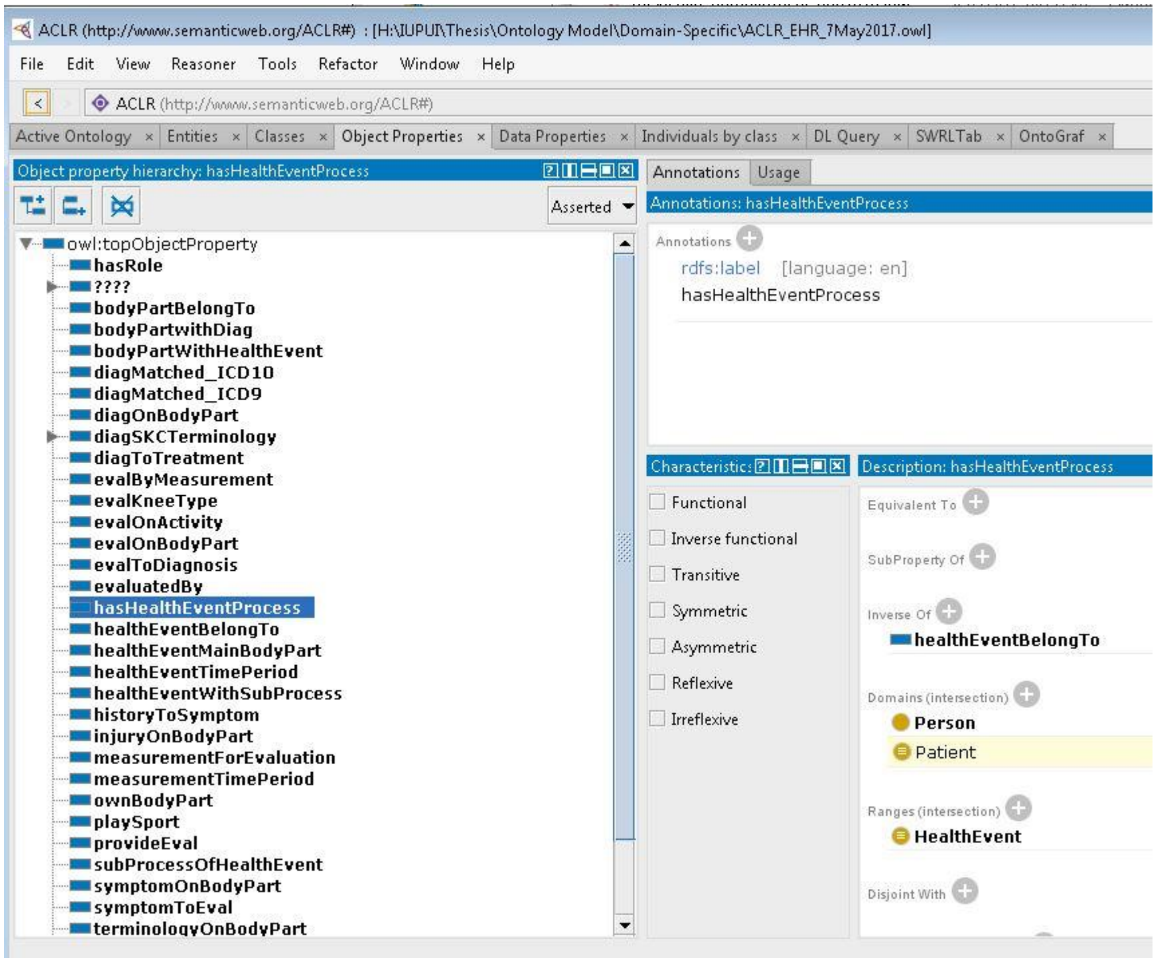


Figure 11 Object property defining the relationship between the patient and health-event concepts via Protégé

- Data Properties

  Data properties are used to describe the characteristics of concepts at the individual level. For instance, every person has the age factor as one of the characteristics defining who we are. Therefore, “hasAge” is an object property for the “Person” class. In Protégé, object properties are required to have ‘domain’ and ‘range’ information. The ‘domain’ determines which concept is the object property belongs to, while the ‘range’ explains its data type, such as numeric, string, or date-time. At this step, the core of the ontology model is capture and can be visualized, as shown in Figure 12. The next step shows the implementation of instances.

4.3.7 Develop Constraints Using Axioms, Rules and Reasoning

Up to the step, the ontology model can present the domain's concepts and relations between them. Its structure demonstrates how the entities existing in the world, and it should be the same structure across practices. That is why the ontology model should be structured as things are without adding an individual's belief. Nevertheless, the ontology model can represent more than a standard structure. With rules and reasoning, the model can add and customize knowledge. Rules and reasoning will allow the model to add belief and create customizations in concepts' relations.

For instance, this ontology can automatically map the practice diagnoses terms to multiple standard terminologies such as ICD-9 and ICD10 even both stand terminologies are different in coding components, as shown in Figure 13. The model can also use rules to add the knowledge of which combination of sport level and sport types can lead to high risk and high level of activity ratings.

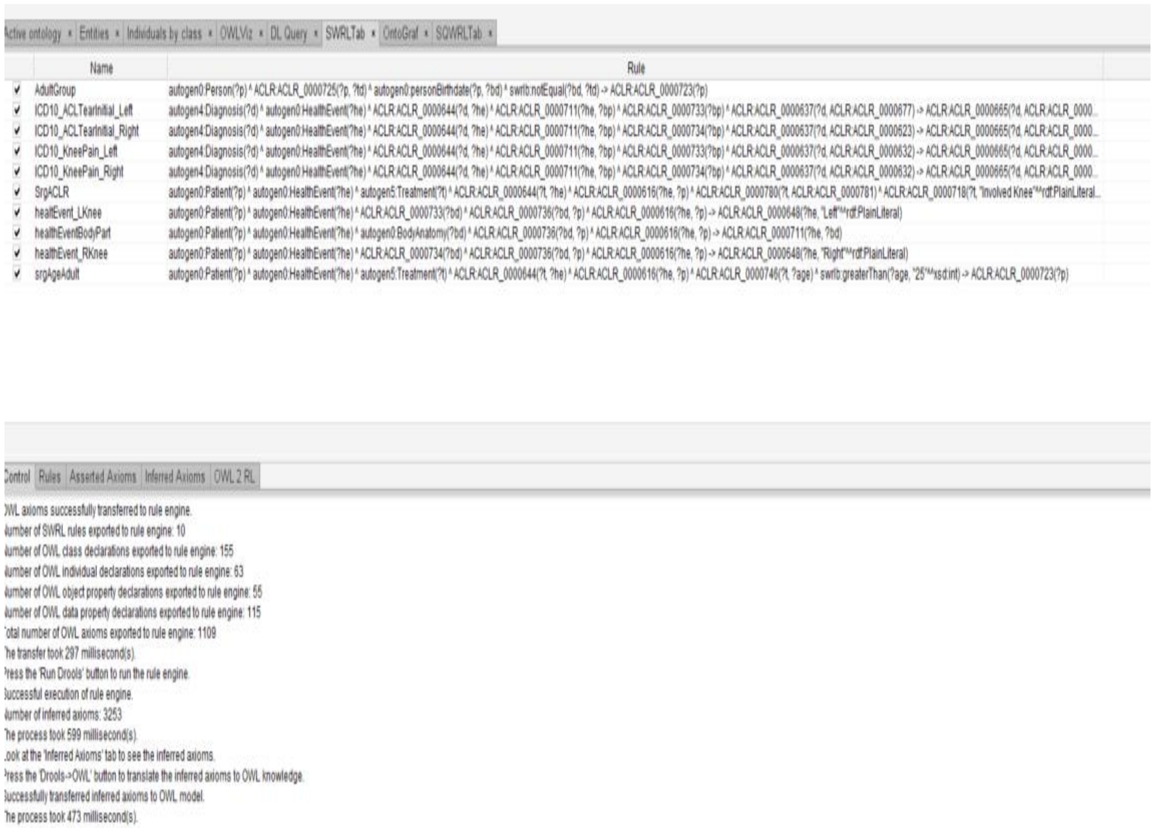


Figure 12 The implementation of SWRL rule for model constraints, axioms, and relation criteria

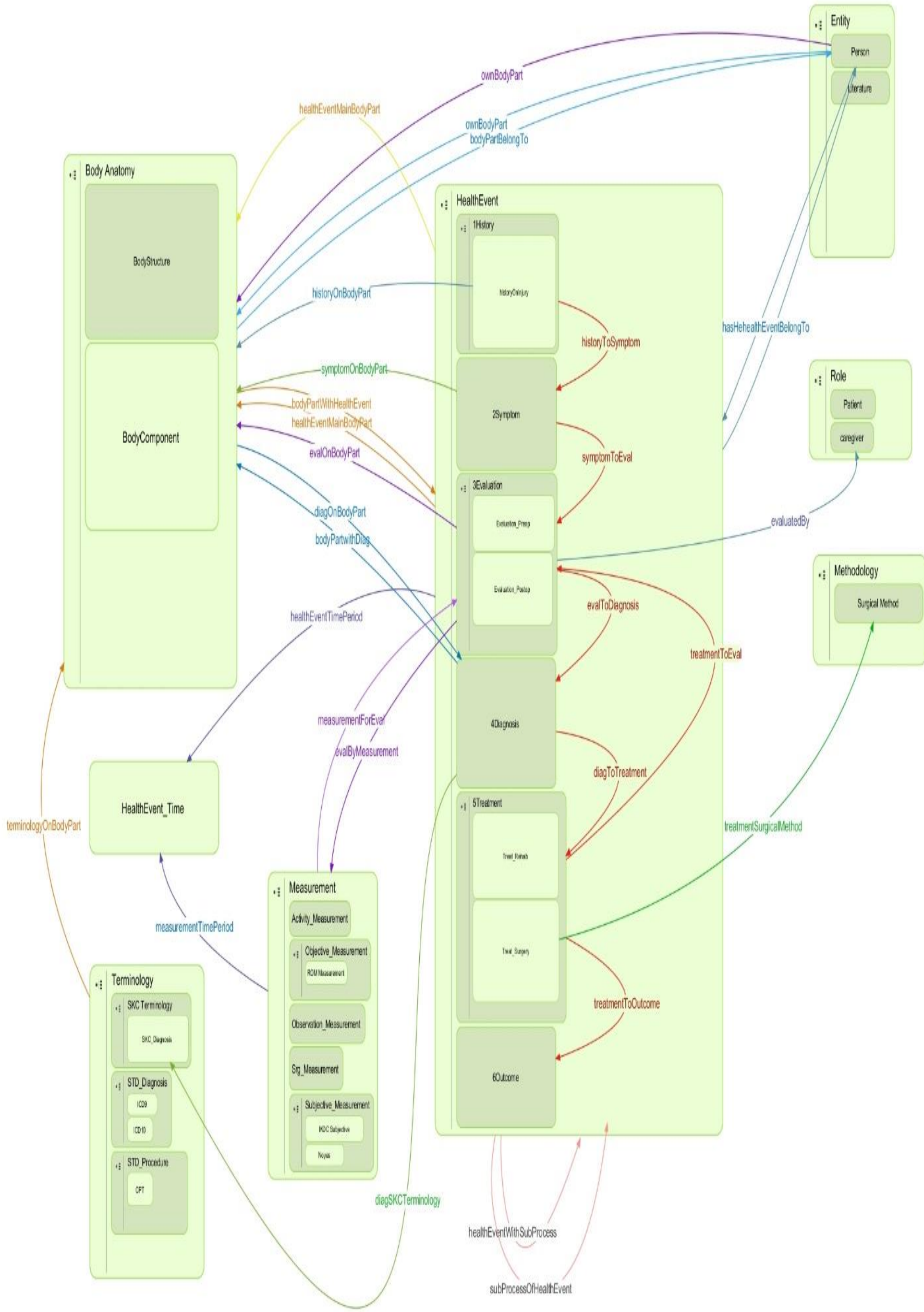


Figure 13 Concept-mapping diagram of ACLRO under the bottom-up approach

4.4 Top-Down Approach

The top-down approach aims to focus on semantic interoperability rather than specific domain-knowledge representation. Its design focuses on technical points of view over domain experts (Klischewski, 2003). The initial step of the top-down approach is to select a suitable foundation ontology in an interesting domain. The next step is to search for existing ontologies under the same foundation framework. These ontologies are called middle-level ontologies. Last, the study performs a gap analysis of overlapped concepts before integrating into the study's model. The details of individual steps are mentioned in section 4.4.1, 4.4.2, and 4.4.3.

4.4.1 Select the Foundation Framework

A foundation ontology, also known as an upper ontology or top-level ontology, is a domain-neutral ontology that comprises of general terms across all domains (Hoehndorf, 2010). There are several top-level ontologies such as Basic Formal Ontology (BFO), Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE), General Formal Ontology (GFO), Suggested Upper Merged Ontology (SUMO), and Cyc Upper Ontology. Their primary purposes are to ensure semantic integration and reusability of specific ontologies (Gibson, 2010). Therefore, the critical factor of selecting one top-level ontology over the others is on the level of interoperability that a specific domain ontology receives. For instance, an ontology in medicine should select a top-level ontology commonly used in the biomedical field.

After thoroughly reviewing the top-level ontologies in biomedicine, this study selects the Basic Formal Ontology (BFO) as the foundation ontology. Barry Smith and his team at the Institute developed BFO for Formal Ontology and Medical Information

Science (IFOMIS) at the University of Leipzig (Grenon et al., 2004). The design of BFO is to support information retrieval, analysis, and integration in scientific and biomedical research (Arp & Smith, 2008). BFO's framework represents entities and their relationships under a single framework of time and space consolidation, i.e., continuant and concurrent (Galton, 2018). Continuant entities carry on their existence through time, while concurrent entities require temporal parts for their existence, such as events and processes. The hierarchy of BFO 2 is shown in Figure 1 (Arp et al., 2015).

4.4.2 Review Middle-Level Ontologies and Existing Concepts

The requirement of middle-level ontologies for this study is to be under the BFO framework. Opportunely, there is a group dedicated to ontologies related to the life sciences and biomedical domain, named the Open Biological and Biomedical Ontologies (OBO) Foundry. The OBO Foundry principles are for a collection of interoperable references ontologies (Smith et al., 2007). The OBO's members have to adhere to its principles and be reviewed before the official acceptance. The summary of OBO Foundry principles composes of openness, standard format, orthogonality, versioning, scope, textual definitions, standardized relations, documentation, a plurality of users, exposure to collaborations, local of authority, naming conventions, and maintenance. The collection of the OBO foundry can be accessed through its site at obofoundry.org and ontogee.org. Ontobee is a linked data server for referencing ontology terms and has been used as the default ontology source for most OBO Foundry library ontologies (Smith et al., 2007). The search on Ontobee sites (ontobee.org) discovered two ontologies, i.e., OBI and IAO, that serve as the study's middle-level ontologies, as mentioned in 3.1.1 and 3.1.2.

4.4.3 Integrate the Middle-Level Ontologies

After reviewing the structure of both the OBI and the IAO model, the study analyzed the concepts involving the study. In the ACLRO, the reused concepts from the OBI are the 'Investigation' and 'Data Transformation' planned processes. The "Investigation" process has two primary relations, "has specified input" and "has specified output' relation connecting to continuous entities (Kong et al., 2011). Additionally, the "Investigation" process can have a relation with another planned process, such as "Study Design Execution" and "Drawing a Conclusion From Data." Figure 14 presents the partial high-level concept map of OBI and IAO under the BFO framework: BFO class are shown in green, IAO classes are shown in blue, and OBI classes are shown in orange. The 'data transformation' process is a process with the raw-data input and the 'analyzed-data' output, which participates in the research process in this study. Besides, the information-content entity plays an essential role in data collection during the health-intervention and data-extraction processes.

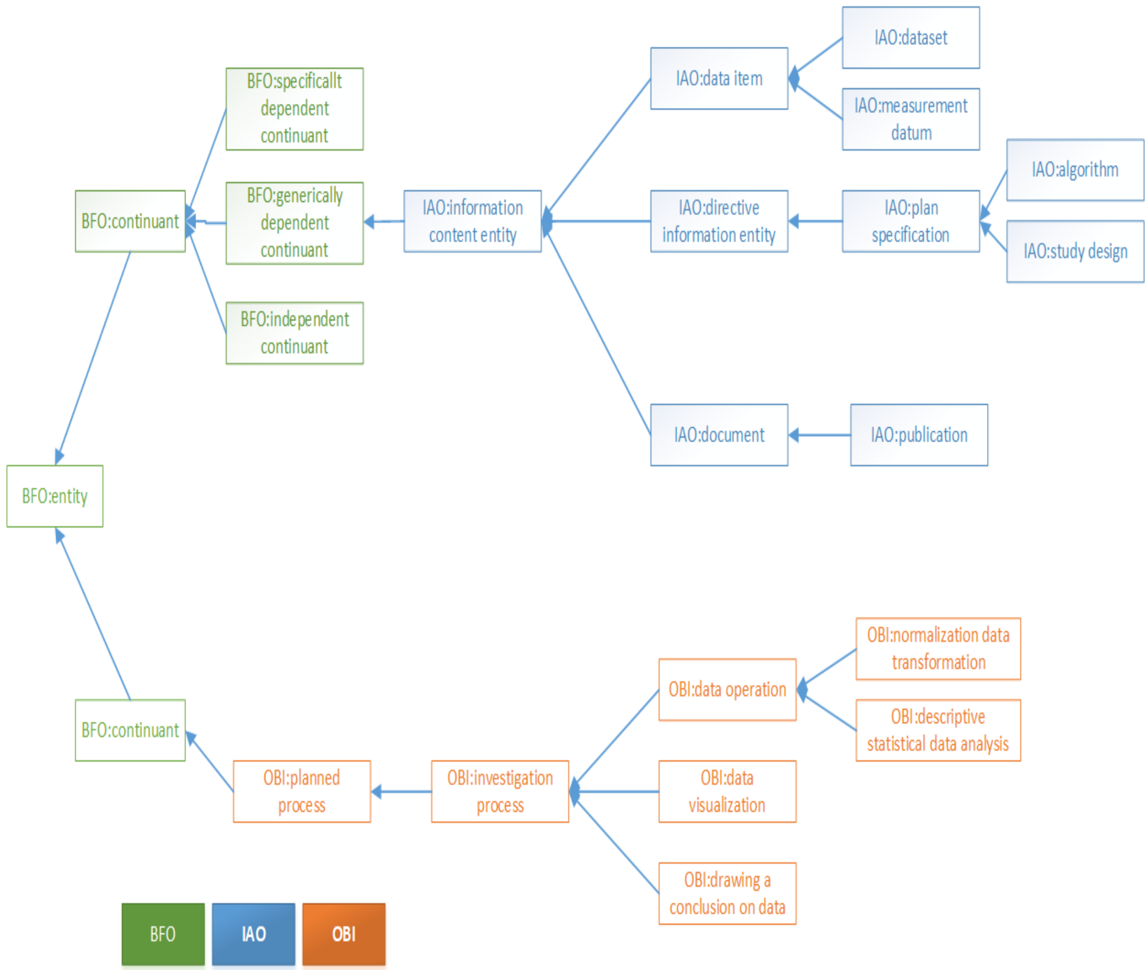


Figure 14 Partial high-level conceptual-mapping diagram of OBI and IAO under the BFO framework

4.5 Integrate the Domain Ontology to the BFO Framework

The ACLRO's concepts are primarily defined in the "Health Intervention" process, the same patient-care process as the nursing process described by the American Nurses Association (ANA), as mentioned in Step 3. The health-intervention process class has the "exists at" relationship with the temporal region, representing treatment-time periods, such as pre- and postoperative. The other key concepts in this study are documented data or information content. However, the information-content entity is not part of BFO2.0. It is an extended entity introduced by a BFO-based Information Artifact Ontology (IAO) that focuses on data collections and associates representational artifacts

(Ceusters, 2012). In ACLRO, the information-content class involved data collected during the health-intervention process and data extracted from the literature.

The structure of ACLRO with the bottom-up approach represented in Figure 13 is redesigned to fit the OBI and IAO under the BFO framework. The visualization of the new concept-mapping diagram is shown in Figure 15. With the top-down approach, the new ACLRO enhances semantic interoperability with more straightforward, organized classes. Besides, the ACLRO structure is more formal, with fewer relationships. Consequently, the new design with the BFO framework enriches the reusability and integration of the domain knowledge with another ontology due to its parent-class equivalence.

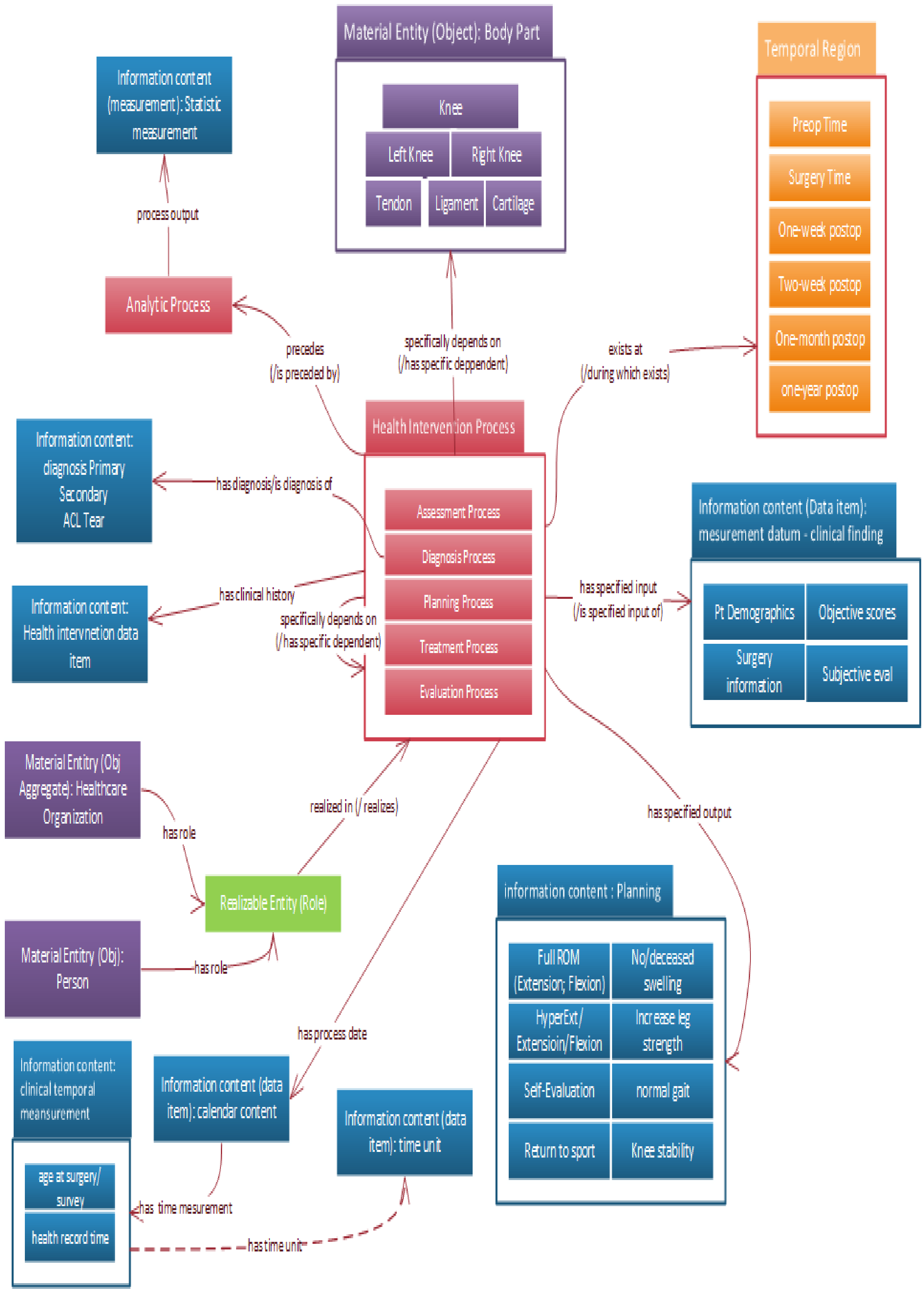


Figure 15 Conceptual-mapping diagram of the ACLRO with the hybrid approach under the BFO framework

4.6 Validation

One of the recommended methods used to capture essential terminologies is to develop multiple scenarios and capture a list of questions from these. The list-of-questions method focuses on the validation of domain conceptualization while bypassing the formal description language scenarios (Abacha, Da Silveira, & Pruski, 2013). Along with reasoning on instance entities, DL queries transformed natural language questions into ontology formal description language. The DL query's results are used to validate the model by comparing it to the expected answers of the domain experts. The following sections are organized into the creation of instances and competency testing.

4.6.1 Create Instances

In this study, the roles of instances demonstrate the design of a domain ontology in real-world settings and validate the model's competency in a scenario-test technique. The two scenarios that are implemented for the validation are below.

- Scenario A: Patient-A is a 19-year-old male. He injured his left knee while playing basketball with friends. He heard a pop at the time of injury. He had a medial and lateral ACL tear. He received a contralateral ACL reconstruction on the left knee. He came for pre-op/post-op rehabilitation at one-week, two-week, one-month, two-month, four-month, and six-month postoperative times. His knee status got evaluated on strength, range of motion, function, and self-assessments during the postoperative visit and one-year, five-year, and ten-year follow-up. He received full ROM and excellent on his knee assessment.

- Scenario B: Patient-B was a 26-year-old female. She had chronic pain on the medial side of her right knee. The pain occurred during full-extension, bending the knee, squatting, and daily activities like walking and biking. She also heard popping sound from moving her knee. She received the ACLR surgery and attended the rehabilitation program. Her right ACL got tore six years after the surgery again.

The three instances are in this paper as follows:

- Patient_1 (Figure 16 and Figure 17)
    - Has birthday on 1/1/1977
    - Is female
    - Injured on right knee

  An example of the development of an instance of the Patient concept with the information of sex and birthdate in Protégé is shown in Figure 17.

- recordDiag_pat1
    - Is a part of encounter of Patient_1
    - Is a diagnosis record of Patient_1
    - Has a primary diagnosis with a local term as (right) Acute ACL Tear, i.e., ICD9: 844.2 and ICD10: S83.511A
    - Has a secondary diagnosis with a local term as (right) Knee Pain, i.e., ICD9: 719.46 and ICD10: M25.562

  Figure 18 demonstrates an example of the development of an instance of the diagnosis-encounter concept in Protégé. Here, the instance, encounter_pat1, is created with the information of primary and secondary diagnosis that can be further mapped to standardized terminologies, i.e., ICD 9 and ICD 10.

- rescordSrg_pat1 (Figure 19)
  - Is a part of the encounter of Patient_1
  - Is a surgery record on Patient_1
  - Is ACLR surgical procedure code (CPT code) 29888
  - With a grafted side on the same knee (Ipsilateral ACLR)

Figure 19 presents an example of the development of an instance of the surgical record via Protégé. Here, the record instance, recordSrg_pat1, is created with the information of surgical details such as graft side and procedure code.

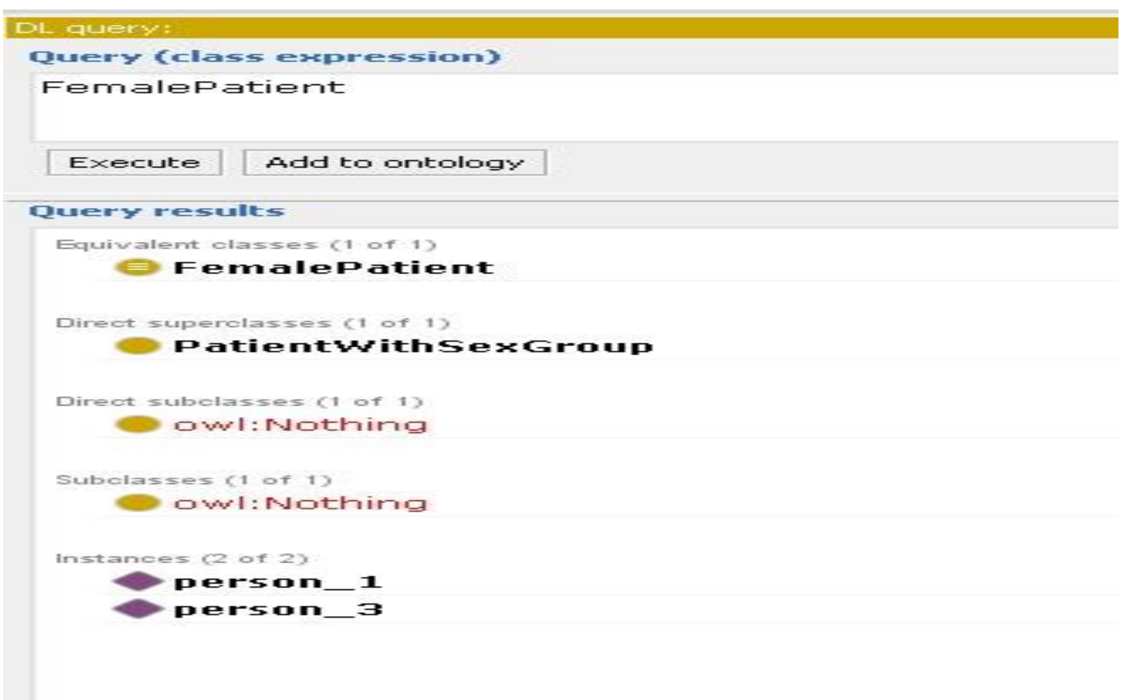


Figure 16 DL query for retrieving the female-patient (FemalePatient) concept

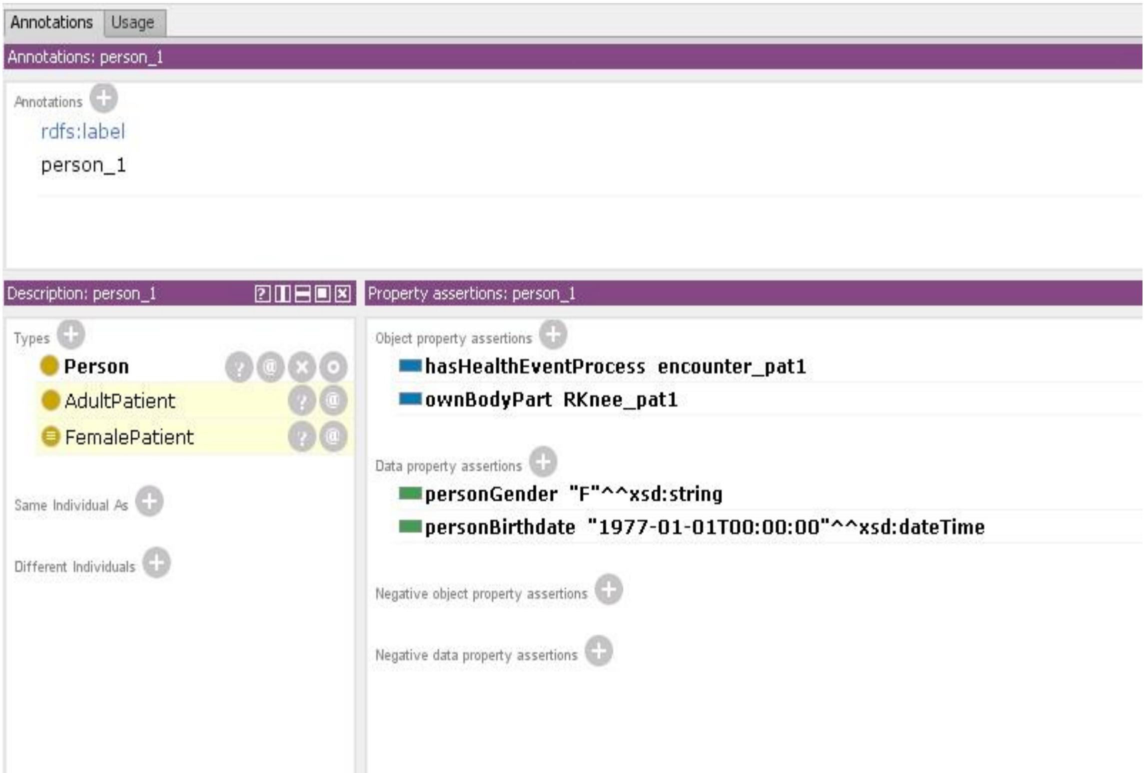


Figure 17 An instance of the Patient concept with the information of sex and birthdate via Protégé

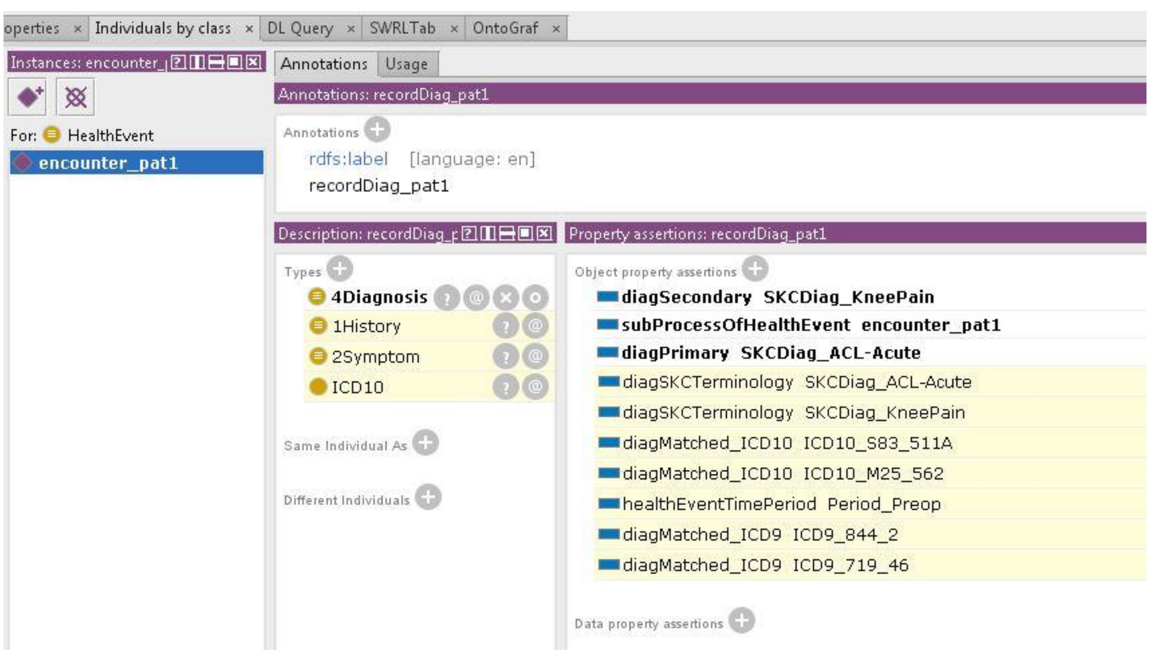


Figure 18 An instance of the diagnosis-encounter concept (encounter_pat1) with the information of primary and secondary diagnosis via Protégé

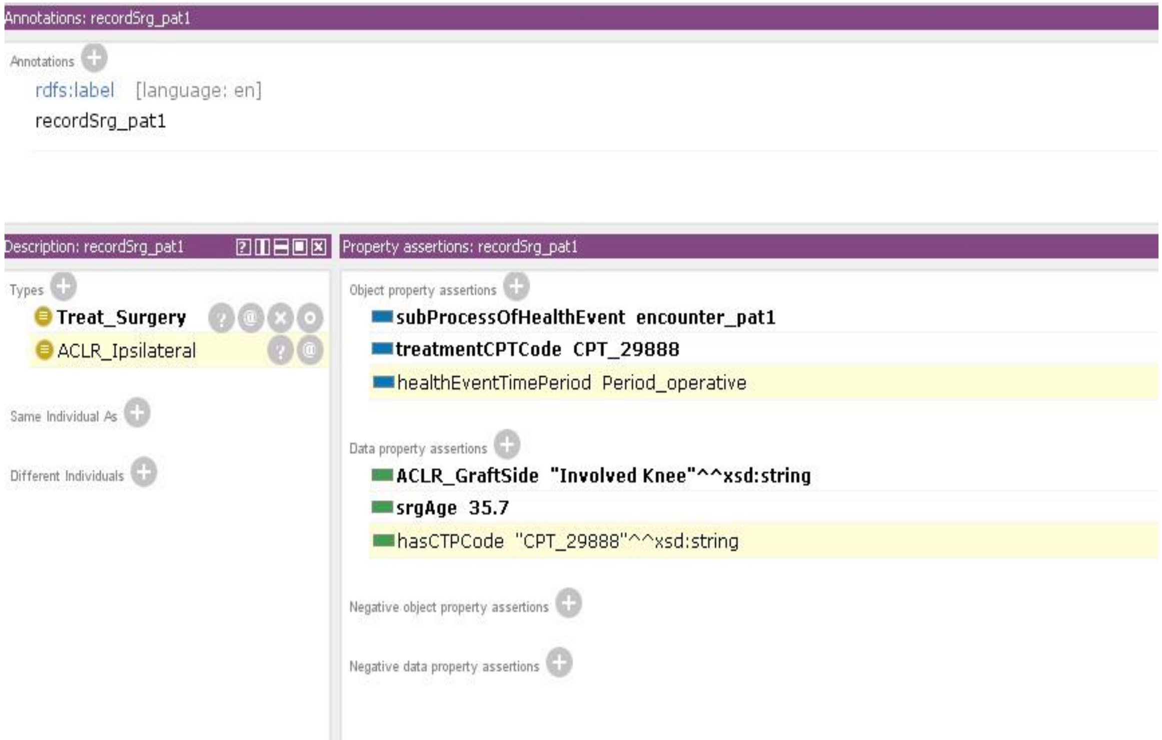


Figure 19 An instance of the surgical-record concept (recordSrg_pat1) with the information of sex and birthdate via Protégé

4.6.2 Competency Testing

- Verify the model's ability to categorize patients based on their demographic information such as age or sex.

  As shown in Figure 16 and Figure 17, the model can categorize all female patients under the 'FemalePatient' class using a piece of information in the 'personGender' property of the 'Person' class. The model can group and identify if a specific patient is grouped into a female or male patient without looking at his/her demographic info. In addition, the patient (Person_1) is also simultaneously grouped as 'AdultPatient'. This query performs the same way as a view in a relational database that selects all patients from the patient table where sex is female. For further benefit, the category method can be applied in a decision support tool to identify if the patient could be in any high-risk diseases based on a set of patient's characteristics.

- Verify the model's ability to categorize patients into a specific treatment type base on the information in their surgery records.

  In this work, the ipsilateral ACLR procedure has a formal definition in the ontology model as a surgical surgery for ACL reconstruction that uses an auto-graft from the same side of an injured knee. Besides, the surgical records do not have a direct relationship with the ipsilateral-ACLR concept. Nevertheless, the ACLR model successfully selects all ACLR surgery using the auto-graft from the same side of the injuries knee into the 'ALCR_Ipsilateral'. As an instance of the ACLRO, the treatment record for a specific patient (recordSrg_pat1) has information about the treatment CPT code (treamentCPTcpde) along with the source of graft side (ACLR_GraftSide), which led the model to identify the surgery record as Ipsilateral ACL Reconstruction (ACLR_Ipsilater), as presented in Figure 16.

- Verify the model's ability to map local terminologies to various standardized terminologies such as ICD 9 to ICD10.

  A standardized terminology offers a classification of common terms that are designed to be shared among users (Iroju, Soriyan, Gambo, & Olaleke, 2013). The primary purpose is to facilitate interoperability and information exchange. However, standardized terminology requires a learning curve and might not be initially recorded in the local documentation. Additionally, the documenting of patient records typically includes only one standardized terminology. Hence, the additional connection to different standardized terminologies is required for the information exchange. In the ACLRO, the local terms can be mapped to both ICD 9 and ICD 10 simultaneously through the concept-mapping technique. As seen in Figure 18, the diagnosis record of Patient_1 (recordDiag_pat1) shows the local primary and secondary diagnosis as Acute ACL Tear and Knee Pain, respectively. On the side note, ICD 10 requires body-side information in their code, which is not determined in either ICD 19 or the local terms. Nevertheless, the ontology model effectively connects the information recorded in the patient encounter to the ICD 10 mapping.

4.7 Conclusion

The domain ACL-Rehabilitation ontology (ACLRO) serves as a proof of concepts that demonstrates how the ontology can be applied to a real-world setup and bring benefits as claimed. The study proposes a practical technique to represent the patient-focus process that utilizes both bottom-up and top-down approaches. The purpose of the hybrid approach is to apply the standard framework of the top-down approach with the specific domain knowledge obtained in the bottom-up approach. The foundation framework helps the domain ontology to define a more explicit structure with fewer adding relationships. As a result, the maintenance of the domain ontology is more straightforward in expanding and integrating more concepts. The outcome of this work shows the success of the hybrid approach that enhances the representation of domain knowledge while improving the reusability and semantic interoperability of the domain ontology.

The ACLRO model can construct the class of data-driven, patient-focus process, which allows Ontology to share the knowledge explicitly from technology or EHR systems. Additionally, the ACLRO can reuse and remap the internal concepts to various external concepts without any edition of the existing structure and relations in the domain.

CHAPTER FIVE AIM 2 STATISTIC ONTOLOGY

5.1 Introduction

Statistics is the branch of science that uses quantitative or mathematical methods to analyze data, and play an essential role in medicine and health sciences. Medical statistics and biostatistics are the study of human health, treatment, and disease, ranging from epidemiology, public health, health promotion, and clinical research. Healthcare organizations engage statistics in decision-making, continuous quality improvement programs, business strategies, health care policy, and financial planning (Deber, Kraetschmer, & Irvine, 1996; Plichta & Garzon, 2009; Rice, 1977; Russell, Gold, Siegel, Daniels, & Weinstein, 1996). The roles of statistics involve collecting, summarizing, presenting, and drawing a conclusion. With the fast progressing in technologies and growing in Big Data, the need for statistics is rising, along with complex problems that require the multidisciplinary team's cooperation from subject-matter experts to advance analytics (Murdoch & Detsky, 2013). With the increasing demand for knowledge sharing, more statistics and machine learnings are utilized to analyze healthcare Big Data. The various statistics applications are designed for researchers who are not experts in statistics fields (Ocaña-Riola, 2016). Statistical analysis can be done in a single click without a supporting document for algorithm selection. Consequently, healthcare is facing a time of "data rich, information poor." The lack of high-quality documentation prevents researchers from sharing knowledge discovery (Sermeus, 2016). Furthermore, the insufficiency of documenting in statistical methods and processes prevents the validation and reproducibility of statistical analysis and a comparison between studies in clinical research (Strasak et al., 2007). It is a known challenge that most publications and external evidence are challenging to reapply their methods to achieve the same outcomes due to the lack of information on study design, statistical criteria, and algorithm applied (Zheng et al., 2016).

5.2 Existing Ontologies

5.2.1 Statistics Ontology (STATO)

STATO was a standalone project presenting data analysis results as part of the community-drive for the International Society of Automation (ISA) in early 2012 (González-Beltrán, Maguire, Sansone, & Rocca-Serra, 2014). STATO has BFO as its top-level model and OBI as its mid-level ontology. It is developed with the Web Ontology Language and follows OBO Foundry principles (Howland, 2007). STATO provides a general structure around common statistical concepts and properties from mathematical terms to statistical-related processes, such as statistical tests, data distribution, variables, and outcome representations in a non-specific domain. The main six STATO's objectives reported on its web page (http://stato-ontology.org) are:

- Serve as a supporting resource for statistical methods in the communications and reporting of scientific results for scientists and researchers by providing guideline compliance and standardizing annotation of result tables.
- Structure the fundamental classes for annotating statistical methods and their assessments; and connect to the associated hypothesis for better representation, interpretation, and review.
- Provide formal definitions of most common univariate statistical tests

- Deliver a formal means of directing and characterizing the criteria of standard statistical tests
- Offer a semantic framework supporting standardized analysis reports
- Recommend specialized formal terminologies for text mining of statistical analysis

STATO organizes the concepts of statistical theories and their components under the information-content class, such as conclusions based on data, data distribution, and hypothesis. The majority of information about statistical-related processes is listed under its ‘data transformation’ process. The process transformed the input data into analyzed or calculated outcomes. Additionally, in this study, ACLRO structure statistical tests as a subclass of the statistical model under “data-item” concepts. STATO also includes concepts of experimental design and graphical visualization for results, as presented in Figure 20.

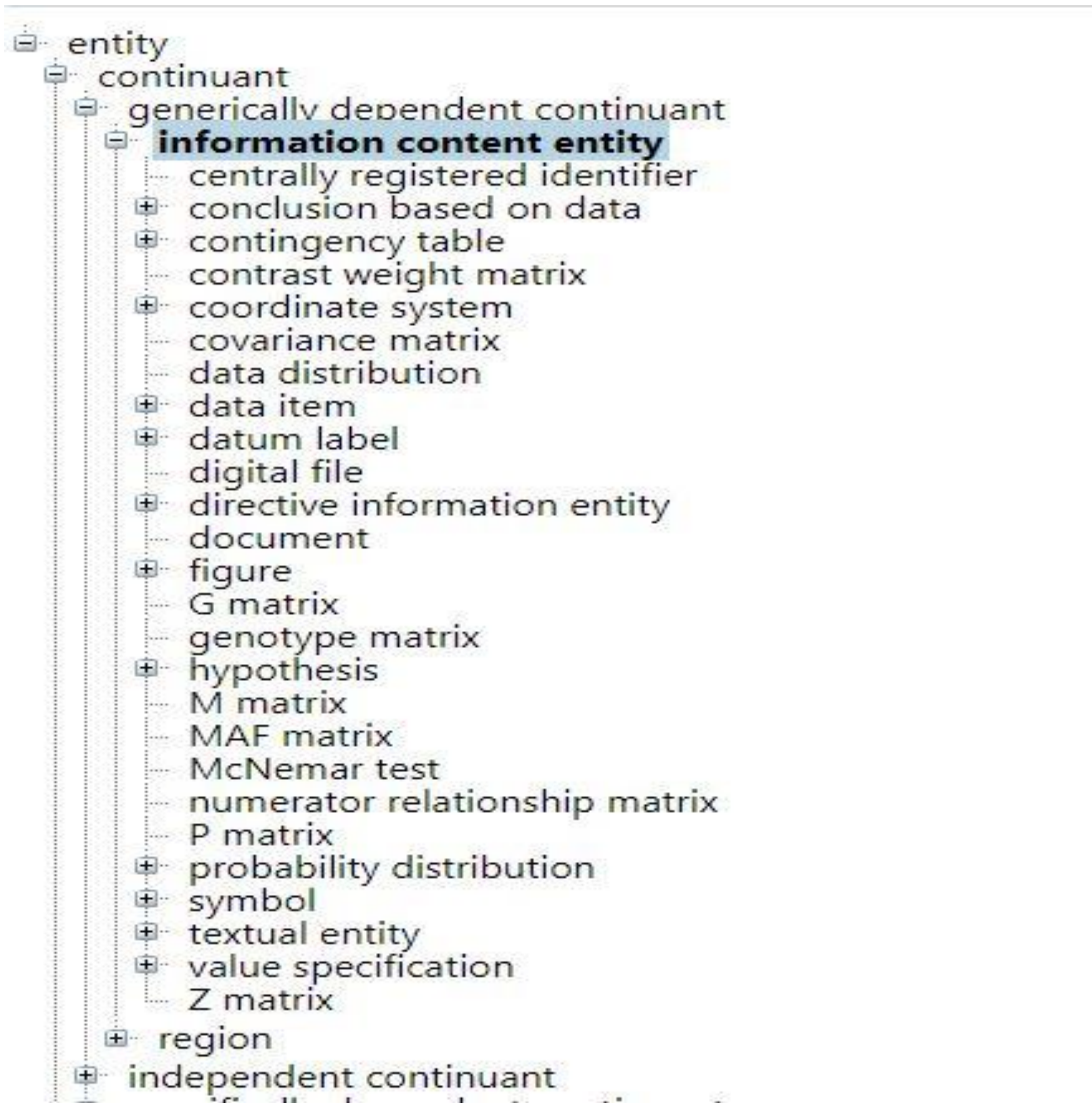


Figure 20 STATO structure of statistics concepts

Furthermore, STATO's web page presents multiple examples of different query cases to answer and demonstrate how the ontology model queries an answer through description logics expressions (http://stato-ontology.org/queryCases.jsp). The sample of questions is grouped into four categories: statistical tests, statistical measures, statistical plots, and study design.

5.2.2 Ontology of Biological and Clinical Statistics (OBCS)

OBCS is a community-based ontology that primarily focuses on statistical representation in biological, biomedical, and clinical domains. It was originated for a study of the Analysis of Variance (ANOVA) meta-analysis of vaccine protection assays in 2010 (He et al., 2010). Its areas expanded to survival rate analysis and advanced more statistical terms in Ontology for Biomedical Investigation (OBI, http://purl.obolibrary.org/obo/obi) (Zheng et al., 2016).

OBCS is written in the W3C standard web ontology Language (OWL 2, http://www.w3.org/TR/owl-guide/), and follows OBO Foundry principles. The development of OBCS is a combination of both top-down and bottom-up approaches. The upper ontology for the top-down approach was imported from the OBI class using the BFO 2.0 classes-only version (BFO, http://purl.obolibrary.org/obo/bfo). Additionally, Information Artifact Ontology (IAO, http://purl.obolibrary.org/obo/iao) was applied as its middle-tier model (Zheng et al., 2014). For that reason, OBCS's ontology structure begins with BFO's continuant and occurrence concepts. Then the classes were extended into information-content and planned-process imported from IAO and OBI, respectively. In addition, OBCS follows ontology guidance by reusing existing terminologies like STATO. As in Figure 21, IAO offers statistics-related terminologies, including probability distribution, statistical variables, and data-collection design, while OBI provides the parent class for the planned process relating to data collection, data analysis, and data transformation. The bottom-up approach was generated using cases that captured further terms related to the basis of statistics workflows through beyond

OBI. The first two use cases studied the systems biology of influenza vaccination (Lin & He, 2012) and clinical outcomes of nursing services data (Needleman et al., 2011).

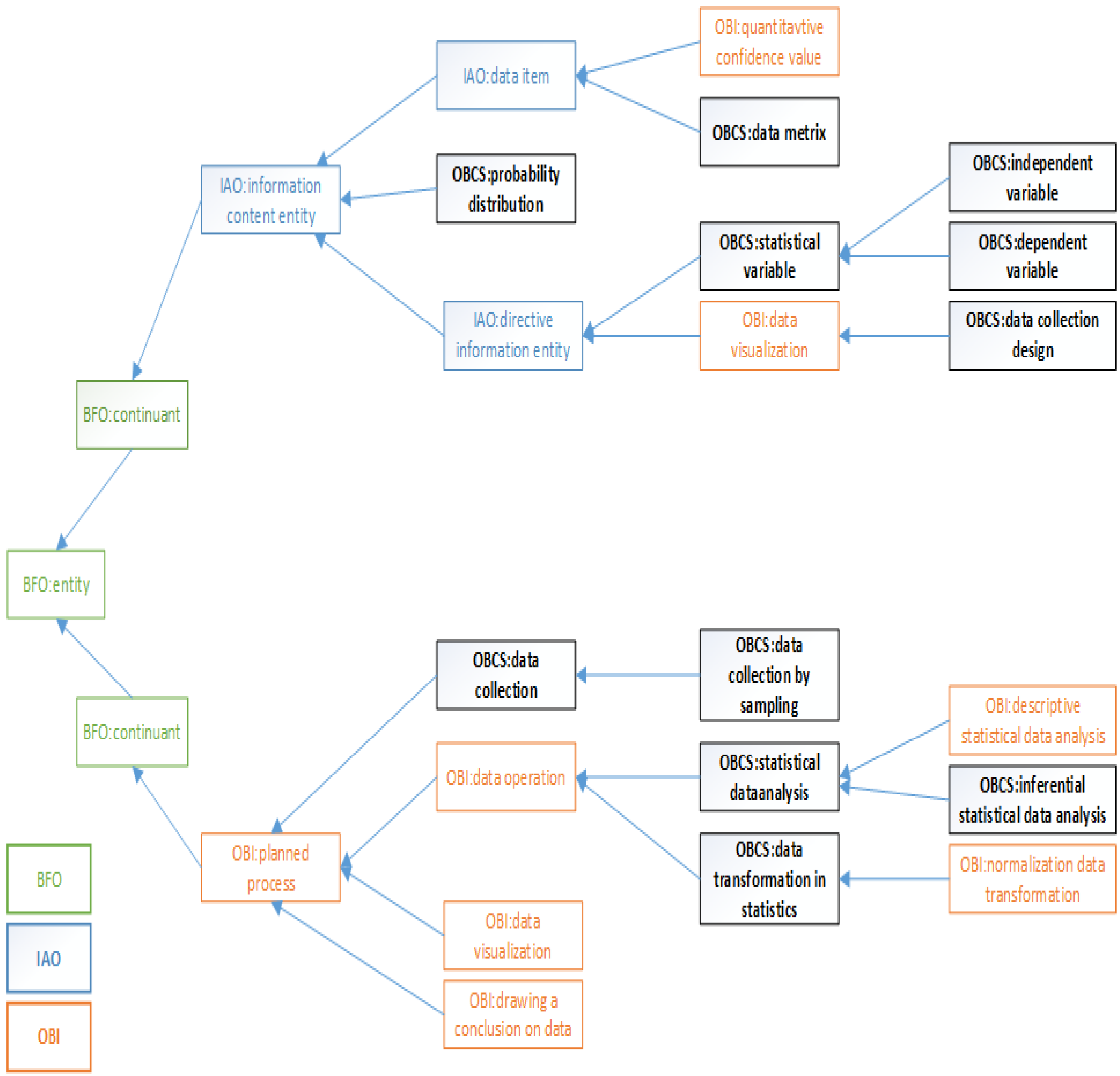


Figure 21 OBCS's hierarchical structure and key ontology terms

OBCS is designed to envision both statistics-mathematics terminologies, such as Information Content Entity, and their general workflow processes, such as Planned Process. The semantic representation of OBCS's statistical concepts is shown in Figure 22. The vital statistics concepts are under the 'information content entity', including 'probability distribution', 'testable hypothesis', 'value specification', 'statistic model', ‘figure’, and ‘conclusion based on data’. The five core processes are ‘data collection’, ‘data visualization’, ‘data transformation in statistics’, ‘(descriptive/inferential) statistical data analysis’, and ‘drawing a conclusion based on data’.

As of 2016, OBCS denoted over 800 terms, 20 BFO’s classes, 403 OBI classes, 229 its classes, and over 100 classes imported from other OBO Ontologies (Zheng, Harris, Masci, Lin, et al., 2016). Bioportal also summarized the ontology metrics of OBCS, as in Table 3 (http://bioportal.bioontology.org/ontologies/OBCS).

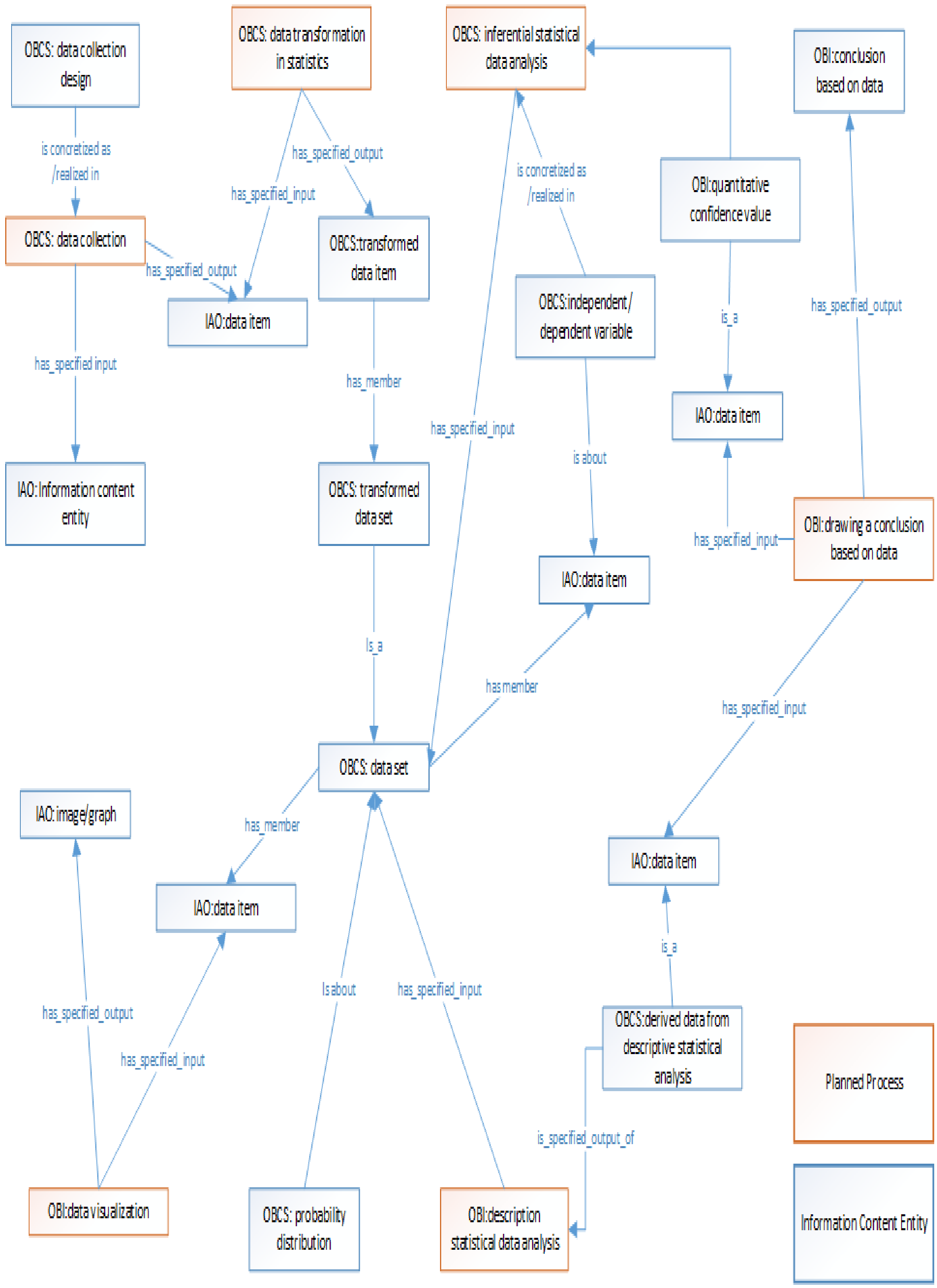


Figure 22 OBCS's semantic representation of statistics-related concepts

5.2.3 APA Statistical Cluster (APASTATISTICAL)

APASTATISTICAL represents three main classes: ‘DesignAnalysisInterpretation’, ‘StatisticalReliabilityValidity’ and ‘VatisticalTheoryExperimentation’. The model does not have a foundation model nor follow a top-down approach. There is no documentation or user group support. As of December 2019, the project status is in alpha state.

5.2.4 Ontology of Clinical Research (OCRE)

OCRE is designed to support a systematic description of, and interoperable queries on, human studies and their elements (http://boiportal.bioontology.org/ontologies/OCRE). There is no top-level ontology used in its framework. As of Dec 2019, the project status is in alpha state.

5.2.5 Mathematical Modelling Ontology (MAMO)

MAMO is designed to classify the most common mathematical models used in the life sciences and other related items, such as variables and relationships (http://boiportal.bioontology.org/ontologies/MAMO). The four main classes are model, modeling entity feature, readout, and variable. There is no top-level ontology used in its framework. As of Dec 2019, the project status is in alpha state.

5.2.6 Semanticscience Integrated Ontology (SIO)

SIO is designed to act as a top-level model for knowledge representation in physical, processual, and information entities (http://boiportal.bioontology.org/ontologies/OCRE). It aims to offer a vocabulary for Bio2RDF (http://bio2rdf.org) and Semantic Automated Discovery and Integration (SADI) (http://sadiframework.org) projects. SADI is a framework for interoperability between distribute data and analytical resources. The three core parent classes are ‘attribute’, ‘object’, and ‘process’.

5.3 Model Comparison

From six statistic-related ontologies, only three are in production and used by other projects, i.e., OBCS, STATO, and SIO. Two out of these three, OBCS and STATO, were implemented with the top-level ontology using BFO and follow the OBO Foundry principles. Both OBCS and STATO emphasize on standardizations and reusability in science domains with the use of a foundation ontology. Nevertheless, there are some dissimilarities between these two. First, their scope and focus domain is different. While STATO has a broader scope in science domains, OBCS’s main scope is in biological and clinical statistics. Second, OBCS and STATO are not alike in their statistics-related characteristics. For instance, under the ‘investigation’ process, STATO extends the class to more biology-related aspects, i.e., ‘acute toxicity study’, ‘genetic association study’, and ‘high throughput screening’ sub-classes). On the other hand, OBCS’s ‘investigation’ process focuses more on research hypothesis parts, as presented in Figure 23 and 24.

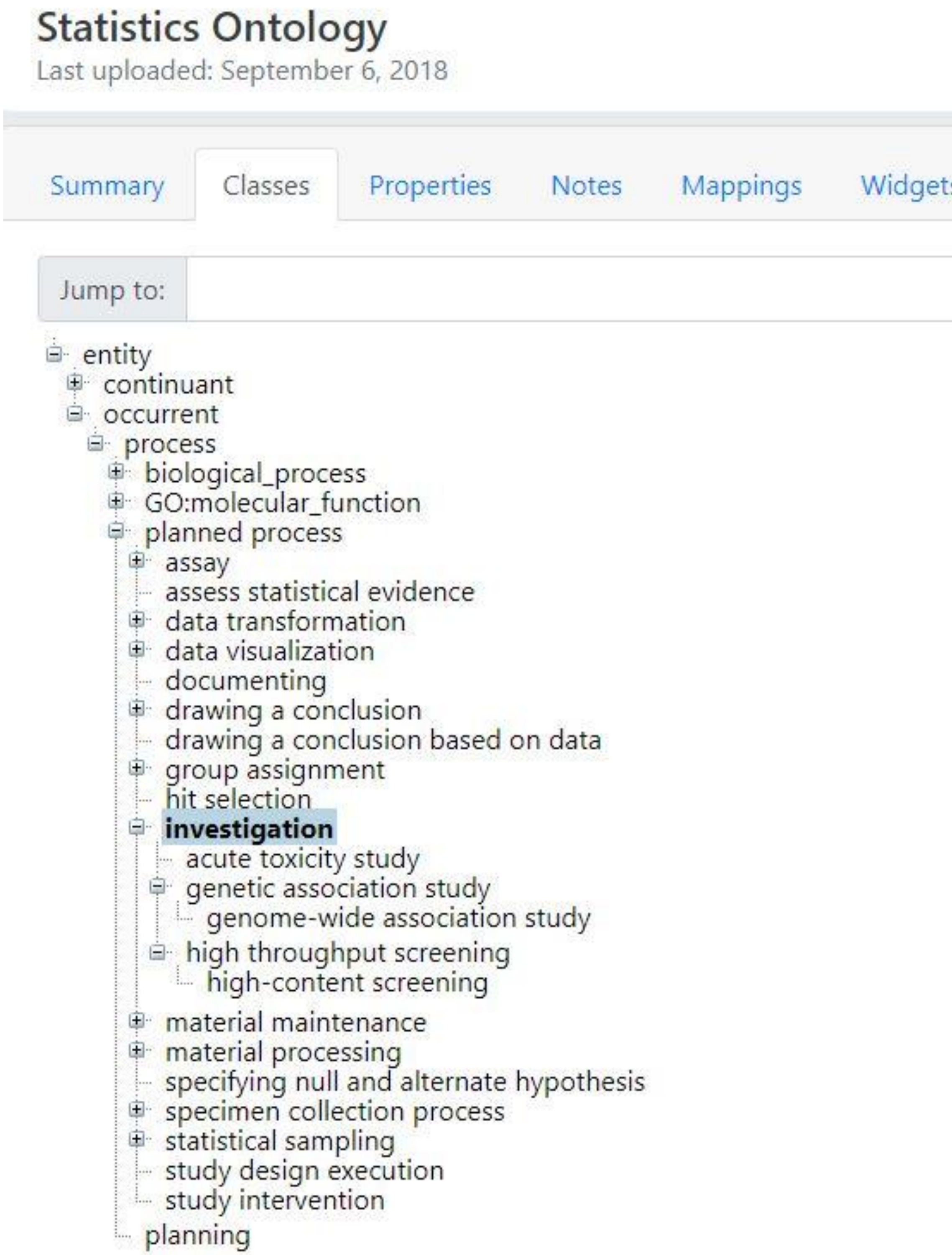


Figure 23 Sub-classes of ‘investigation’ process in STATO

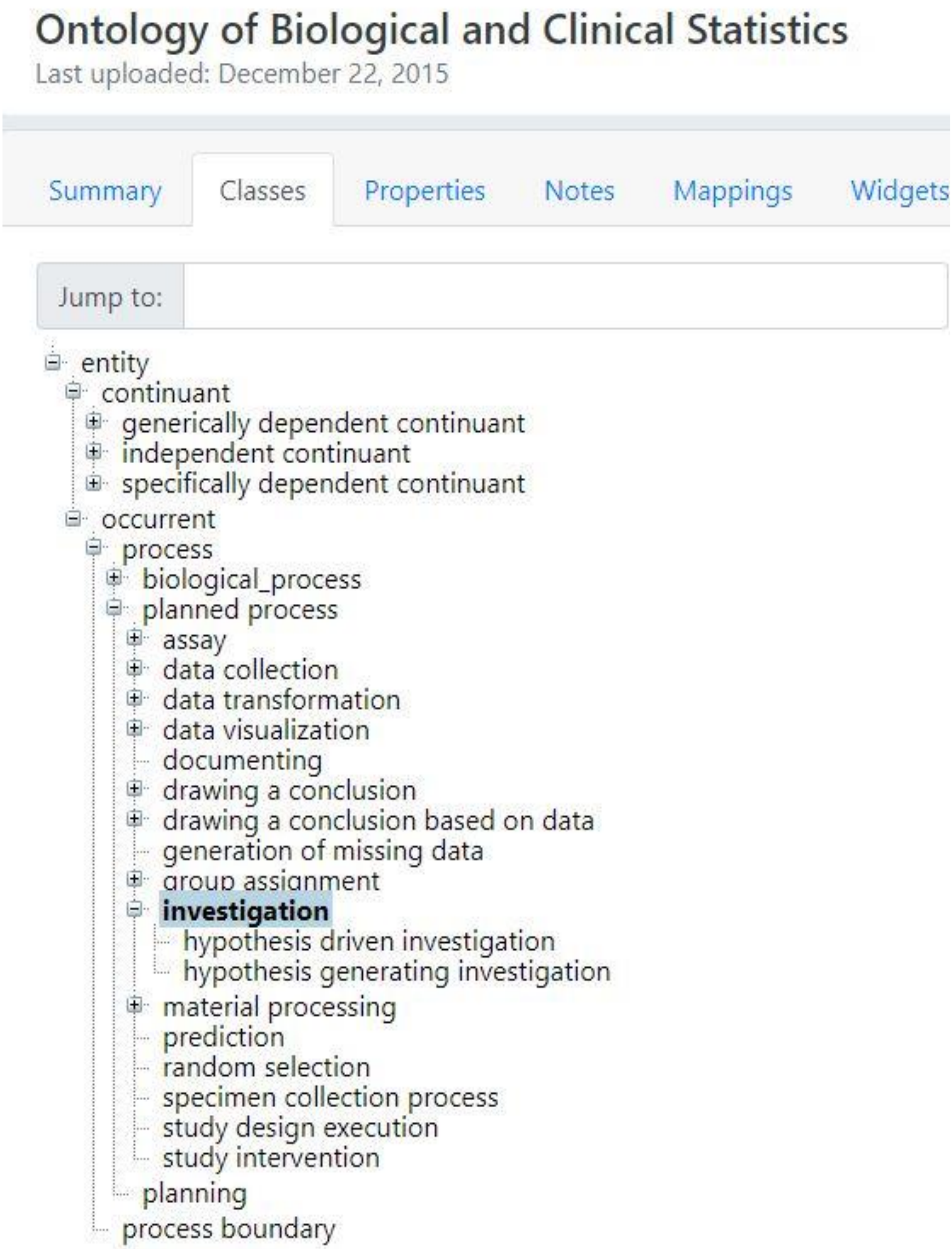


Figure 24 Sub-classes of 'investigation' process in OBCS

As of December 2019, BioPortal reported the summary metrics of the statistics-related ontology, indicating the models' quality and maturity. The "classes with only one subclass" metric reports the number of classes with only one subclass in the is-a relation. The high number of this metric implies that the structure is under-specified, or the classification is not appropriate. In contrast, when the classes have many subclasses as measured in the "classes with more than 25 subclasses" metric, the measurement indicates that the additional distinction might be needed.

The comparison of the statistics-related ontologies is summarized in Table 5.1.

Metrics	OBCS	STATO	SIO	APA-STAT	OCRE	MAMO
Last uploaded date	22/12/15	17/10/15	4/12/18	1/11/15	6/21/13	10/17/15
Status	Prod	Prod	Prod	Alpha	Alpha	Alpha
Foundation Model	BFO	BFO	N/A	N/A	N/A	N/A
Number of Classes	779	100	1,544	140	389	100
Number of Individuals	25	0	0	0	39	0
Number of Properties	48	3	212	0	220	3
Maximum depth	10	5	10	1	6	5

Maximum number of children	27	12	118	87	19	12
Average number of children	3	2	3	35	3	2
Classes with a single child	71	13	174	0	19	13
Class with more than 25 children	3	0	1	2	0	0
Classes with number definition	43	1	165	140	68	1

Table 3 Ontology metrics reported by BioPortal on Dec 2019 (http://bioportal.bioontology.org)

5.4 ACLRO's Statistics-Concept Modeling

The statistical vocabularies and concepts in the thesis are imported from four statistic-related ontologies, i.e., IAO, OBI, OBCS, and STATO. Each ontology suits the study needs in different aspects. For instance, STATO offers the 'hypothesis' class, which is essential to ACLRO. While other ontologies only have the 'null hypothesis' subclass, STATO includes more based on research aims and statistics models, such as 'goodness of fit hypothesis', 'presence of association hypothesis', and 'absence of difference hypothesis'. The study borrows multiple terms from OBCS that are not available in the others. Some examples are terms under ‘statistical model’ and ‘statistical variable’ classes.

In addition to the above existing ontologies, ACLRO also developed more specific terms to meet the study’s needs. For instance, the ‘healthcare variable’ subclass is added to the ‘variable’ class. The high-level structure is shown in Figure 25 and Figure 26.

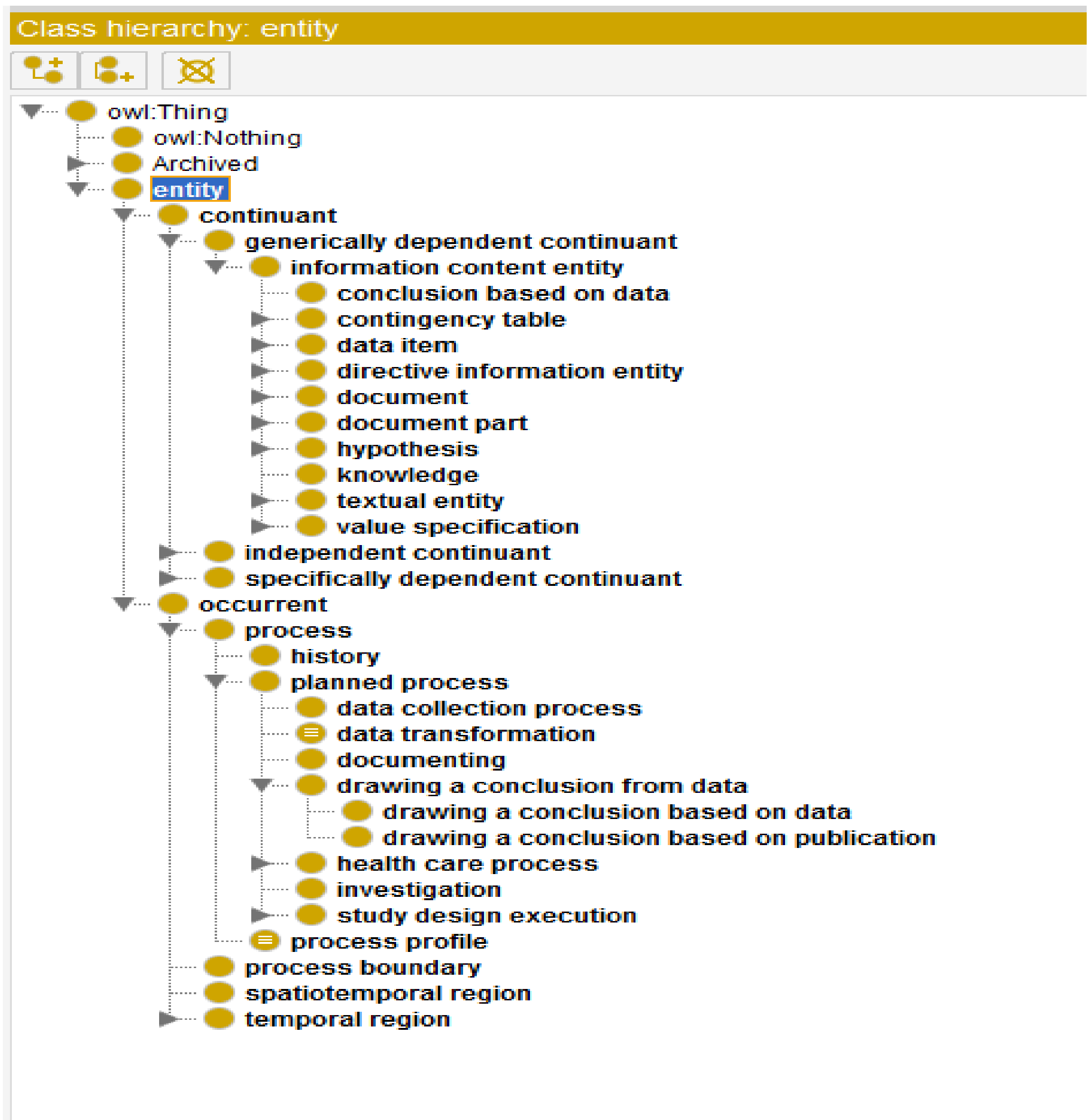


Figure 25 High level of statistics-related hierarchical structure

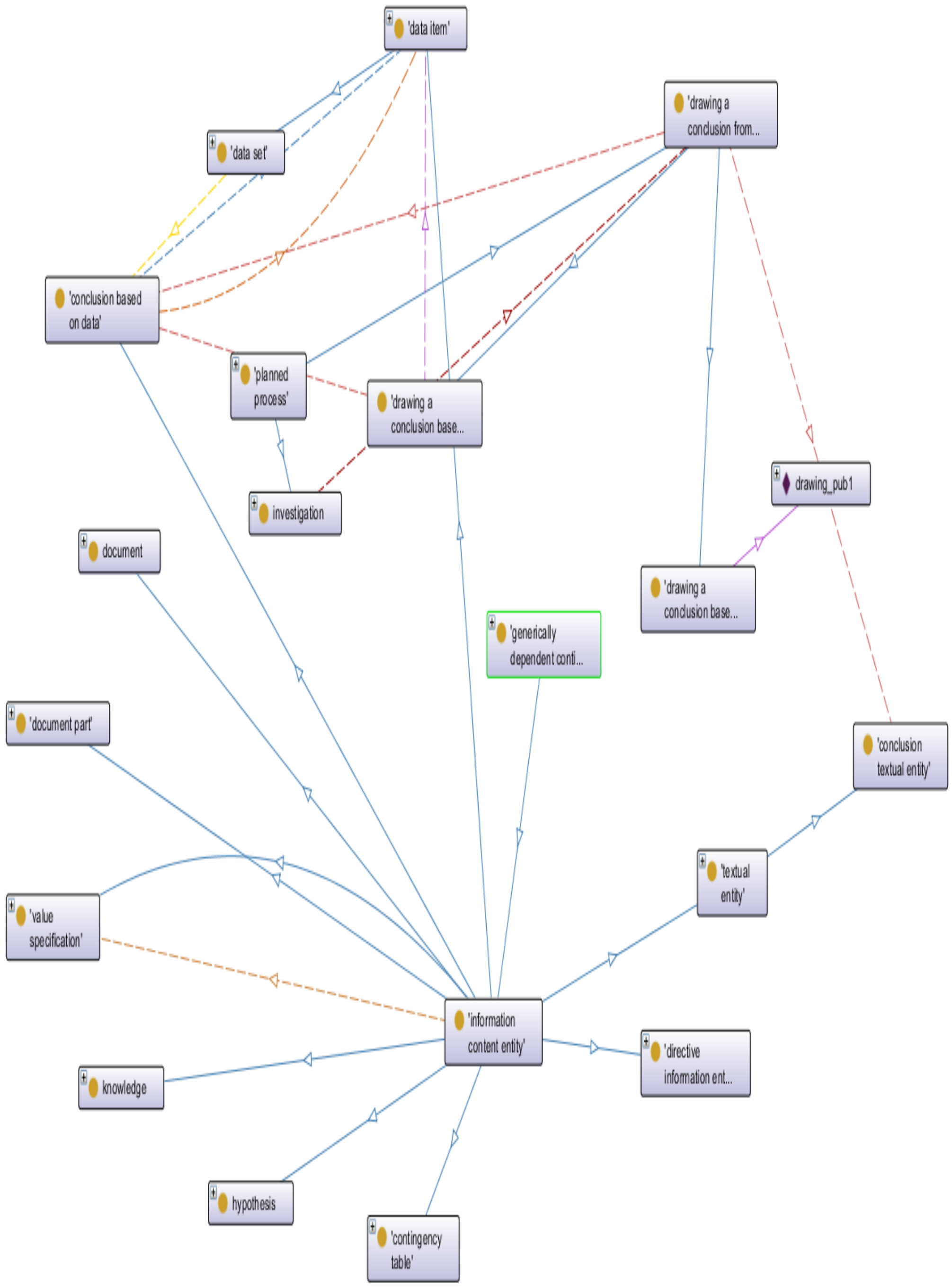
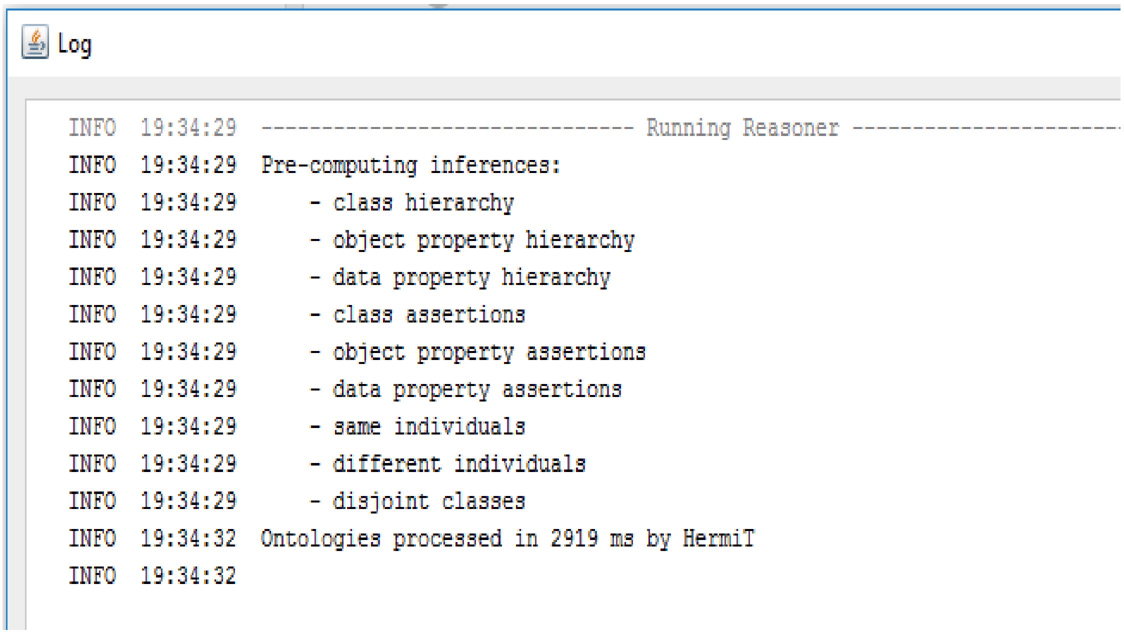


Figure 26 OntoGraph: High level of statistics-related classes

5.5 Consistency Testing

For consistency testing of the model, ACLRO used the HermiT reasoner. HemitT ran successfully without any error message in 2,919 milliseconds, as shown in Figure 27.



```
Log
INFO  19:34:29  ------------------------------ Running Reasoner ----------------------
INFO  19:34:29  Pre-computing inferences:
INFO  19:34:29      - class hierarchy
INFO  19:34:29      - object property hierarchy
INFO  19:34:29      - data property hierarchy
INFO  19:34:29      - class assertions
INFO  19:34:29      - object property assertions
INFO  19:34:29      - data property assertions
INFO  19:34:29      - same individuals
INFO  19:34:29      - different individuals
INFO  19:34:29      - disjoint classes
INFO  19:34:32  Ontologies processed in 2919 ms by HermiT
INFO  19:34:32
```

Figure 27 Reasoner log of ontology processing by HemiT

5.6 Validation

The study presents various ontology DL queries, aiming to answer questions with multi-aspects to demonstrate how the ontology model implements formal representation and semantic relationships between statistics and research design. First, the study organizes the questions into three categories: statistics-related features, research design, and knowledge discovery. Then, the study creates a set of questions that captures aspect(s) for each category as below. Note that the answers here are not complete since ACLRO is built as a proof of concept. As the model expands, more information will be added.

5.6.1 Statistic Parameters

- About statistic parameters, which probability distribution that requires both shape and location parameters?

  DL query: 'probability distribution' and 'is denoted by' some ('probability distribution shape parameter' and 'probability distribution location parameter')

  Result: Weibull probability distribution (Figure 28)

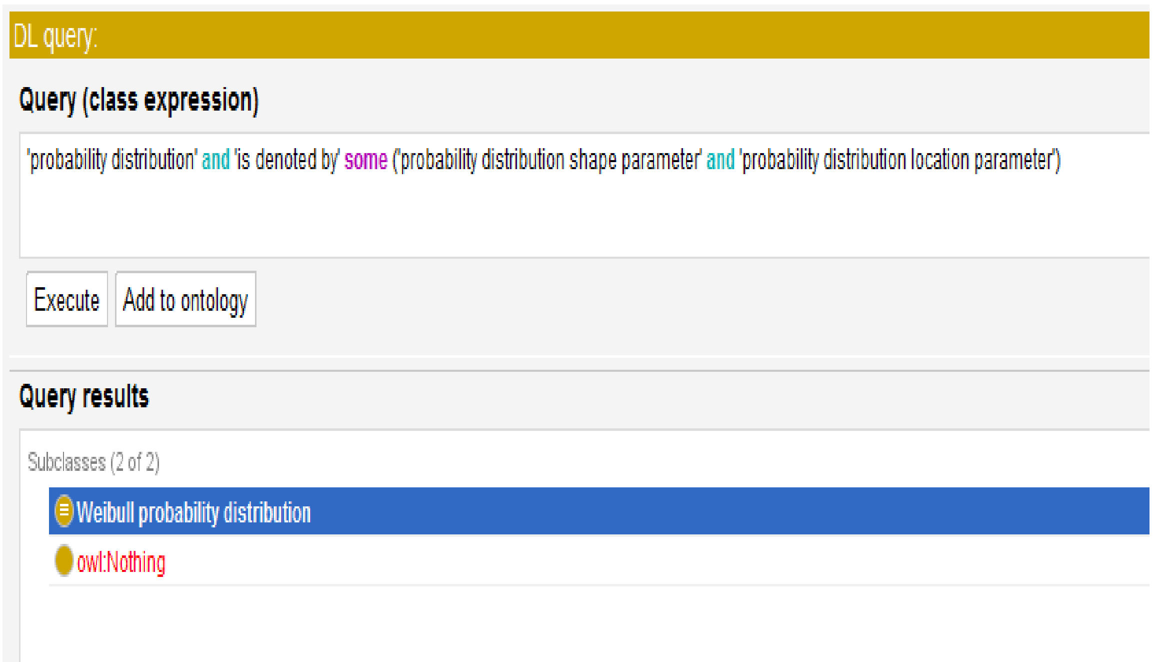


Figure 28 DL query on probability distribution and parameter criteria

5.6.2 Statistical Test and Parameters

- Which statistic tests are suitable on a probability distribution that requires all shape, scale, and location parameters?

  DL query: 'statistical model' and 'is denoted by' some ('is denoted by' some ('probability distribution location parameter' and 'probability distribution scale parameter' and 'probability distribution shape parameter'))

Result: Kruskal Wallis test; Mann-Whitney U-test (Figure 29)

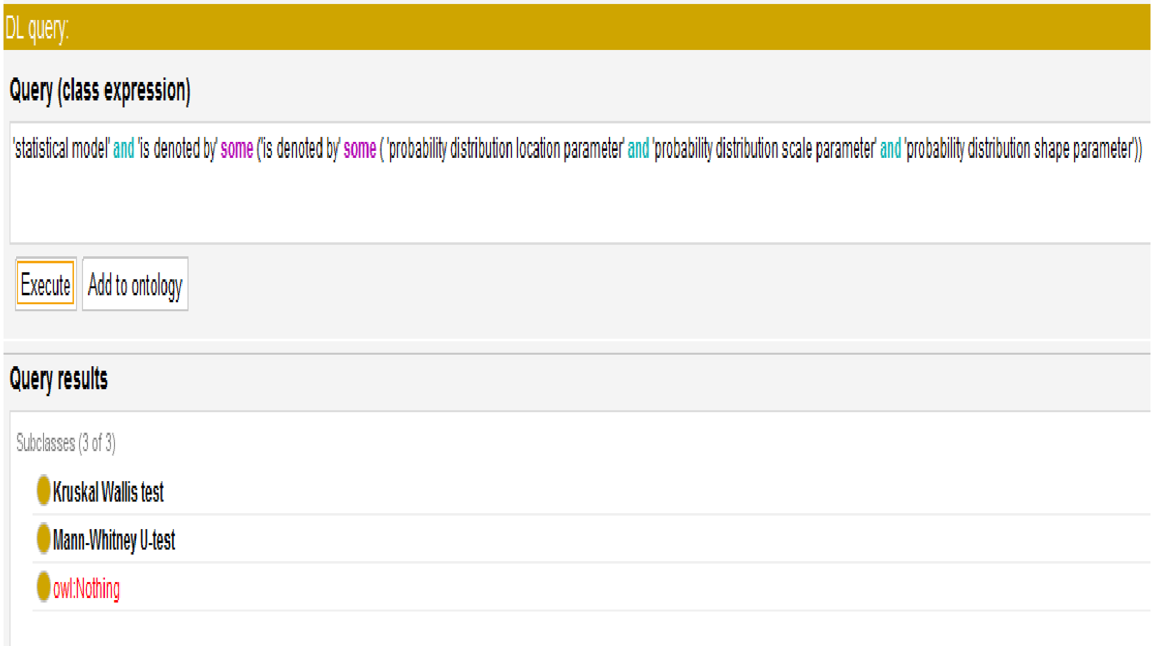


Figure 29 DL query on statistic test and parameters

5.6.3 Statistical Test and Data Types

- Which statistical tests can be used for ranking variables?

  DL query: 'statistical model' and 'has part' some ranking

  Result: Kruskal Wallis test; Mann-Whitney U-test; Wilcoxon signed-rank test; Non-parametric test (Figure 30)

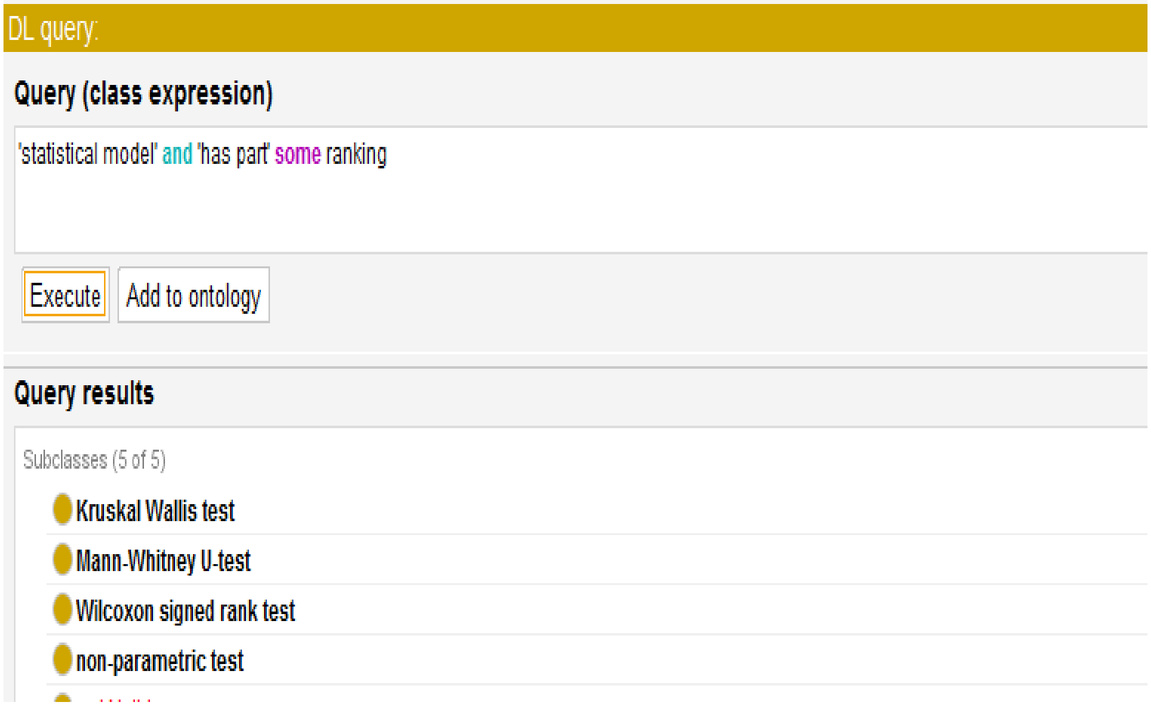


Figure 30 DL query on the statistic model for ranking

- Which correlation coefficient can be calculated using two ordinal variables?

  DL query: 'correlation coefficient' and 'is about' exactly 2 'ordinal variable'

  Result: Kendall’s correlation coefficient (Figure 31)

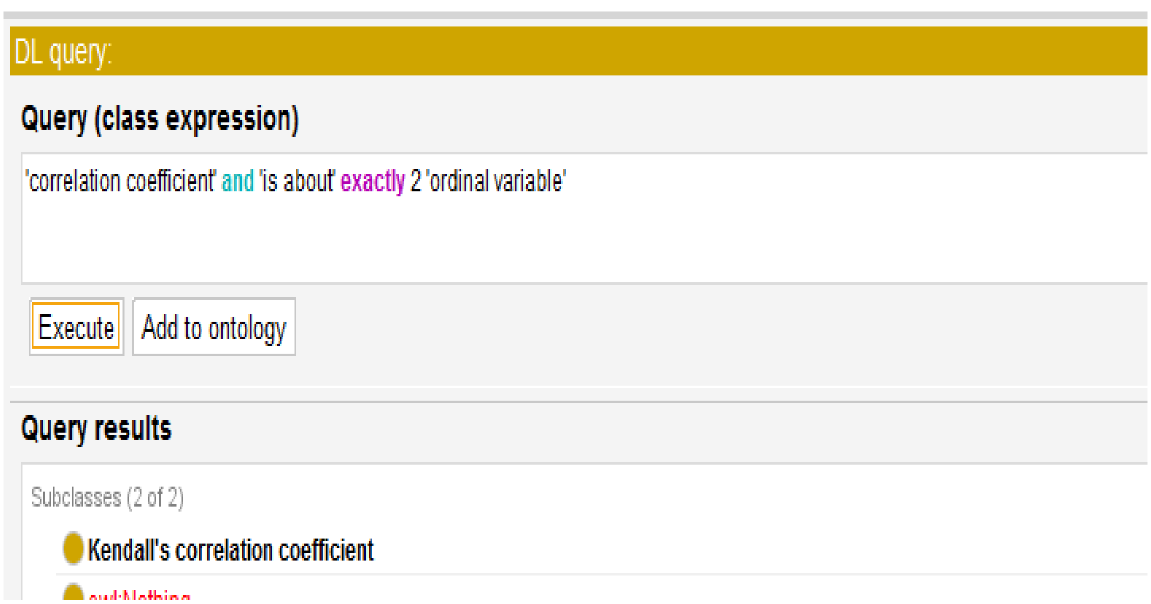


Figure 31 DL query on the statistic model for correlation

5.6.4 Statistical Test, Distribution and Hypothesis Test

- Which statistical tests can test for distribution fitting?

  DL query: 'statistical model' and achieves_study_objective some 'goodness of fit testing objective'

  Result: Exact binomial test, F-test (Figure 32)

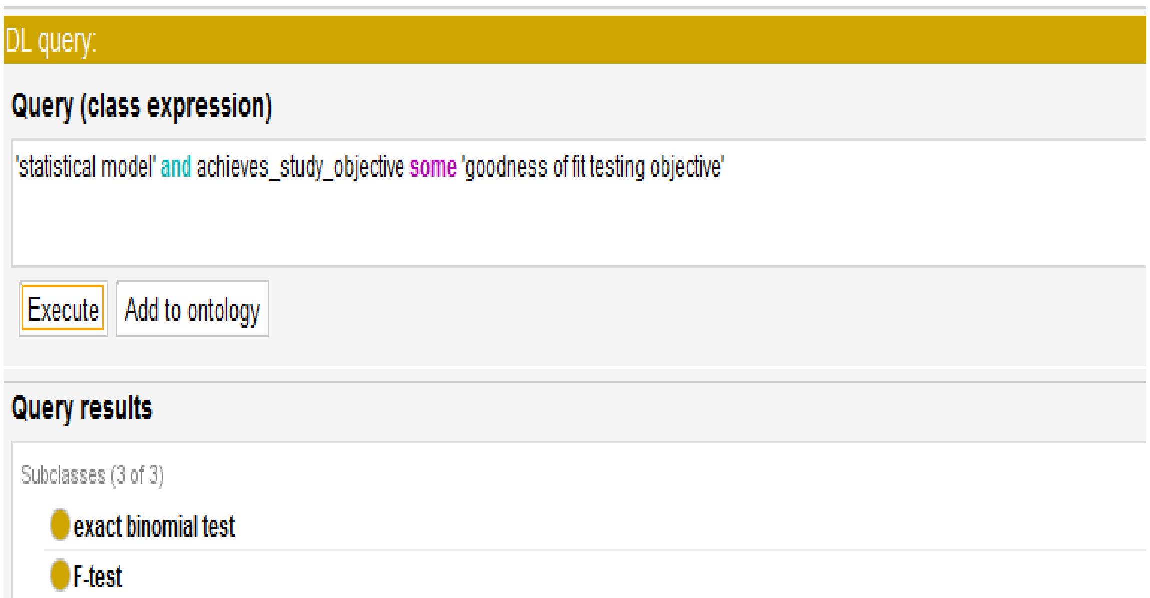


Figure 32 DL query on the statistic model for the goodness of fit test

- Which statistical tests can test for group comparison?

  DL query: 'statistical model' and achieves_study_objective some 'between-group comparison objective'

  Result: Kruskal Wallis test; Mann-Whitney U test; Non-parametric test; ANOVA (Figure 33)

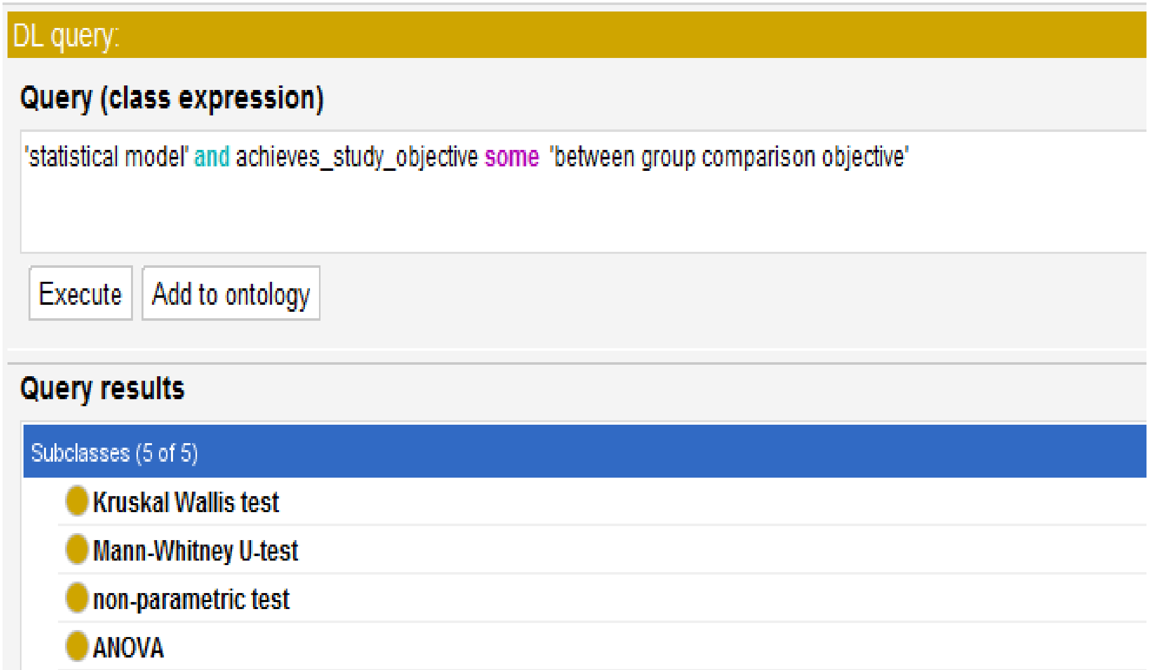


Figure 33 DL query on the statistic model for group comparisons

- Which statistical tests can test for group comparison with F-distribution?

  DL query: 'statistical model' and 'is about' some F-distribution and achieves_study_objective some 'between-group comparison objective'

  Result: ANOVA (Figure 34)

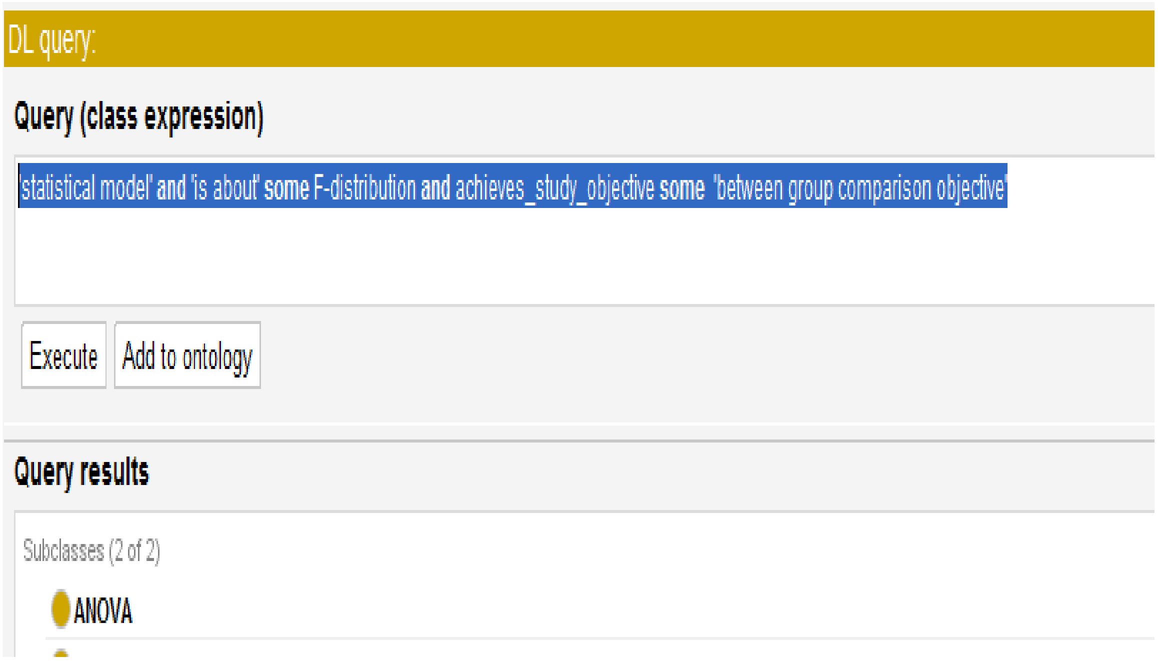


Figure 34 DL query on the statistic model for group comparison with the F distribution

5.6.5 Statistics and Research Design

- Which statistical test is eligible to test the relationship between categories variable in a randomized complete block design?

  DL query: 'statistical model' and 'is about' some 'randomized complete block design'

  Result: Cochran's q test for heterogeneity; ANOVA (Figure 35)

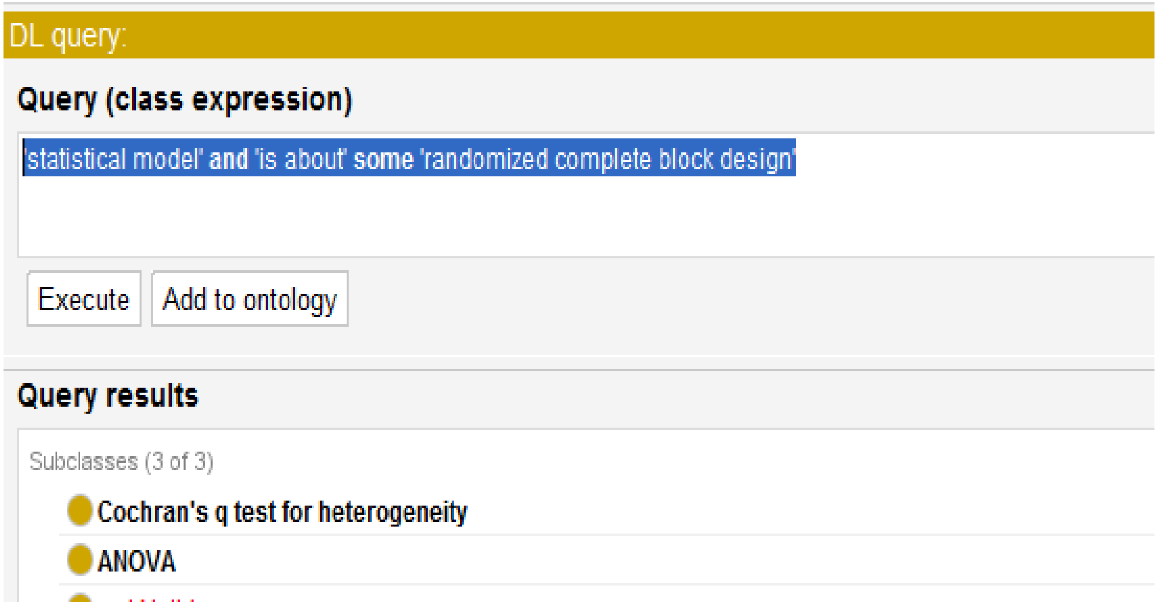


Figure 35 DL query on the statistic model for randomized complete block design studies

- Which statistical test is suitable for crossover study?

  DL Query: 'statistical model' and 'is about' some 'cross over design'

  Result: McNeMar's test; ANOVA (Figure 36)

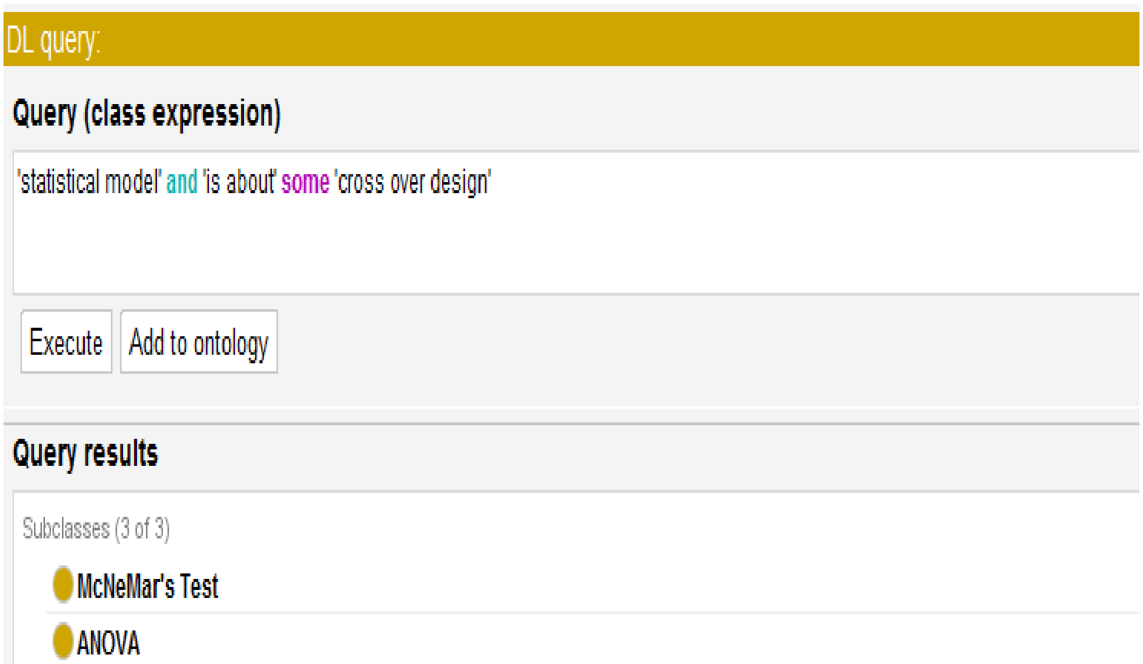


Figure 36 DL query on the statistic model for cross-over design studies

5.7 Conclusion

After reviewing existing statistics-related ontologies, the study discovered that OBCS and STATO are suitable for the ACLRO due to their alignment with BFO and OBI, as well as OBO Foundry principles. OBCS is primarily designed for statistics in biological, biomedical, and clinical fields, while STATO provides more statistics terms that are common for all domains. In addition, OBCS offers the terminologies related to clinical research, allowing the study to expand more specific terms. As a result, in the validation section, the study presented that the ontology model can be used for information queries not only for the statistic model but also for research design. The statistical concepts in ACLRO will continue to expand to more advanced mathematics like machine learning and deep learning to be compatible with the growth of Big Data.

CHAPTER SIX AIM 3 EVIDENCE-BASED PRACTICE ONTOLOGY

6.1 Introduction

Medicine and healthcare are a domain of complex systems due to its dynamic relations influencing the changes in the properties of other entities (Boon et al., 2007). Therefore, knowledge in medicine not only influences one another but also is evolving and adaptive. Evidence-based practice (EBP) has been considered the most standard knowledge-sharing approach in biomedicine and other domains by integrating the best external evidence into clinical expertise (Melnyk et al., 2014) that provides the outcomes for personalizing treatment (Oman, Duran, & Fink, 2008). The EBP term was introduced in 1991 as a novel approach in problem-solving and clinical decision making with scientific, clinically relevant research that replaces "intuition, unsystematic clinical experience, and pathophysiologic rationale (Waite & Killian, 2016).

Regardless of its long-term involvement and potentials, the implementation of EBP is not a simple process facing many challenges. The EBP process can be divided into five essential steps: (1) Formulate a clinical research question, (2) Search for the relevant evidence for the research question, (3) Evaluate and select the best evidence for validity and applicability, (4) Apply the findings to the local environment, and (5) Assess the result (Jenson & Howard, 2013). The first challenge is an inadequate understanding of the EBP process. Each step in the EBP process brings a specific outcome as an input of the next step. Without the right training, the missing actions might cause unsuccessful implementation (Rousseau & Gunia, 2016). The second challenge of EBP depends on the efficient information retrieval (IR) strategy, which relies on the clarity of the research question and the understanding of search engines. Additionally, the performance of a complete-article search is not offered in all search engines. For instance, PubMed and Google Scholar offer a search on title and abstract, but not on a full article. The issue can reduce the numbers of retrieved documents, leading to the IR performance in both precision and recall rate (Anders & Evans, 2010; Bramer, Giustini, Kramer, & Anderson, 2013). Another challenge is the hidden information of the knowledge of domain criteria, research design, and statistics method of the studies is not fully reported (Patrick et al., 2004), which can reduce the accuracy of the validation process and prevent the success of implementing the EBP findings to the local environment (Majid et al., 2011).

An ontology has been involved in medicine and healthcare, such as standard terminologies, such as SNOMED CT (El-Sappagh et al., 2018), and decision-making (Ishizu, Gehrmann, Minegishi, & Nagai, 2008). It offers a formal structure of entities and the relationship between them as existing in a specific domain. Each entity is defined with a set of properties representing its existence and characteristics. The properties also allow the integration of entities and concepts across disciplinary domains. As a result, the barriers to EBP can be lifted. The application of document search and evaluation can be embedded in the ontology and make the EBP process seamless and effective by the integration of concepts in medicine, research design, and statistic method into one framework. Furthermore, with the upper-level ontology, i.e., BFO, the ontology ensures a shareability of semantic information among other ontologies across domains.

With the BFO framework, the input and output of these two processes are considered continuant entities. Furthermore, the “information content” entity imported from the IAO serves as parent concepts of all related data involved in EBP processes.

6.2 Existing Ontologies

6.2.1 Evidence and Conclusion Ontology (ECO)

The Evidence & Conclusion Ontology (Riaño et al.) describes scientific-evidence types collected from laboratory experiments, computational methods, and literature within the biological research domain (Chibucos et al., 2017). ECO was initially developed using the Open Biological and Biomedical Ontologies (OBO) edit tool. Since 2016, ECO implemented development in the Web Ontology Language (Howland) using Protégé for viewing and editing on a small scale and ROBOT (http://robot.obolibrary.org) on a large scale. ECO also reuses and collaborates with other ontologies, such as Gene Ontology (GO), Ontology for Biomedical Investigations (OBI), Ontology of Microbial Phenotypes (OMP), and Synapse Gene Ontology Annotation Initiative (SynGO). ECO terms are grouped mainly based on (1) the biological context of the evidence, and (2) the technique used to generate the evidence. Some terms related to both categories. Therefore, ECO develops logical definitions of these terms under technique concepts linked to relevant assay-based OBI terms. As a result, the ECO model reduced the issue of ambiguous classes. ECO's aim is not abstract, nor is our study's purpose of storing knowledge conclusions for research evidence. See Figure 37.

Currently, the structure of ECO does not apply a foundation ontology in its model. Besides, most of ECO's terms are in biomedicine, such as generic fields rather than healthcare. Consequently, ECO is not compatible with the aim of this dissertation.

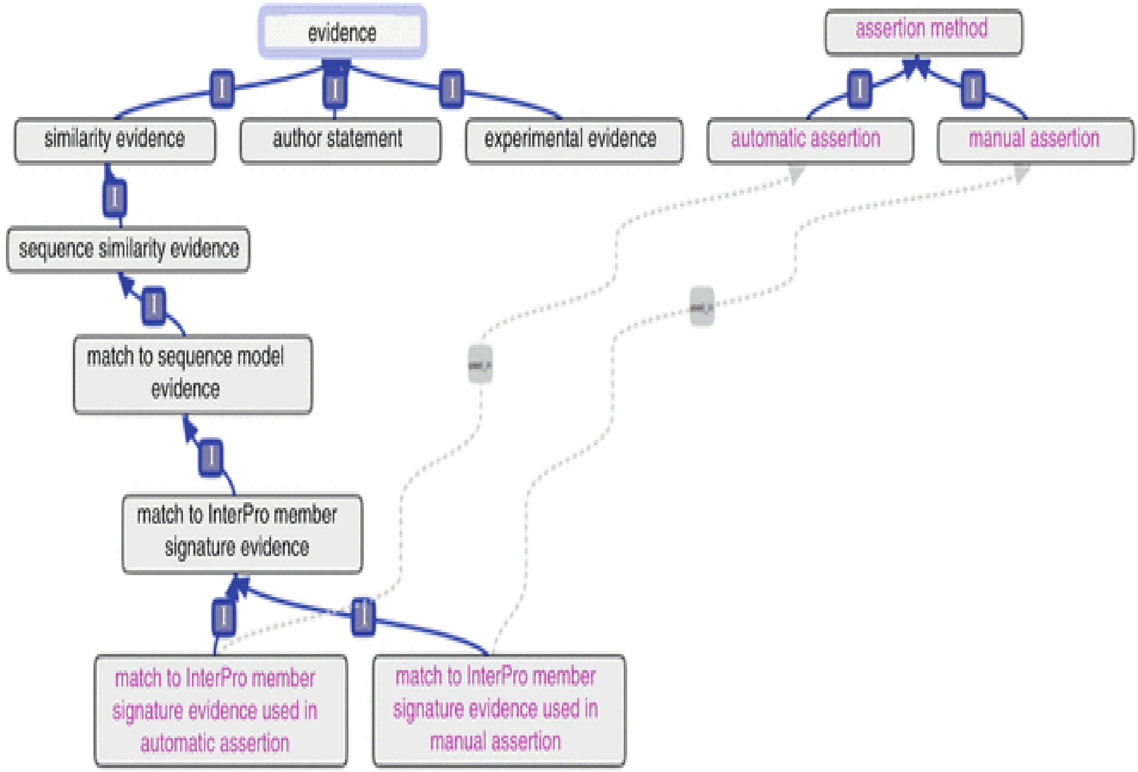


Figure 37 ECO's structure of evidence concepts (Chibucos et al., 2017)

6.2.2 Medical Literature Search Agent (MELISA)

MELISA is a prototype of an ontology-based information retrieval agent that recognizes the IR process's challenge due to the retrieval environment (Abasolo & Gmez, 2000). To formulate a well-designed query for IR, a detailed understanding of the retrieval environment is necessary. MELISA implemented an ontology model that generates queries and evaluates the results. It also allows a user to reformulate and review the results in the model. The overview structure of MELISA is shown in Figure 38.

Nevertheless, MELISA does not apply a top-level ontology to enhance its interoperability. Besides, the IR process is beyond the scope of this work. Therefore, there is no integration between the dissertation and MELISA at the current stage.

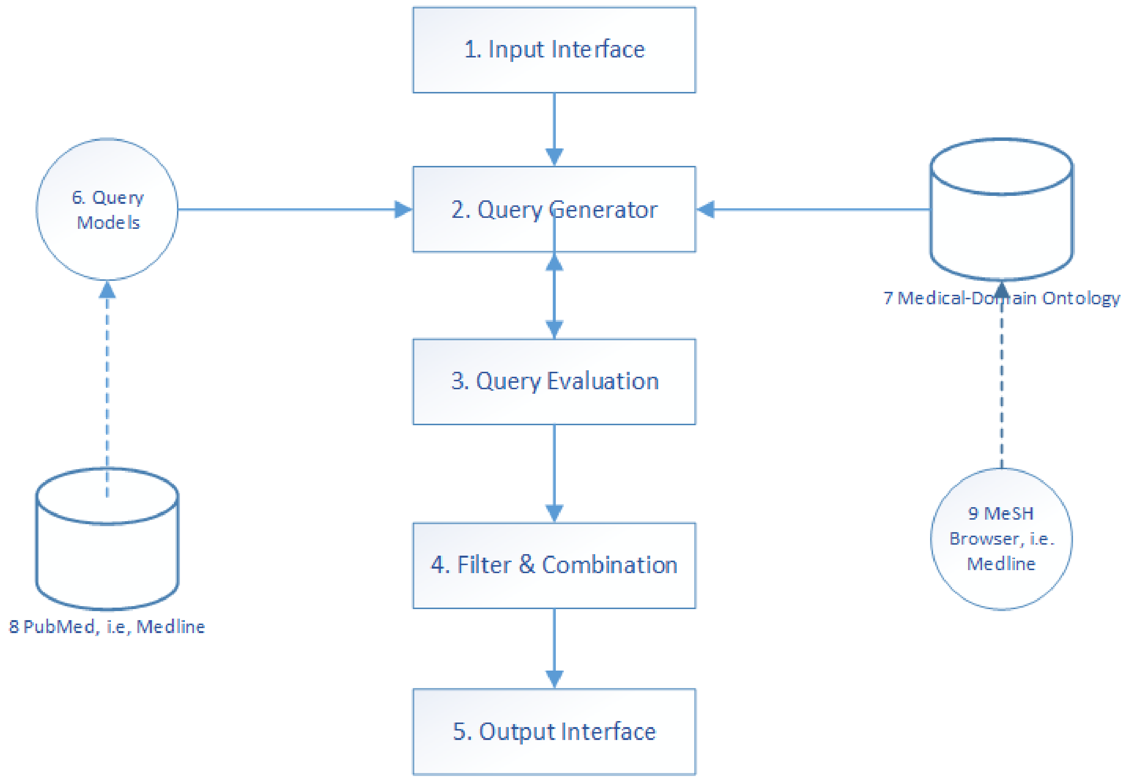


Figure 38 Overview structure of MELISA

6.2.3 EMBRACE Data and Methods (EDAM)

EDAM is an ontology of common concepts that have been developed in the scope (EMBRACE) to represent operations, data types, data identifiers, data formats, topics, and applications used in the bioinformatics community (Abasolo & Gmez, 2000). It provides controlled terms as semantic information that bridges the gap between service registries and service composition methodologies. EDAM contains over 3,400 terms and is considered the most comprehensive ontology for semantic annotations of web services that improve the data exchange between services. The EDAM structure can be categorized into five main concepts: Topic, Operation, Data, Format, and Identifier, as presented in Figure 39. The 'Topic' concept represents fields of bioinformatics study. The 'Operation' concept includes functions and methodologies of tools or services like web service operations. The 'Data' concept references to formal definitions of common data entities in the bioinformatics domain. The 'Identifier' concept presents the specific identification of bioinformatics entities. Lastly, the 'Format' concept describes data format specifications.

Due to the focus of EDAM in web services, its scope is not the dissertation's application. However, the structure of the ACLRO is compatible with EDAM and suitable for future integration.

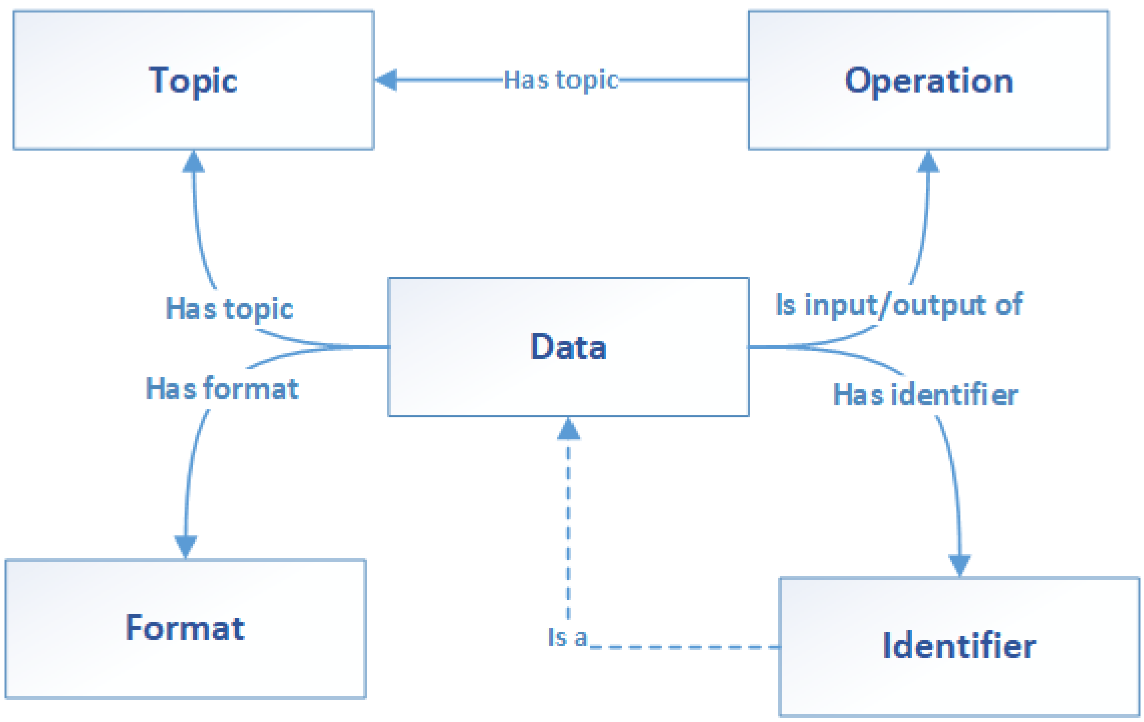


Figure 39 The overview concept diagram of EDAM ontology

6.3 Resource Description (RDF) Triples and Concept Mapping Diagram

In ACLRO, the EBP concept started at publication and ended at knowledge extraction. This required two processes, which performed extractions: "information extraction" and "drawing conclusion from data" processes. The information-extraction process allowed us to link data from the literature and external evidence to the local practice data. Binding these two together, we can develop a formal framework of an evidence-based model. The other key concepts here are documented data or information content. However, the information-content entity is not part of BFO2.0. It is an extended entity introduced by a BFO-based Information Artifact Ontology (IAO) that focuses on data collections and associates representational artifacts (Ceusters, 2012). In this study, the information-content class involved data collected during the health-intervention process and data extracted from the literature, as mentioned in Chapter 4.

6.3.1 Information Extraction Process

Ontology formal language is structured in Resource Description (RDF) triples (https://www.w3.org/TR/2002/WD-rdf-primer-20020319/):

<subject> -> [*predicate]* -> *<object>*

The RDF triples represent the information-extraction process of EBP can be reviewed as below:

Publication -> [part of] -> Information Extraction Process

Information Extraction Process -> [is about] -> Study Design

Study Design -> [declares] -> Objective Specification

Objective Specification -> [denotes] -> Stat Model

Stat Model -> [is model for] -> Study Variable

Study Variable -> [has role] -> Dependent Variable

Study Variable -> [is about] -> Data Set

6.3.2 Drawing a Conclusion from the Data Process

Publication -> [part of] -> Drawing a Conclusion from the Data Process

Drawing a Conclusion from the Data Process -> [has specified output] -> Document Parts

Drawing a Conclusion from the Data Process -> [has specified output] ->

Conclusion Based on Data

Conclusion Based on Data -> [denotes] -> Knowledge

6.3.3 EBP Concept-Mapping Diagram

The concept-mapping diagram presents the overview of EBP components, including the entities of publications, research design, research outcomes, and data points. The diagram successfully confirms the capability of the study's EBP on integrating all EBP's crucial entities into one framework following the structure of BFO, as presented in Figure 40.

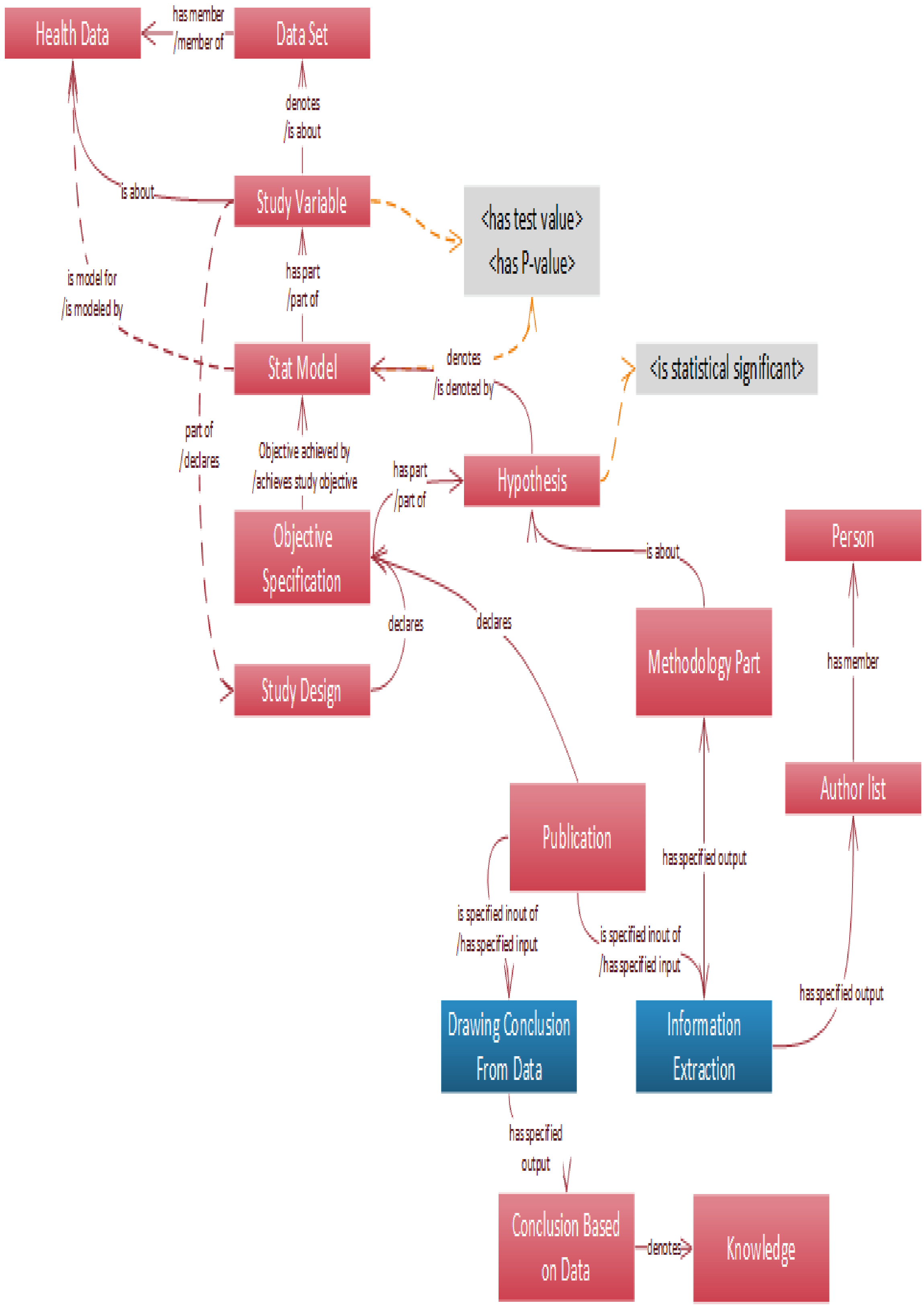


Figure 40 The overview of EBP’s concepts and their relationships in ACLRO

EBP's taxonomy structure under the BFO's continuant concept and occurrence concept is shown in Figure 41 and Figure 42, respectively.

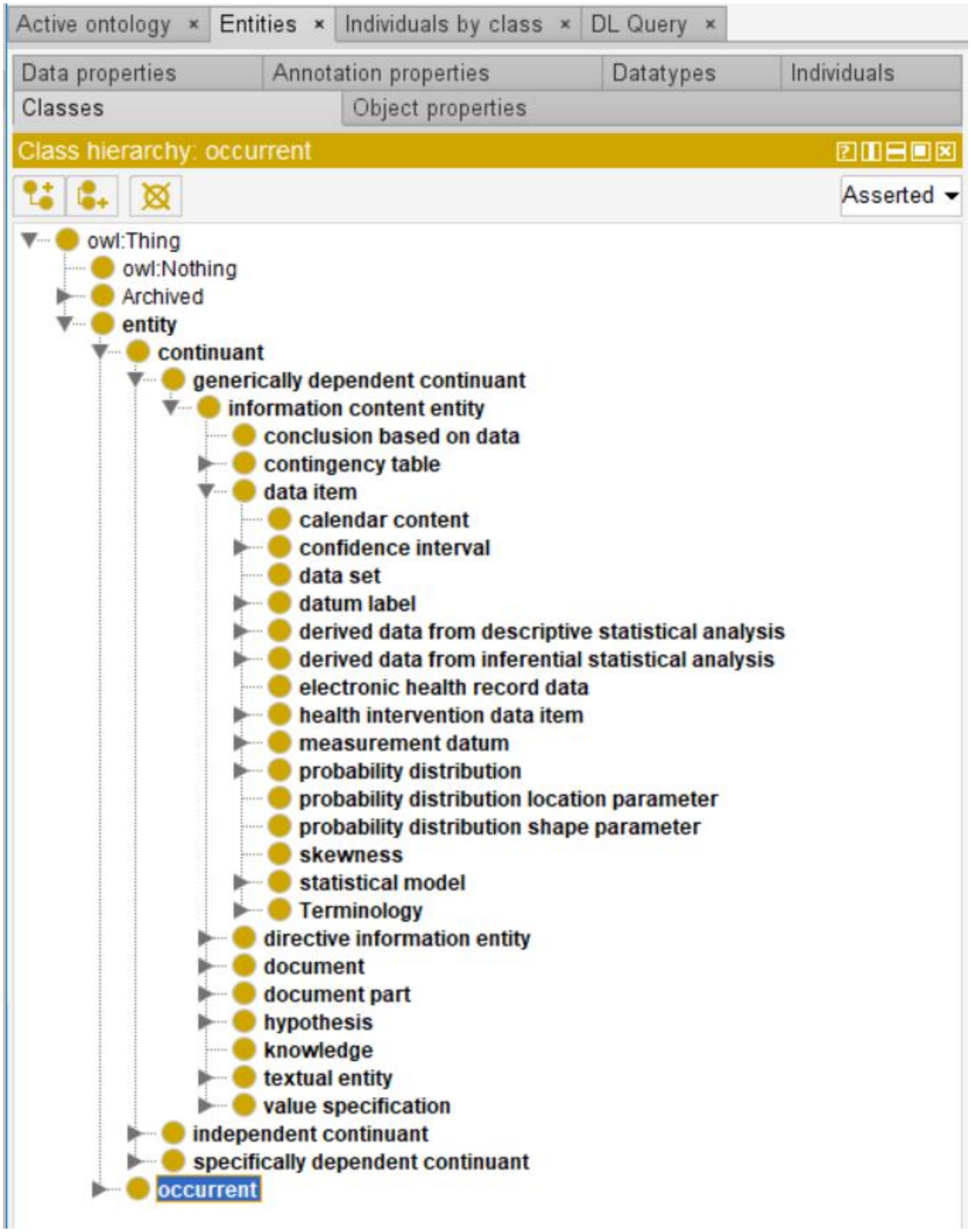


Figure 41 EBP's taxonomy structure under BFO's continuant concept

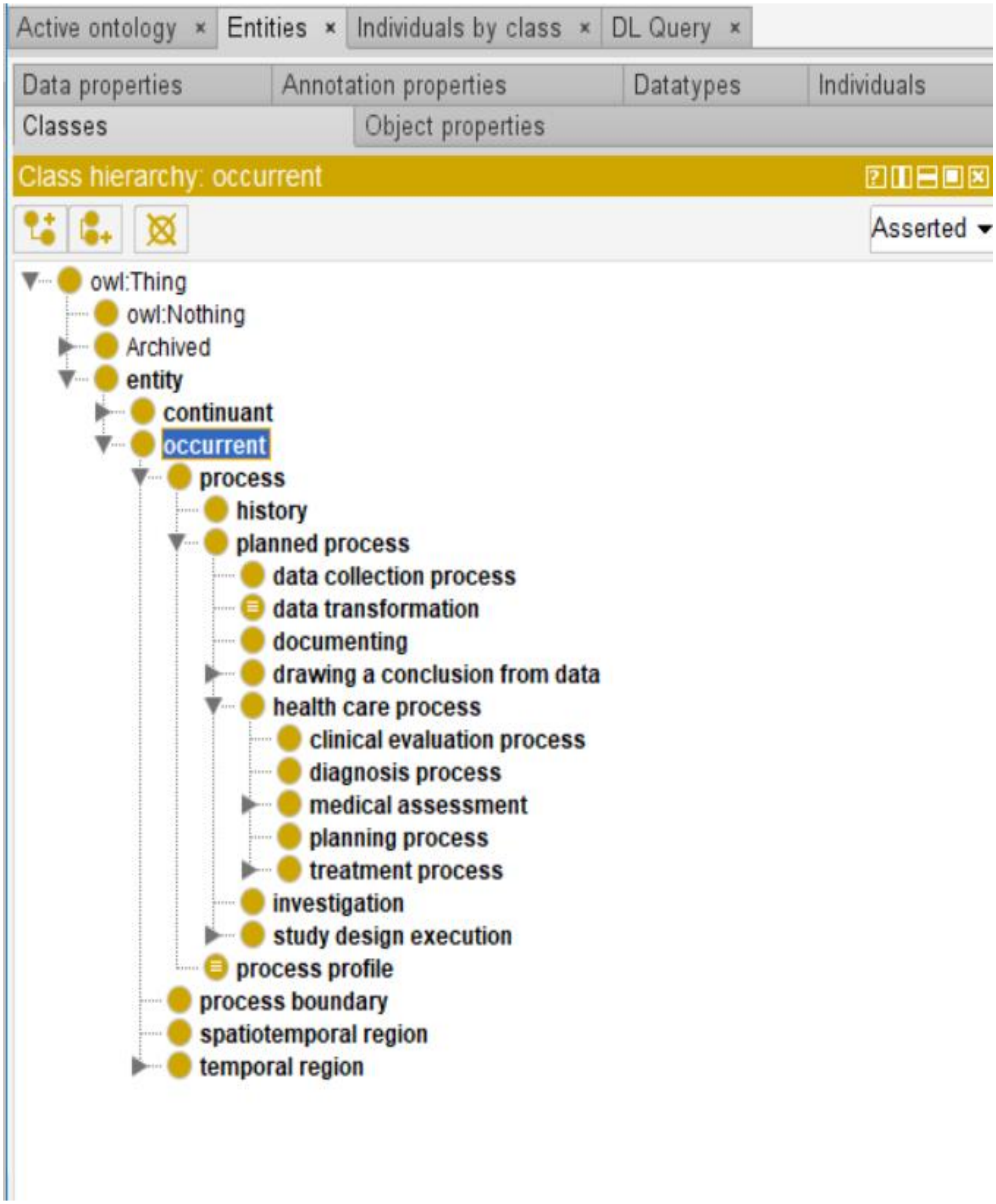


Figure 42 EBP's taxonomy structure under BFO's occurrent concept

6.4 Validation

6.4.1 Instance Creation

Instances or individuals are the smallest units of an ontology. While classes represent concepts, individuals are represented by a unique entity in a specific concept. The instances usually are not included in the domain representation. In this study, two instances of ACLR-related literature are implemented to support the validation process.

1. Publication A: Patient Characteristics and Predictors of Return to Sport at 12 Months After Anterior Cruciate Ligament Reconstruction: The Importance of Patient Age and Postoperative Rehabilitation (Edwards et al., 2018).
   Objective: To investigate factors predictive of return to sport 12 months after ACLR. The factors specifically evaluated were strength, hop function, self-reported knee function, patient age, and quality of postoperative rehabilitation.

   Result: Complete rehabilitation

   OR = 7.95; P-value = 0.009

   Age <= 25 yr – OR = 3.84; P-value = 0.024

   IKDC score – P-value = <0.001

   Conclusion:

   Higher IKDC scores were predictive of return to sports (RTS)

   Younger patients were predictive of RTS

2. Publication B: Return to Play and Future ACL Injury Risk After ACL Reconstruction in Soccer Athletes from the Multicenter Orthopedic Outcomes Network (Smucny, Westermann, Group, & Group) Group (Brophy et al., 2012)

Objective:

a) Test the hypothesis that player's sex does not influence RTP

b) Test the hypothesis that player's age does not influence RTP

Result:

a) OR = 0.3; P-value = 0.037

b) TTEST; P-value = 0.006

Conclusion:

a) Females were less likely to RTS than males

b) Older athletes were less likely to RTS

Figure 43 shows the Publication-A instance's formal definition and its attributes through Protégé, an open-source ontology editor.

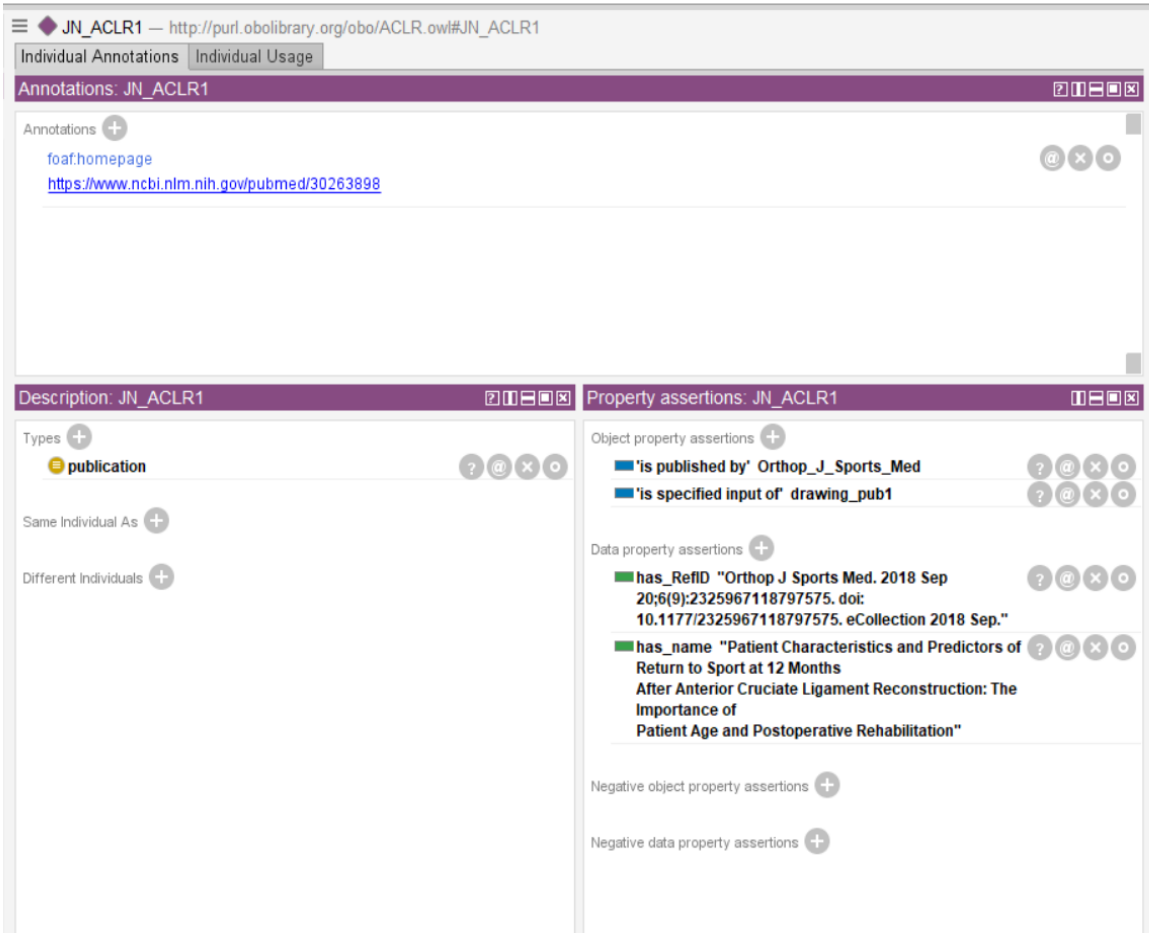


Figure 43 Formal structure and attributes of Publication-A

6.4.2 Competency Testing

In this step, the study used the list of questions to validate and check for the model implementation. The two questions to be decided should detect relations across concepts. For instance, the two questions below detect whether the publication and knowledge extractions are connected and queried as below:

1. Which publications were researched on predictive risk analysis related to return to the sport in ACLR surgery?

   DL query: publication and declares some predictive_risk_analysis and is_relevant some ACLR_surgery and is_relevant some return_to_sport_month

Result: The model can select the publications that are supportive of the criteria. In this study, Publication A is the only study on the predictive risk analysis of RTS after ACLR (Figure 44)

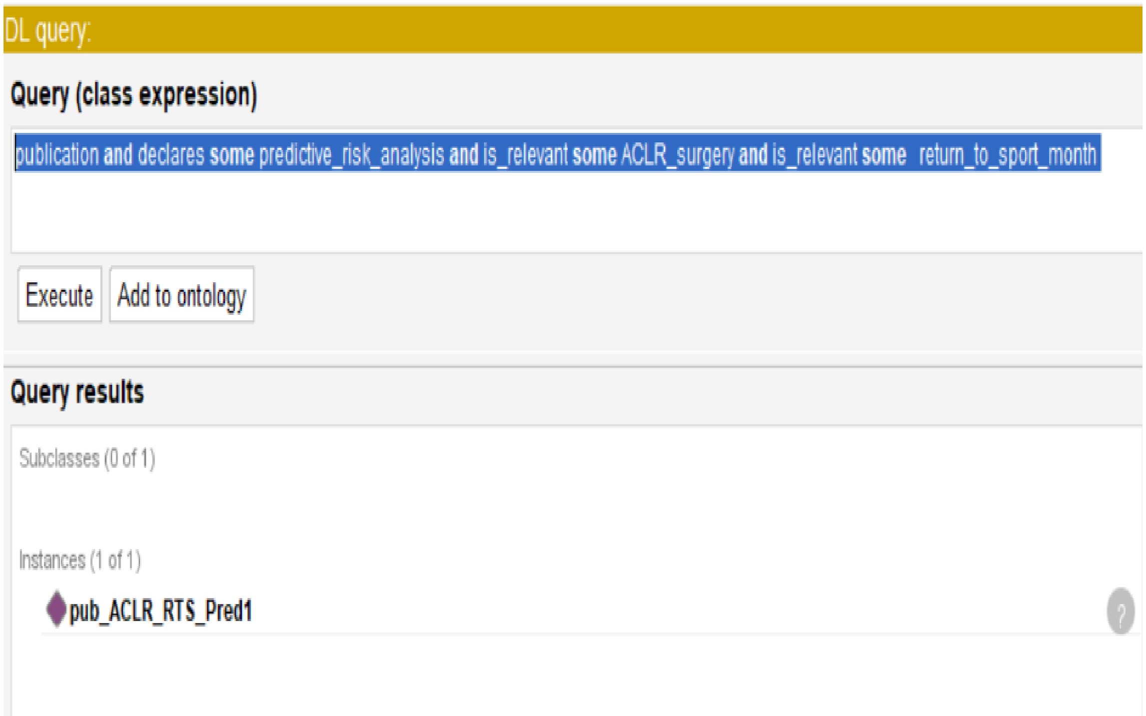


Figure 44 DL query on predictive risk analysis of RTS in ACLR

2. Which EBP knowledge was discovered in return to sport after ACLR surgery?

   DL query: knowledge and is_relevant some ACLR_surgery and is_relevant some return_to_sport_month

   Result: The model can select all knowledge to be reported in the literature relevant to RTS after ACLR surgery (Figure 45)

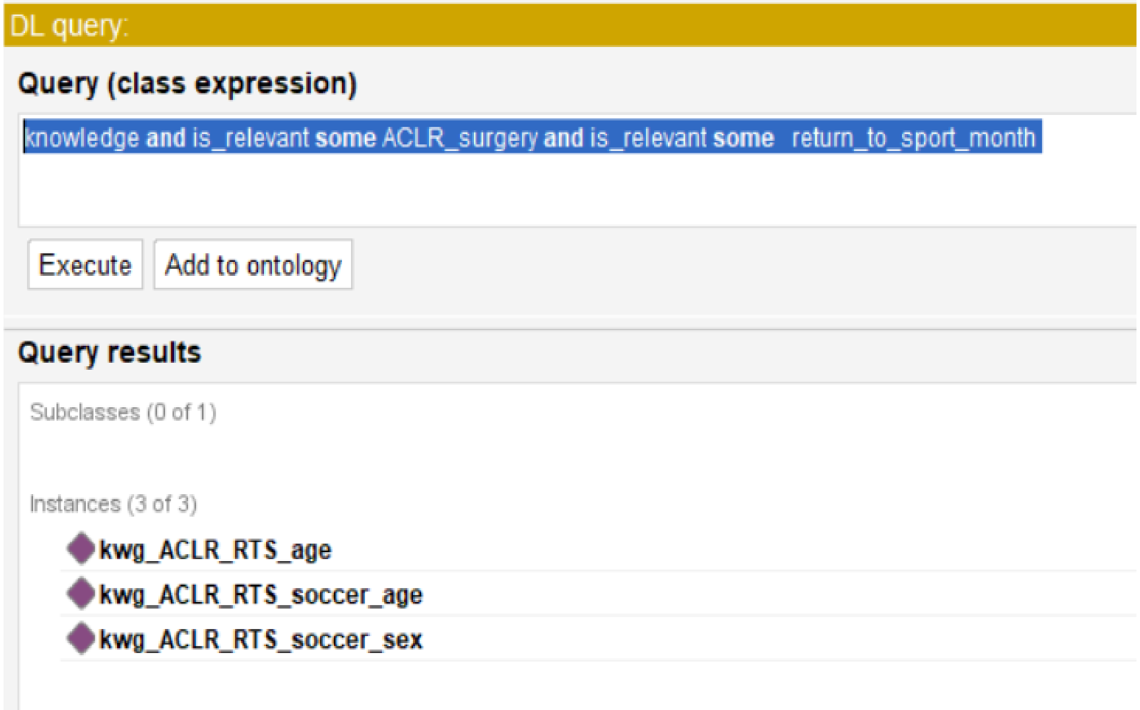


Figure 45 DL query for knowledge discovery on RTS in ACLR

6.5 Conclusion

The study presented the implementation of an evidence-based knowledge model in the ACLR rehabilitation domain to comply with the shareable semantic ontological framework. The model simultaneously emphasized both patient-focus treatment processes and external evidence. With the foundation ontology, individual ontology models can be implemented separately and then merged for greater use and a more meaningful model later, as shown in this study. It not only allows the model to be exchangeable but also eases the process of expanding and extending both domain concepts in the model since the classes can grow independently. The study also enhanced the ontology model's functionalities, allowing knowledge queries derived from both domains. Further study related to adding a practice-based domain can lead to an ideal structure of base practice in medicine.

CHAPTER SEVEN AIM 4 SHARABLE BEST-PRACTICE ONTOLOGY

7.1 Introduction

The medicine and healthcare domain is a complex system composed of many agents and components that interact with one another. The primary goal of medicine is not to build a standardization, but to understand the disease, find a cure, improve patients' health outcomes, and improve healthcare quality (Black, 2013). Researchers profoundly invest in new technology to find more effective drugs and robust devices to treat patients. As a result, there are multiple protocols, medications, and treatments to deal with the same disease. The discovered knowledge might be agreeable or disagreeable (Kelley, Moy, Stryer, Burstin, & Clancy, 2005). With a flush of electronic health record data, the healthcare domain is in the world of big data. The most challenging aspect of big data is analyzing it (Labrinidis & Jagadish, 2012). An ontology plays a vital role in organizing big data in an efficient format that can be understood by both human and computer systems. It can also make information interpretatively exchangeable and sharable (Gruber, 2018). An ontology captures the semantic information of concepts used in a specific domain. The critical success of ontology implementation depends on its consistency, completeness, and granularity of the model, leading to exchangeability (Obrst et al., 2014). A domain-specific ontology model should be implemented based on a formal foundation, i.e., an upper-level or top-level model. The core concepts of upper-level models provide standard terms and formal structures across domains. The sufficient upper-level ontology must be small and generalized, allowing for consistency and shareability across different domains (Degen et al., 2001). The dissertation applies both top-down and bottom-up approaches to enhance the domain's semantic definition and shareability of the ACLRO.

Another crucial component of the ACLRO is to design an ontology model for the best-practice setup. The best practice approach aims to enrich both EBP and PBE. EBP applies scientific evidence as to its guidance and decision support. The goal of EBP is to interpret the best suitable evidence from systematic research and integrate the results into clinical expertise and the environment (Sackett et al., 1996). Systematic research collects and summarizes all reported evidence to answer a defined research question linking to a knowledge area. With the BFO framework, the input and output of these two processes are considered continuant entities. Furthermore, the "information content" entity imported from the IAO serves as parent concepts of all related data involved in EBP processes. The predicate of the EBP processes is presented in Chapter 4 - 4.3.2.

7.2 Practice-Based Evidence (PBE) Ontology

Some might say the use of EBP is becoming a standard of patient care (Ellis, 2019). PBE is defined as a relatively new procedure for gathering good-quality data from routine practices in real-word settings with trial and error (Margison et al., 2000). PBE initiates innovation and knowledge discovered in healthcare that has objective support based on community values.

In the study, the core entity of PBE is the "Investigation" process (OBI_0000066) imported from OBI. The investigation process is a planned process that consists of parts: planning, study design execution, documentation, and concluding. Therefore, the investigation process is an assembly of sub-processes, contributing to a research study. The visualization of the PBE concept diagram is shown in Figure 46.

EBP and PBE share some similarities and differences. The information content entities, in place of study design, objective specification, statistical model, and variables, are consistent in both models. Both models contain the “Drawing Conclusion from Data” planned process that links to “Study Design.”

The main difference between both models is where the information initially generates. EBP starts at the “Publication” entity, while PBE begins at the “Investigation” process. Using the upper-level ontology, ACLRO can reuse the entities in similar ways, and then merge the differences into one framework.

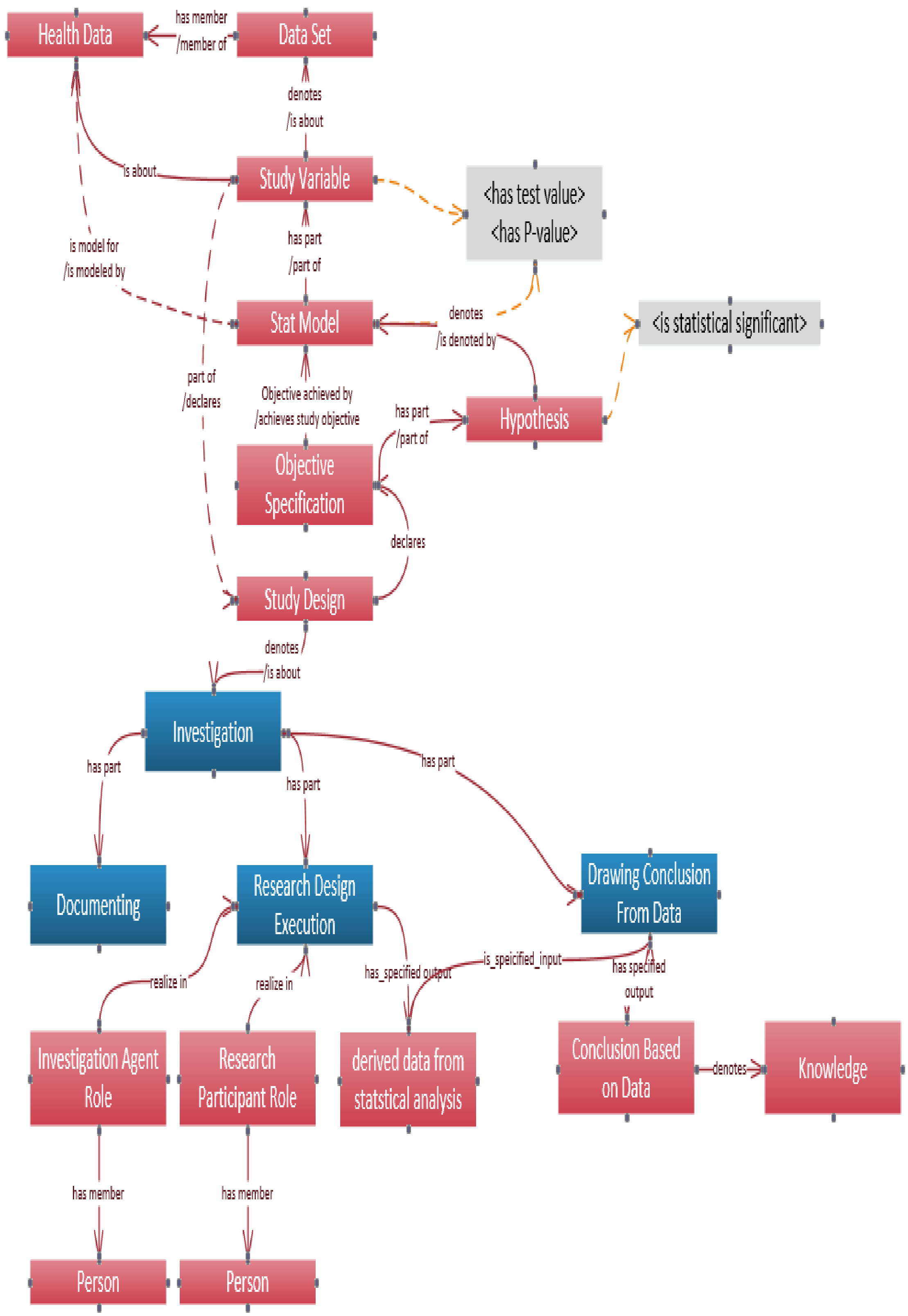


Figure 46 The concept diagram of PBE ontology

7.3 Sharable Best Practice (BP) Ontology

BP is defined through comprehensive existing research studies and experts' experience. The combination of EBP and PBE increases patient care quality and leads healthcare to the "best practice" concept (Melnyk & Fineout-Overholt, 2011). In general, it can be a challenge to directly apply the knowledge gained from EPB to a practice's workflow. Furthermore, the lack of understanding of individual research environments can prevent a consistent outcome with the original study (Perleth, Jakubowski, & Busse, 2001). On the other hand, best practices syndicate EBP and PBE by customizing existing knowledge and proven evidence to its practice environment and continuing the knowledge discovery without requiring meeting the scientific standard. Hence, best practice integrates EBP and PBE knowledge into one single outcome representing the reusable and sharable ontology framework, obtaining external and internal evidence from both. The visualization of the syndicated model is presented in Figure 47.

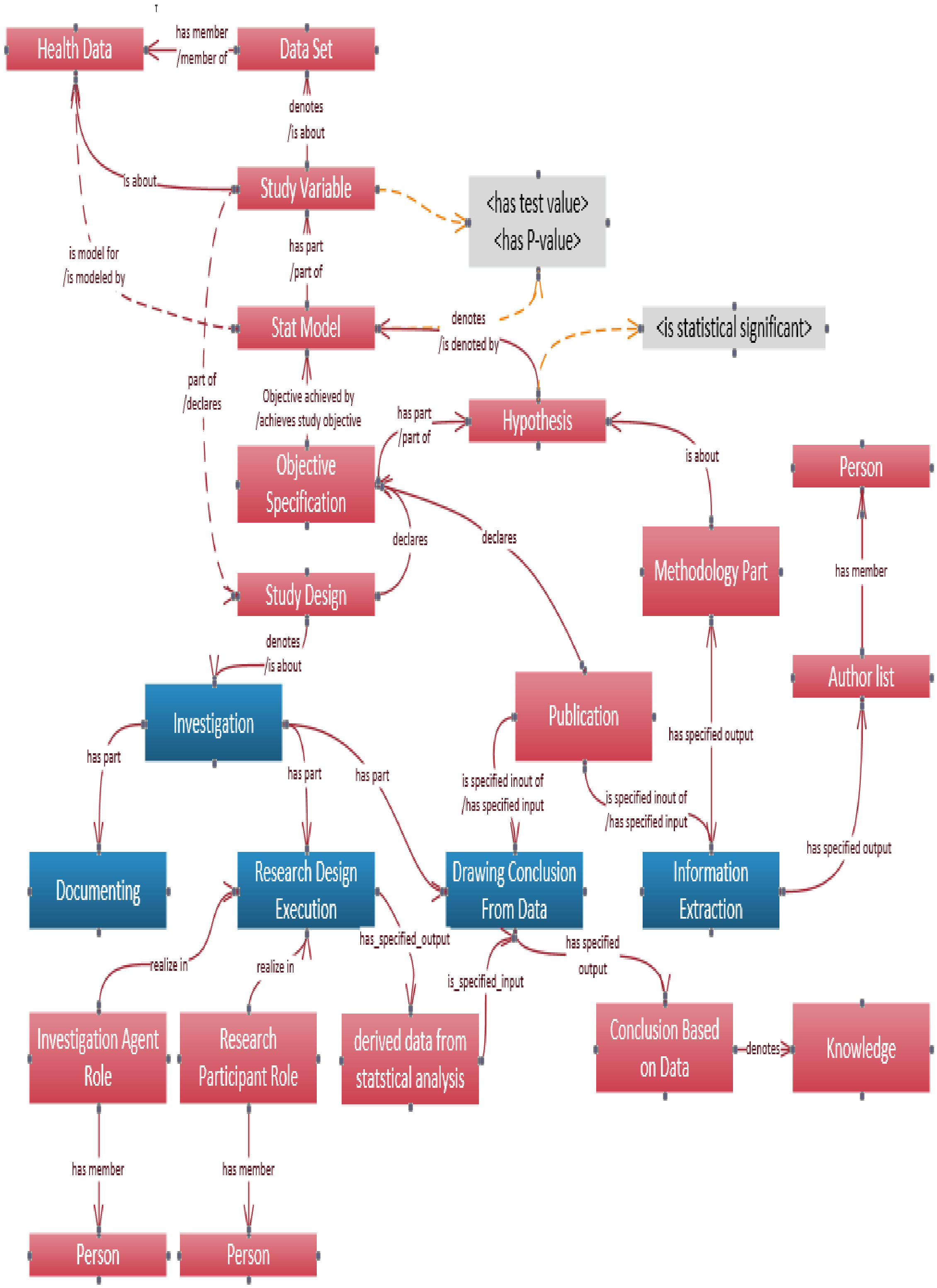


Figure 47 The concept diagram of BP ontology with the integration of EBP and PBE

7.4 Conclusion

The study demonstrates that upper-level ontology increases reusability and shareability over the button-up approach. With foundation ontology, individual ontology models can be implemented separately and then later was merged for greater use and a more meaningful model, as shown in this study. EBP and PBE have different unique characters. EBP offers knowledge derived from scientific research, while PBE captures knowledge gained from practices over time. Combining PBE and EBP processes allows individual practices to customize their scientific learnings for optimal outcomes, called the best practice approach. In addition, the study enhances the ontology model's functionalities, allowing knowledge queries across multi-disciplinary areas, such as medical data, statistical models, study design, and scientific knowledge.

CHAPTER EIGHT CONCLUSION AND FUTURE WORKS

8.1 Introduction

Healthcare is one of the industries impacted by a fast change in technology, as shown in a rapid EHRs adoption rate since 2011. Accordingly, the data and information's availability increase in volume, variety, and velocity are a definition of big data. Big data have been an infamous term in medicine analytics, leading to a demand for knowledge extraction and exchange. There are two main approaches to knowledge discovery in medicine: evidence-based practice (EBP) and practice-based evidence (PBE). EBP allows the practices to learn from the best external evidence, whereas PBE gains knowledge from its researches. The semantic mapping between EBP and PBE can combine both advantages for the best-practice knowledge model. The reusability and exchangeability of knowledge are the critical success of semantic mapping, which is not a simple task due to the variations in healthcare. The lack of standardization, such as in terminologies and patient-intervention protocols, prevents the industry from moving forward in knowledge sharing across practices.

The study implemented ontological engineering for knowledge architecture design that provides a sharable best-practice framework and the support of the foundation ontology model. First, the study presents how the traditional knowledge model is built for the ACLR domain while discussing the limitations of its design with the bottom-up approach. Then, the study proposes the top-down approach with the foundation ontology to improve the shareability and reusability of the ontology model. Finally, the ACLR-Rehabilitation Ontology (ACLRO) is implemented to apply both the top-down and bottom-up approaches.

8.2 Impact of the Research

1) The study demonstrated the guidance of knowledge capturing and formal representation in a triplet format using a combination of Toronto Virtual Enterprise (TOVE) ontologies and the foundation ontology, i.e., Basic Formal Ontology (BFO).

2) The comparison between the bottom-up and top-down approaches evaluated the advantages and disadvantages of both approaches. The bottom-up approach is easier to use a starter for domain-knowledge abstracting, while the top-down approach requires a new understanding of the foundation ontologies. Some of the domain concepts might not fit in any class definition of the foundation ontologies. However, the shareability requires the formal definitions provided in the foundation ontologies.

3) The study revealed a unique benefit of ontological knowledge engineering that permitted EBP, PBE, domain, and statistic models to be built separately and integrated into the final model without re-work on existing models. This work proved that the ontology model with the foundation structure could be reusable and sharable.

4) The study proposed the final ACLR ontology (ACLRO) model as a proof-of-concept prototype to demonstrate how the ontology can be applied to a real-world setup with the ability to link different types of semantic knowledge storing in EBP and PBE for the success of the best-practice model.

8.3 Challenges and Limitations

1) The first challenging is a learning curve of BFO and ontology languages like OWL and SWRL. BFO is very strict on the class definition. Also, some domain entities do not fit well in the BFO framework. As a result, it is challenging to classify domain entities under upper ontology classes correctly.
2) Due to its popularity, the study selected Protégé as ACLRO's implementation tool. Whereas Protégé has many useful built-in functions like annotation, query, visualization, reasoning, and importing, it also offers some essential plugins to the tool such as RDBMS integrating and decision trees. Nonetheless, many of these plugins are not reliable and lack of supporting documents. For that reason, the author avoided employing additional plugins.
3) The study manually created instances in Protégé for the use cases for publications and research processes. The processes were time-consuming and caused the model bloated.
4) The last limitation is that the ACLRO model is designed and implemented by a single researcher, and the scope of study involves multiple areas, i.e., ACLR protocols, statistics, research designs, and ontology implementation.

8.4 Future Works

1) The model can improve its shareability by adding meta-annotations, which add semantic information to the model. This process is significant for the semantic web.
2) As mentioned above, one of the study's challenges is that the current model cannot connect to RDBMS. A triplestore is a database management system

for the RDF data model that can connect the ACLRO model to RDBMS by storing the model in the RDF triplet format.

3) The study will continue adding statistics and machine-learning models into its statistic framework.

APPENDICES

Appendix A

R Code for Random Forest

```
---
title: "R-Random Forest (Caret)"
output:
  html_notebook: default
  html_document: default
---

##### ---------------------------------------------------------------------------
##   Load Data:ACL Reconstruction

```{r echo=FALSE, message=FALSE, warning=FALSE, cache=FALSE,
paged.print=FALSE}
setwd("I:\\IUPUI\\A Thesis\\R\\Data")
aclr.imp <- read.csv("aclrNoyes_Dec2019.csv",header=TRUE, sep=",")

```

Create Train and Test datasets

```{r warning=FALSE}
if (!require("caret")) install.packages(caret)
library(caret)
#if (!require("lattice")) install.packages(lattice)
#library(lattice)

# Medv is continuous variable
set.seed(123)
split <- createDataPartition(y=aclr.imp$pain_2yr, p=0.7, list=FALSE)
train <- aclr.imp[split,]
test <- aclr.imp[-split,]

```

Decision Tree (Tree)

```{r warning=FALSE}
library(tree)
aclrTotal.tree = tree(total_2yr ~ SEX + TYPE + cartilage+srgage+ injsrg_mth + inj_type
                  + lat_rem + med_rem
           , data=train)
```
```

```
summary(aclrTotal.tree)
plot(aclrTotal.tree)
text(aclrTotal.tree, pretty=0)

aclrPain.tree = tree(pain_2yr ~ SEX + TYPE + cartilage+srgage+ injsrg_mth + inj_type
                + lat_rem + med_rem
          , data=train)
summary(aclrPain.tree)
plot(aclrPain.tree)
text(aclrPain.tree, pretty=0)

aclrStability.tree = tree(stability_2yr ~ SEX + TYPE + cartilage+srgage+ injsrg_mth +
inj_type
                + lat_rem + med_rem
          , data=train)
summary(aclrStability.tree)
plot(aclrStability.tree)
text(aclrStability.tree, pretty=0)

```
```
##   Decision Tree (RPART)

```{r warning=FALSE}

library(rpart)
aclrTotal.rpart = rpart(total_2yr ~ SEX + TYPE + cartilage+srgage+ injsrg_mth +
inj_type
                + lat_rem + med_rem
                    , data=train, method = "anova")
summary(aclrTotal.rpart)
plot(aclrTotal.rpart)
text(aclrTotal.rpart)
printcp(aclrTotal.rpart)
plotcp(aclrTotal.rpart)

aclrPain.rpart = rpart(pain_2yr ~ SEX + TYPE + cartilage+srgage+ injsrg_mth +
inj_type
                + lat_rem + med_rem
                    , data=train, method = "anova")
summary(aclrPain.rpart)
plot(aclrPain.rpart)
text(aclrPain.rpart)
printcp(aclrPain.rpart)
plotcp(aclrPain.rpart)
```
```

````
aclrStability.rpart = rpart(stability_2yr ~ SEX + TYPE + cartilage+srgage+ injsrg_mth +
inj_type
              + lat_rem + med_rem
                 , data=train, method = "anova")
summary(aclrStability.rpart)
#plot(aclrStability.rpart)
#text(aclrStability.rpart)
printcp(aclrStability.rpart)
plotcp(aclrStability.rpart)

```
##### --------------------------------------------------------------------------------
##   RANDOM FOREST
#####  RandomForest (Package: randomForest)
##### (1) Build Random Forest Trees
#####   Note: mtry = the number of parameters used in each split.  The recommedation
is p/3 for regression trees, and sqrt(p) for classification trees
#####   In this example, we set mtry=4
#####  (2) Calculate the Yhat value
#####  (3) Calculate the residuals (=error rate)
#####  (4) View the importance of each variable
#####   %IncMSE is based on the mean decrease in accuracy in predictions on the out of
bag samples, when the given variable was excluded from the model.
#####   IncNodePurity is a measure of the total decrease in node impurity that results
from splits over that variable, averaged over all trees. f
#####  (5) Plot the importance

Total
```{r warning=FALSE}
if (!require("randomForest")) install.packages(randomForest)
library(randomForest)
# (1) Build Random Forest Trees
aclrpain.rf <- randomForest(pain_2yr ~ SEX + injsrg_mth + cartilage+srgage+ inj_type
              + lat_rem + med_rem
              , data=train, mtry=4, importance=TRUE,na.action = na.omit)
# (2) Calculate the Yhat value
yhat <- predict(aclrpain.rf,test)
# (3) Calculate the means of errors (=error rate)
mean((yhat-test$total_2yr)^2)
# (4) View the importance of each variable (Note:if the importance=TRUE is omitted,
the %IncMSE will not be shown in importance()
importance(aclrpain.rf)
# (5) Plot the importance
varImpPlot(aclrpain.rf)

plot(aclrpain.rf)
````

````
summary(aclrpain.rf)

```

```{r}
#########

aclrTotal.lm <- lm(total_2yr ~ SEX + injsrg_mth  + cartilage+srgage+ inj_type
              + lat_rem + med_rem
              , data=train)

summary(aclrTotal.lm)
```
````

Stability

````
```{r warning=FALSE}
if (!require("randomForest")) install.packages(randomForest)
library(randomForest)
# (1) Build Random Forest Trees
aclrStability.rf <- randomForest(stability_2yr ~ SEX + TYPE + cartilage+srgage+
inj_type
              + lat_rem + med_rem
              , data=train, mtry=4, importance=TRUE,na.action = na.omit)
# (2) Calculate the Yhat value
yhat <- predict(aclrStability.rf,test)
# (3) Calculate the means of errors (=error rate)
mean((yhat-test$stability_2yr)^2)
# (4) View the importance of each variable (Note:if the importance=TRUE is omitted,
the %IncMSE will not be shown in importance()
importance(aclrStability.rf)
# (5) Plot the importance
varImpPlot(aclrStability.rf)

plot(aclrStability.rf)

summary(aclrStability.rf)

```
````

Pain

````
```{r warning=FALSE}
if (!require("randomForest")) install.packages(randomForest)
library(randomForest)
# (1) Build Random Forest Trees
aclrPain.rf <- randomForest(pain_2yr ~ SEX + TYPE + cartilage+srgage+ inj_type
              + lat_rem + med_rem
````

~~~
              , data=train, mtry=4, importance=TRUE,na.action = na.omit)
# (2) Calculate the Yhat value
yhat <- predict(aclrPain.rf,test)
# (3) Calculate the means of errors (=error rate)
mean((yhat-test$pain_2yr)^2)
# (4) View the importance of each variable (Note:if the importance=TRUE is omitted,
the %IncMSE will not be shown in importance()
importance(aclrPain.rf)
# (5) Plot the importance
varImpPlot(aclrPain.rf)

plot(aclrPain.rf)

```
##### ---------------------------------------------------------------------------------
###  RandomForest (Package: caret)
##### (1) Build Random Forest Trees
#####  "caret" package will figure the best mtry
#####  (2) Calculate the Yhat value
##### (3) Calculate the means of errors (=error rate)
#####  (4) View the importance of each variable (Note:if the importance=TRUE is
omitted, the %IncMSE will not be shown in importance()
#####  (5) Plot the importance
##### (6) Use Package "doMC" to calculate final model (Best mtry)

```{r warning=FALSE}
if (!require("caret")) install.packages(caret)
library(caret)
# (1) Build Random Forest

rf.caret <- train(total_2yr ~ SEX + TYPE + cartilage+srgage+ injsrg_mth + inj_type
              + lat_rem + med_rem
            , train, method='rf',preProc = c('center','scale'),importance=TRUE,
na.action=na.omit)
# (2) Calculate the Yhat value
yhat.caret <- predict(rf.caret,test)
# (3) Calculate the means of errors (=error rate)
mean((yhat.caret-test$medv)^2)
# (4) View the importance of each variable (Note:if the importance=TRUE is omitted,
the %IncMSE will not be shown in importance()
importance(rf.caret$finalModel)
# (5) Plot the importance
varImpPlot(rf.caret$finalModel)
# (6) Use Package "doMC" to calculate final model (Best mtry)
if (!require("doMC")) install.packages("doMC", repos="http://R-Forge.R-project.org")
library(doMC)
~~~

~~~~
registerDoMC(6)
rf.caret <- train(total_2yr ~ SEX + TYPE + cartilage+srgage+ inj_type
                 + lat_rem + med_rem
                  , train, preProc = c('center','scale'),method='rf',importance=TRUE,
na.action=na.omit)
rf.caret
# (7)
library(randomForest)
rf.caret.best = randomForest(total_2yr ~ SEX + TYPE + BILAT +cartilage+srgage+
injsrg_mth + inj_type
                 + lat_rem + med_rem
            , data=train, mtry=2,importance=TRUE,na.action = na.omit)
rf.caret.best
# (2) Calculate the Yhat value
yhat.rf.caret.best <- predict(rf.caret.best,test)
# (3) Calculate the means of errors (=error rate)
mean((yhat.rf.caret.best-test$diffext_1mth)^2)
# (4) View the importance of each variable (Note:if the importance=TRUE is omitted,
the %IncMSE will not be shown in importance()
importance(rf.caret.best)
# (5) Plot the importance
varImpPlot(rf.caret.best)

plot(rf.caret.best)
summary(rf.caret.best)
```
##### ----------------------------------------------------------
##   BOOSTING
##### (1) For boosting, the gmb package and gbm() function is used to fit a boosted
regression tree.
#####    - distribution='gaussian' is used for regression tree, while
distribution='bernoulli' is used for binary classification
#####     - n.trees= . indicates the number of trees
#####     - interaction.depth limits the depth of each tree
##### (2) we can plot  'partial depedence' for the top variables, which illustrates the
marginal effect of the selected variables on the response after integrating out the other
variables.
#####     - Here, we can see that the median house prices increase with 'rm, and decrease
with 'lstat'
##### (3) Calculate the MSE
##### (4) Tune parameters (The best model with maxdepth and nstop)
##### (5) Re-calculate MSE from the new tunned boost model
##### (6) Summary the R-Squared Variable Importance

```{r warning=FALSE}
~~~~

```
if (!require("gbm")) install.paclages(gbm)
### (1) For boosting, the gmb package and gbm() function is used to fit a boosted
regression tree.
boost1mth <- gbm(total_2yr ~ SEX + TYPE + cartilage+srgage+ inj_type
                 + lat_rem + med_rem
             , data=train, distribution='gaussian',n.trees=5000,interaction.depth=4)
summary(boost1mth)

### (2) we can plot  'partial depedence' for the top variables, which illustrates the
marginal effect of the selected variables on the response after integrating out the other
variables.
par(mfrow=c(1,2)) ## to put the two plots below side-by-side
plot.gbm(boost1mth,i="srgage",n.trees = boost1mth$n.trees)
#?plot.gbm ## plot marginal effect

### (3) Calculate the MSE
boost1mth.pred <- predict(boost1mth,test,n.trees=500)
mean((boost1mth.pred - test$diffext_1mth)^2)

### (4) Tune parameters (The best model with maxdepth and nstop)
if (!require("e1071")) install.packages(e1071)
library(e1071)
if (!require("bst")) install.packages(bst)
library(bst)
if (!require("plyr")) install.packages(plyr)
library(plyr)

ctr <- trainControl(method = "cv", number=10)

boost1mth.caret <- train(total_2yr ~ SEX + TYPE + cartilage+srgage+ inj_type
                 + lat_rem + med_rem
                    , train, method='bstTree',preProc=c('center','scale'),trControl=ctr,
na.action=na.omit)
boost1mth.caret

### Plot
plot(boost1mth.caret)

### (5) Re-calculate MSE from the new tunned boost model
boost1mth.caret.pred <- predict(boost1mth.caret,test,n.trees=100)
mean((boost1mth.caret.pred - test$diffext_1mth)^2)

### (6) Summary the R-Squared Variable Importance
varImp(boost1mth.caret)
```

REFERENCES

Abacha, A. B., Da Silveira, M., & Pruski, C. (2013). *Medical ontology validation through question answering.* Paper presented at the Conference on Artificial Intelligence in Medicine in Europe.

Abasolo, J., & Gmez, M. (2000). MELISA. An ontology-based agent for information retrieval in medicine.

Adler-Milstein, J., Bates, D. W., & Jha, A. K. (2013). Operational health information exchanges show substantial growth, but long-term funding remains a concern. *Health Affairs, 32*(8), 1486-1492.

Allen, I. E., & Seaman, C. A. (2007). Likert scales and data analyses. *Quality progress, 40*(7), 64-65.

Álvez, J., Lucio, P., & Rigau, G. (2019). A framework for the evaluation of SUMO-based ontologies using WordNet. *IEEE Access, 7*, 36075-36093.

Anders, M. E., & Evans, D. P. (2010). Comparison of PubMed and Google Scholar literature searches. *Respiratory care, 55*(5), 578-583.

Anitha, B., & Rajagopalan, S. (2011). Computer decision support systems for evidence-based medicine–an overview. *European Journal of Scientific Research, 50*(3), 444-452.

Arp, R., & Smith, B. (2008). Function, role, and disposition in basic formal ontology. *Nature Precedings*, 1-1.

Arp, R., Smith, B., & Spear, A. D. (2015). *Building ontologies with basic formal ontology*: Mit Press.

B. Bolbel, K. E., P Pharow. (2006). Semantic Interoperability. *Methods Inf Med, 2006*(45), 343-353.

Bandrowski, A., Brinkman, R., Brochhausen, M., Brush, M. H., Bug, B., Chibucos, M. C., . . . Dumontier, M. (2016). The ontology for biomedical investigations. *PloS one, 11*(4).

Barkham, M., & Mellor-Clark, J. (2003). Bridging evidence-based practice and practice-based evidence: Developing a rigorous and relevant knowledge for the psychological therapies. *Clinical Psychology & Psychotherapy: An International Journal of Theory & Practice, 10*(6), 319-327.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific american, 284*(5), 34-43.

Biolchini, J. (2002). Developing a UMLS-based Ontology of Cardiology Procedures for Cognitive Support in Medical Decision Making. *Lister Hill Nat'l Ctr., Nat'l Lib. Med*.

Black, N. (2013). Patient reported outcome measures could help transform healthcare. *Bmj, 346*, f167.

Boon, H., MacPherson, H., Fleishman, S., Grimsgaard, S., Koithan, M., Norheim, A. J., & Walach, H. (2007). Evaluating complex healthcare systems: a critique of four approaches. *Evidence-Based Complementary and Alternative Medicine, 4*.

Bramer, W. M., Giustini, D., Kramer, B. M., & Anderson, P. (2013). The comparative recall of Google Scholar versus PubMed in identical searches for biomedical systematic reviews: a review of searches used in systematic reviews. *Systematic reviews, 2*(1), 115.

Brophy, R. H., Schmitz, L., Wright, R. W., Dunn, W. R., Parker, R. D., Andrish, J. T., . . . Spindler, K. P. (2012). Return to play and future ACL injury risk after ACL reconstruction in soccer athletes from the Multicenter Orthopaedic Outcomes Network (MOON) group. *The American journal of sports medicine, 40*(11), 2517-2522.

Brusa, G., Laura Caliusco, M., & Chiotti, O. (2008). Towards ontological engineering: a process for building a domain ontology from scratch in public administration. *Expert Systems, 25*(5), 484-503.

Calegari, S., & Ciucci, D. (2007). *Fuzzy ontology, fuzzy description logics and fuzzy-owl.* Paper presented at the International Workshop on Fuzzy Logic and Applications.

Cercone, N., An, X., Li, J., Gu, Z., & An, A. (2011). Finding best evidence for evidence-based best practice recommendations in health care: the initial decision support system design. *Knowledge and information systems, 29*(1), 159.

Ceusters, W. (2012). *An information artifact ontology perspective on data collections and associated representational artifacts.* Paper presented at the MIE.

Chalortham, N., Buranarach, M., & Supnithi, T. (2009). *Ontology development for type II diabetes mellitus clinical support system.* Paper presented at the Proceedings of 4th international conference on knowledge information and creativity support systems.

Chen, Z., Xue, C., Zhu, S., Zhou, F., Xuefeng Bruce, L., Liu, G., & Chen, L. (2005). GoPipe: Streamlined gene ontology annotation for batch anonymous sequences with statistics. *Sheng wu hua xue yu sheng wu wu li jin zhan, 32*(2), 187-191.

Chiarugi, F., Colantonio, S., Emmanouilidou, D., Moroni, D., Perticone, F., Sciacqua, A., & O., S. (2008). ECG and Echocardiography Processing for Decision Support in Heart Failure. *Computers in Cardiology*, 649-652.

Chibucos, M. C., Siegele, D. A., Hu, J. C., & Giglio, M. (2017). The evidence and conclusion ontology (ECO): supporting GO annotations. In *The Gene Ontology Handbook* (pp. 245-259): Humana Press, New York, NY.

Chung, K. C., Swanson, J. A., Schmitz, D., Sullivan, D., & Rohrich, R. J. (2009). Introducing evidence-based medicine to plastic and reconstructive surgery. *Plastic and reconstructive surgery, 123*(4), 1385.

Cole, C. (2004). Using a Terminology Server and Consumer Search Phrases to Help Patients Find Physicians with Particular Expertis. *MedInfo*.

Consortium, G. O. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research, 32*(suppl_1), D258-D261. doi:10.1093/nar/gkh036

Cornet, R. (2015). Toward a patient safety upper level ontology. *Digital Healthcare Empowering Europeans: Proceedings of MIE2015, 210*, 160.

D. Pappa, P. T., Lampros K Stergioulas. (2006). Management of Medicines Information for Patient Safety. *Journal of Telemedicine and Telecare, 12*, 34-36.

Da Costa, P. C. G., Laskey, K. B., & Laskey, K. J. (2006). PR-OWL: A Bayesian ontology language for the semantic web. In *Uncertainty Reasoning for the Semantic Web I* (pp. 88-107): Springer.

Dang, J., Hedayati, A., Hampel, K., & Toklu, C. (2008). An ontological knowledge framework for adaptive medical workflow. *Journal of Biomedical Informatics, 41*(5), 829-836.

Deber, R. B., Kraetschmer, N., & Irvine, J. (1996). What role do patients wish to play in treatment decision making? *Archives of internal medicine, 156*(13), 1414-1420.

Degen, W., Heller, B., Herre, H., & Smith, B. (2001). *GOL: toward an axiomatized upper-level ontology.* Paper presented at the Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001.

Ding, L., Finin, T., Joshi, A., Pan, R., Cost, R. S., Peng, Y., . . . Sachs, J. (2004). *Swoogle: a search and metadata engine for the semantic web.* Paper presented at the Proceedings of the thirteenth ACM international conference on Information and knowledge management.

Doyle, T. J., Ma, H., Groseclose, S. L., & Hopkins, R. S. (2005). PHSkb: A knowledgebase to support notifiable disease surveillance. *BMC Medical Informatics and Decision Making, 5*(1), 27.

Edwards, P., Ebert, J., Joss, B., Ackland, T., Annear, P., Buelow, J.-U., & Hewitt, B. (2018). Patient Characteristics and Predictors of Return to Sport at 12 Months After Anterior Cruciate Ligament Reconstruction: The Importance of Patient Age and Postoperative Rehabilitation. 2325967118797575. doi:10.1177/2325967118797575]

El-Sappagh, S., Franda, F., Ali, F., & Kwak, K.-S. (2018). SNOMED CT standard ontology based on the ontology for general medical science. *BMC Medical Informatics and Decision Making, 18*(1), 76. doi:10.1186/s12911-018-0651-5

Elkin, P., Brown, S. H., Bauer, B. A., Husser, C. S., Carruth, W., Bergstrom, L., & Wahner-Roedler, D. (2005). A Controlled Trial of Automated Classification of Negation from Clinical Notes. *BMC Medical Informatics and Decision Making 10.1186/1472-6947.*

Ellis, P. (2019). *Evidence-based practice in nursing*: Learning Matters.

Elmhadhbi, L., Karray, M.-H., & Archimède, B. (2019). Toward the Use of Upper-Level Ontologies for Semantically Interoperable Systems: An Emergency Management Use Case. In *Enterprise Interoperability VIII* (pp. 131-140): Springer.

Evans, J. M.-C., Frank Margison, Michael Barkham, Kerry Audin, Janice Connell, Graeme McGrath, Chris. (2000). CORE: clinical outcomes in routine evaluation. *Journal of Mental Health, 9*(3), 247-255.

Faber, P., Mairal, R., & Magaña, P. (2011). *Linking a domain-specific ontology to a general ontology.* Paper presented at the Twenty-Fourth International FLAIRS Conference.

Fadel, F. G., Fox, M. S., & Gruninger, M. (1994). *A generic enterprise resource ontology.* Paper presented at the Proceedings of 3rd IEEE Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises.

Farfan, F., Hristidis, V., Ranganathan, A., & Burke, R. P. (2008, 7-12 April 2008). *Ontology-Aware Search on XML-based Electronic Medical Records.* Paper presented at the Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on.

Farion, K., Michalowski, W., Wilk, S., O'Sullivan, D., Rubin, S., & Weiss, D. (2009). Clinical decision support system for point of care use. *Methods of information in medicine, 48*(04), 381-390.

Flouris, G., Plexousakis, D., & Antoniou, G. (2006). *Evolving ontology evolution.* Paper presented at the International Conference on Current Trends in Theory and Practice of Computer Science.

Forbes, D., Sidhu, A., & Singh, J. (2011). Coupling Indigenous Patient-Friendly Cultural Communications with Clinical Care Guidelines for Type 2 Diabetic Mellitus. *In Proc. Australasian Workshop on Healthmatics and Knowledge Management (HIKM 2011) CRPIT, 120*(ACS), 31-26.

Fox, M. S., Chionglo, J. F., & Fadel, F. G. (1993). *A common-sense model of the enterprise.* Paper presented at the Proceedings of the 2nd Industrial Engineering Research Conference.

Fried, R., Klas, W., & Westermann, W. (2003). The CadioOP-Data Clas (CDC). *Methods Inf Med, 2003*(42), 68-78.

Galton, A. (2018). *The Treatment of Time in Upper Ontologies.* Paper presented at the FOIS.

Gibson, F. (2010). Upper level ontologies. *Ontogenesis*.

González-Beltrán, A., Maguire, E., Sansone, S.-A., & Rocca-Serra, P. (2014). linkedISA: semantic representation of ISA-Tab experimental metadata. *BMC bioinformatics, 15*(14), S4.

Grenon, P., Smith, B., & Goldberg, L. (2004). Biodynamic ontology: applying BFO in the biomedical domain. *Studies in health technology and informatics*, 20-38.

Gruber, T. (2018). Ontology. In.

Grüninger, M., Atefi, K., & Fox, M. S. (2000). Ontologies to support process integration in enterprise engineering. *Computational & Mathematical Organization Theory, 6*(4), 381-394.

Guarino, N., Oberle, D., & Staab, S. (2009). What is an ontology? In *Handbook on ontologies* (pp. 1-17): Springer.

Haendel, M. A., Neuhaus, F., Osumi-Sutherland, D., Mabee, P. M., Mejino, J. L., Mungall, C. J., & Smith, B. (2008). CARO–the common anatomy reference ontology. In *Anatomy Ontologies for Bioinformatics* (pp. 327-349): Springer.

Hastings, J., Batchelor, C., Neuhaus, F., & Steinbeck, C. (2012). What's in anis about'link? Chemical diagrams and the Information Artifact Ontology. *arXiv preprint arXiv:1204.4805*.

He, Y., Xiang, Z., Todd, T., Courtot, M., Brinkman, R., Zheng, J., . . . Sansone, S.-A. (2010). Ontology representation and ANOVA analysis of vaccine protection investigation.

Hoehndorf, R. (2010). What is an upper level ontology? *Ontogenesis*.

Horn, S. D., & Gassaway, J. (2007). Practice-based evidence study design for comparative effectiveness research. *Medical care, 45*(10), S50-S57.

Howland, R. H. (2007). Limitations of evidence in the practice of evidence-based medicine. *Journal of psychosocial nursing and mental health services, 45*(11), 13-16.

Hudson, K. L., & Collins, F. S. (2015). Sharing and reporting the results of clinical trials. *Jama, 313*(4), 355-356.

Hwang, K. S., Park, K. S., Lee, S. H., Kim, K. I., & Lee, K. M. (2018, 5-8 Dec. 2018). *Autonomous Machine Learning Modeling using a Task Ontology.* Paper presented at the 2018 Joint 10th International Conference on Soft Computing and Intelligent

Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS).
Iroju, O., Soriyan, A., Gambo, I., & Olaleke, J. (2013). Interoperability in healthcare: benefits, challenges and resolutions. *International Journal of Innovation and Applied Studies, 3*(1), 262-270.
Ishizu, S., Gehrmann, A., Minegishi, J., & Nagai, Y. (2008). *Ontology-driven decision support systems for management system audit.* Paper presented at the Proceedings of the 52nd Annual Meeting of the ISSS-2008, Madison, Wisconsin.
J. Effken, B. C. (2002). The Era of Patient Safety: Implications for Nursing Informatics Curricula. *J AM Med Inform Assoc.* Retrieved from
Jarrar, M. (2005). Towards Methodological Principles for Ontology Engineering. *Mustafa Jarrar: Towards methodological principles for ontology engineering. Ph.D. Thesis. Vrije Universiteit Brussel.(May 2005)*.
Jenson, J. M., & Howard, M. O. (2013). Evidence-based practice. In *Encyclopedia of social work*.
Jill Brown, J. S., Jason McClaren. (2006). Waiting for a Liver Transplant. *Qulitative Health Research, 16*(1), 119-136.
Jones, J., Phalakornkule, K., Fitzpatrick, T., Iyer, S., & Ombac, C. Z. (2011). *Developing protégé to structure medical report.* Paper presented at the International Conference on Universal Access in Human-Computer Interaction.
Jovic, A., Prcela, M., & Gamberger, D. (2007). *Ontologies in medical knowledge representation.* Paper presented at the 2007 29th International Conference on Information Technology Interfaces.
Jovic, A., Prcela, M., & Gamberger, D. (2007, 25-28 June 2007). *Ontologies in Medical Knowledge Representation.* Paper presented at the Information Technology Interfaces, 2007. ITI 2007. 29th International Conference on.
K. Kim, H.-J. C. (2007). Design of a Clinical Knowledge Base for Heart Disease Detection. *IEEE Computer Society*, 610-615.
Kazdin, A. E. (2008). Evidence-based treatment and practice: new opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American psychologist, 63*(3), 146.
Kelley, E., Moy, E., Stryer, D., Burstin, H., & Clancy, C. (2005). The national healthcare quality and disparities reports: an overview. *Medical care, 43*(3), I-3-I-8.
Klischewski, R. (2003). *Top down or bottom up? How to Establish a Common Ground for Semantic Interoperability within e-Government Communities.* Paper presented at the Proceedings of 1st International Workshop on E-Government at ICAIL.
Kong, F., Liu, P., & Wang, X. (2011). Experimental investigation of evolution process of nonlinear characteristics from chatter free to chatter. *Journal of Modern Physics, 2011*.
L. Ahmadian, R. C., N. Keizer. (2009). Facilitating Pre-Operative Assessment Guidelines Representation using SNOMED CT. *Journal of Biomedical Informatics, 2010*(43), 883-890.
Labrinidis, A., & Jagadish, H. V. (2012). Challenges and opportunities with big data. *Proceedings of the VLDB Endowment, 5*(12), 2032-2033.
Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news, 2*(3), 18-22.

Liaw, A., & Wiener, M. (2018). Randomforest, R Package Version 4.6-14. *R Core Team, Vienna, Austria*.

Lin, Y., & He, Y. (2012). Ontology representation and analysis of vaccine formulation and administration and their effects on vaccine immune responses. *Journal of Biomedical Semantics, 3*(1), 17.

Loane, M. A., Bloomer, S., Corbett, R., Eedy, D., Hicks, N., Lotery, H., . . . Wootton, R. (2000). A comparison of real-time and store-and-forward teledermatology: a cost–benefit study. *British Journal of Dermatology, 143*(6), 1241-1247.

López-Pellicer, F., Vilches-Blázquez, L., Nogueras-Iso, J., Corcho, O., Bernabé, M., & Rodríguez, A. (2007). *Using a hybrid approach for the development of an ontology in the hydrographical domain.* Paper presented at the Proceedings of 2nd Workshop of COST Action C21-Towntology Ontologies for urban development: conceptual models for practitioners.

Lucock, M., Leach, C., Iveson, S., Lynch, K., Horsefield, C., & Hall, P. (2003a). A Systematic Approach to Practice-Based Evidence in a Psychological Therapies Service. *Clinical Psychology and Psychotherpy, 10*, 389-399. doi:10.1002/cpp.385

Lucock, M., Leach, C., Iveson, S., Lynch, K., Horsefield, C., & Hall, P. (2003b). A systematic approach to practice-based evidence in a psychological therapies service. *Clinical Psychology & Psychotherapy: An International Journal of Theory & Practice, 10*(6), 389-399.

Majid, S., Foo, S., Luyt, B., Zhang, X., Theng, Y.-L., Chang, Y.-K., & Mokhtar, I. A. (2011). Adopting evidence-based practice in clinical decision making: nurses' perceptions, knowledge, and barriers. *Journal of the Medical Library Association: JMLA, 99*(3), 229.

Maramis, C., Falelakis, M., Lekka, I., Diou, C., Mitkas, P., & Delopoulos, A. (2013). Applying semantic technologies in cervical cancer research. *Data & Knowledge Engineering, 86*, 160-178. doi:https://doi.org/10.1016/j.datak.2013.02.003

Margison, F. R., Barkham, M., Evans, C., McGrath, G., Clark, J. M., Audin, K., & Connell, J. (2000). Measurement and psychotherapy: Evidence-based practice and practice-based evidence. *The British Journal of Psychiatry, 177*(2), 123-130.

Mascardi, V., Cordì, V., & Rosso, P. (2007). *A Comparison of Upper Ontologies.* Paper presented at the Woa.

McDonald, P., & Viehbeck, S. (2007). From Evidence-Based Practice Making to PRactice-Based Evidence Making: Creating Cimmunities of (research) and Practice. *Health Promotion Practice, 8*(140). doi:10.1177/1524839906298494

McGrath, P. (2002). Creating a Language for 'Spiritual Pain' through Research: a Beginning. *Support Care Cancer, 2002*(10), 637-646.

Melnyk, B. M., Gallagher-Ford, L., Long, L. E., & Fineout-Overholt, E. (2014). The establishment of evidence-based practice competencies for practicing registered nurses and advanced practice nurses in real-world clinical settings: Proficiencies to improve healthcare quality, reliability, patient outcomes, and costs. *Worldviews on Evidence-Based Nursing, 11*(1), 5-15.

Mitrović, J., Pease, A., & Granitzer, M. Modeling Legal Terminology in SUMO.

Mohammadi, M., & Mukhtar, M. (2013). A review of SOA modeling approaches for enterprise information systems. *Procedia Technology, 11*(0), 794-800.

Monachesi, P., Cristea, D., Evans, D., Killing, A., Lemnitzer, L., Simov, K., & Vertan, C. (2006). *Integrating language technology and semantic web techniques in elearning.* Paper presented at the Proceedings of ICL.

Montani, S., Bellazzi, R., Riva, A., Larizza, C., Portinale, L., & Stefanclli, M. Artificial Intelligence Techniques for Diabetes Managements: the T-IDDM Project.

Morris, Z. S., Wooding, S., & Grant, J. (2011). The answer is 17 years, what is the question: understanding time lags in translational research. *Journal of the Royal Society of Medicine, 104*(12), 510-520.

Moss, L., Sleeman, D., Sim, M., Booth, M., Daniel, M., Donaldson, L., . . . Kinsella, J. (2010). Ontology-driven hypothesis generation to explain anomalous patient responses to treatment. In *Research and Development in Intelligent Systems XXVI* (pp. 63-76): Springer.

Mungall, C. J., Torniai, C., Gkoutos, G. V., Lewis, S. E., & Haendel, M. A. (2012). Uberon, an integrative multi-species anatomy ontology. *Genome biology, 13*(1), R5.

Murdoch, T. B., & Detsky, A. S. (2013). The inevitable application of big data to health care. *Jama, 309*(13), 1351-1352.

Needleman, J., Buerhaus, P., Pankratz, V. S., Leibson, C. L., Stevens, S. R., & Harris, M. (2011). Nurse staffing and inpatient hospital mortality. *New England Journal of Medicine, 364*(11), 1037-1045.

Noy, N. F., & McGuinness, D. L. 1 Why develop an ontology?

Noy, N. F., & McGuinness, D. L. (2001). Ontology development 101: A guide to creating your first ontology. In: Stanford knowledge systems laboratory technical report KSL-01-05 and ….

O'Neill, E. S., Dluhy, N. M., Fortier, P. J., & Michel, H. E. (2004). Knowledge acquisition, synthesis, and validation: a model for decision support systems. *Journal of advanced nursing, 47*(2), 134-142.

O. Nee, A. H. (2010). Clinical Decision Support with Guidelines and Bayesian Networks. *INTECH*, 342.

Obrst, L., Grüninger, M., Baclawski, K., Bennett, M., Brickley, D., Berg-Cross, G., . . . Kutz, O. (2014). Semantic web and big data meets applied ontology.

Ocaña-Riola, R. (2016). The Use of Statistics in Health Sciences: Situation Analysis and Perspective. *Statistics in Biosciences, 8*, 204–219. doi:10.1007/s12561-015-9138-4

Oman, K. S., Duran, C., & Fink, R. (2008). Evidence-based policy and procedures: an algorithm for success. *JONA: The Journal of Nursing Administration, 38*(1), 47-51.

Ong, E., Xiang, Z., Zhao, B., Liu, Y., Lin, Y., Zheng, J., . . . He, Y. (2017). Ontobee: a linked ontology data server to support ontology term dereferencing, linkage, query and integration. *Nucleic Acids Research, 45*(D1), D347-D352.

Overton, J. A., Romagnoli, C., & Chhem, R. (2011). Open Biomedical Ontologies applied to prostate cancer. *Applied Ontology, 6*, 35-51. doi:10.3233/AO-2010-0081

P. Jannin, X. M. (2007). Surgical Models for Computer-Assisted Neurosurgery. *NeuroImage, 37*, 783-791. Retrieved from

Panzarasa, S., Madde, S., Quaglini, S., Pistarini, C., & Stefanelli, M. (2002). Evidence-based Careflow Management SystemsL the Case of Post-Stroke Rehabilitation. *Journal of Biomedical Informatics, 2002*(35), 123-139.

Park, K., Kim, S., Lee, B., Park, J., Shim, B., Choi, S., & Rodrigues, J. (2009). Categorical Structure for a Terminology System in Traditional Medicine Symptoms, Signs, and their Combination Patterns. *WHO-FIC 2009, D033*.

Patel, M. R., Vichich, J., Lang, I., Lin, J., & Zheng, K. (2017). Developing an evidence base of best practices for integrating computerized systems into the exam room: a systematic review. *Journal of the American Medical Informatics Association, 24*(e1), e207-e215.

Patrick, T. B., Demiris, G., Folk, L. C., Moxley, D. E., Mitchell, J. A., & Tao, D. (2004). Evidence-based retrieval in evidence-based medicine. *Journal of the Medical Library Association, 92*(2), 196.

Pease, A., & Niles, I. (2002). IEEE standard upper ontology: a progress report. *The Knowledge Engineering Review, 17*(1), 65-70.

Peleg M., K. S., Denekamp Y. (2008). Mapping computerized clinical guidelines to electronic medical records: knowledge-data ontological mapper (KDOM). *J Biomed Inform, 41*(1), 180-201.

Pellegrin, L., Bonnardel, N., Antonini, F., Albanese, J., Martin, C., & Chaudel, H. (2007). Event Oriented Representation for Collaboratve Activites. *Methods Inf Med, 46*, 506-515.

Pellegrin, L., Bonnardel, N., Antonini, F., Albanese, J., Martin, C., & Chaudet, H. (2007). Event Oriented représentation for collaborative activities (EORCA). *Methods of information in medicine, 46*(05), 506-515.

Phalakornkule, K., Jones, J. F., & Finnell, J. T. (2013). *Ontological model for CDSS in knee injury management.* Paper presented at the International Conference on Universal Access in Human-Computer Interaction.

Phillimore, J., & Goodson, L. (2004). Progress in qualitative research in tourism: Epistemology, ontology and methodology. In *Qualitative research in tourism* (pp. 21-23): Routledge.

Pisanelli, D. M., Gangemi, A., & Steve, G. (1998). *An ontological analysis of the UMLS Metathesaurus.* Paper presented at the Proceedings of the AMIA symposium.

Plichta, S. B., & Garzon, L. S. (2009). *Statistics for nursing and allied health*: Lippincott Williams & Wilkins.

Prakash M Nadkarni, J. A. D. (2010). Migrating Existing Clinical Content from ICD-9 to SNOMED. *J AM Med Inform Assoc., 2010*(17), 602-607.

Purcell, G. P. (2005). What makes a good clinical decision support system. In: British Medical Journal Publishing Group.

Ramakrishnan, S., & Vijayan, A. (2014). A study on development of cognitive support features in recent ontology visualization tools. *Artificial Intelligence Review, 41*(4), 595-623.

Rasmussen, K. E., & Dolan, M. E. (2013). *OncoCL: A Cancer Cell Ontology.* Paper presented at the ICBO.

Riaño, D., Real, F., López-Vallverdú, J. A., Campana, F., Ercolani, S., Mecocci, P., . . . Caltagirone, C. (2012). An ontology-based personalization of health-care knowledge to support clinical decisions for chronically ill patients. *Journal of Biomedical Informatics, 45*(3), 429-446.

Rice, D. P. (1977). The role of statistics in the development of health care policy. *The American Statistician, 31*(3), 101-106.

Robinson, J. (1997). Evidence-based medicine how to practice & teach EBM: David L. Sackett, W. Scott Richardson, William Rosenberg R. Brian Haynes Churchill Livingstone 1997. *Australian College of Midwives Incorporated journal, 10*(2), 5.

Romero-Tris, C., Riaño, D., & Real, F. (2010). *Ontology-based retrospective and prospective diagnosis and medical knowledge personalization.* Paper presented at the International Workshop on Knowledge Representation for Health Care.

Rousseau, D. M., & Gunia, B. C. (2016). Evidence-based practice: The psychology of EBP implementation. *Annual Review of Psychology, 67*, 667-692.

Russell, L. B., Gold, M. R., Siegel, J. E., Daniels, N., & Weinstein, M. C. (1996). The role of cost-effectiveness analysis in health and medicine. *Jama, 276*(14), 1172-1177.

S Colantonio, D. C., M Martinelli, D Moroni, F Perticone, O Salvetti, A Sciacqua. (2008). An Intelligent and Integrated Paltform for Supporting the Management of Chronic Heart Failure Patients. *Computers in Cardiology, 35*, 897-900.

S. Abidi, S. S. R. A., Lorna Butler, Sajjad Hussain. (2009). Operationalizing Prostate Cancer Clinical Pathways: An Ontological Model to Computerize, Merge and Execute Institue-Specific Clinical Pathways. *Springer-Verlag Berlin Heidelberg, 2009*, 1-12.

S. Doan, A. K., Mike Conway, Nigel Collier. (2009). Towards Role-Based Filtering of Disease Outbreak Reports. *Journal of Biomedical Informatics, 2009*(42), 773-780.

S. Meghani, A. H. (2007). The Meanings of and Attitudes about Cancer Pain Among African Americans. *Oncology Nursing Forum, 34*(6), 1179-1186.

S. Mersmann, M. D. (2004). SmartCare - Automated Clinical Guidelines in Critical Care. *ECAI*.

Sackett, D. L., Rosenberg, W. M., Gray, J. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't. In: British Medical Journal Publishing Group.

Sermeus, W. (2016). Transforming clinical documentation in EHRs for 2020: Recommendations from University of Minnesota’s big data conference working group. *Nursing Informatics 2016*, 18.

Shaw, M., Detwiler, L. T., Brinkley, J. F., & Suciu, D. (2008). *Generating application ontologies from reference ontologies.* Paper presented at the AMIA Annual Symposium Proceedings.

Sim, I., Gorman, P., Greenes, R., Haynes, B., Kaplan, B., Lehman, H., & Tang, P. (2001). Clinical Decision Support Systems for the Practice of Evidence-based Medicine. *Journal of the American Medical Informatics Association, 8*(6), 527-534.

Sistrom, L. (2005). The ACR Appropriateness Criteria: Translation to Practice and Research. *Americal College of Radiology*.

Slim, K. (2005). Limits of evidence-based surgery. *World journal of surgery, 29*(5), 606-609.

Smith, B., Almeida, M., Bona, J., Brochhausen, M., Ceusters, W., Courtot, M., . . . Hastings, J. (2015). Basic formal ontology 2.0–specification and user’s guide

(2015). *URL: https://github. com/BFO-ontology/BFO/raw/master/docs/bfo2-reference/BFO2-Reference. pdf, 755*.
Smith, B., Almeida, M., Bona, J., Brochhausen, M., Ceusters, W., Courtot, M., . . . Hastings, J. (2015). Basic Formal Ontology 2.0: Specification and user's guide. *Online: https://github. com/BFO-ontology/BFO*.
Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., . . . Mungall, C. J. (2007). The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology, 25*(11), 1251-1255.
Smith, B., Malyuta, T., Rudnicki, R., Mandrick, W., Salmen, D., Morosoff, P., . . . Parent, K. (2013). IAO-Intel: an ontology of information artifacts in the intelligence domain.
Smucny, M., Westermann, R. W., Group, M. K., & Group, M. (2018). The MOON and Marsteam based Research: Formation, Challenges, and Success. *Sports Medicine and Rehabilitation Journal, 3(2) 1033*.
Sowa, J. F. (2000). *Knowledge representation: logical, philosophical, and computational foundations* (Vol. 13): MIT Press.
Stiwne, D., & Abrandt Dahlgren, M. (2004). Challenging evidence in evidence-based practice.
Strasak, A. M., Zaman, Q., Pfeiffer, K. P., Göbel, G., & Ulmer, H. (2007). Statistical errors in medical research-a review of common pitfalls. *Swiss medical weekly, 137*(0304).
Tiscornia, D. (2006). *The LOIS project: Lexical ontologies for legal information sharing.* Paper presented at the Proceedings of the V Legislative XML Workshop.
Titler, M. G. (2008). The Evidence for Evidence-Based Practice Implementation. Retrieved from http://www.ahrq/gov/qual/nurseshdbk
Tonelli, M. (1998). The philosophical limits of evidence-based medicine.
Uschold, M., & Gruninger, M. (1996). Ontologies: Principles, methods and applications. *The knowledge engineering review, 11*(02), 93-136.
V. Mago, M. S. D. (2007). A Multi-Agent Medical System for Indian Rural Infant and Child Care. *IJCAI*.
Van Der Straeten, R., Mens, T., Simmonds, J., & Jonckers, V. (2003). *Using description logic to maintain consistency between UML models.* Paper presented at the International Conference on the Unified Modeling Language.
Van Der Vet, P. E., & Mars, N. J. (1998). Bottom-up construction of ontologies. *IEEE Transactions on Knowledge and data Engineering, 10*(4), 513-526.
Waite, R., & Killian, P. (2016). Evidence-based practice. *Translation of Evidence into Nursing and Health Care. 2nd ed. New York: Springer Publishing Company*.
Wang, S., & Noe, R. A. (2010). Knowledge sharing: A review and directions for future research. *Human resource management review, 20*(2), 115-131.
Wilkinson, E. K., McColl, A., Exworthy, M., Roderick, P., Smith, H., Moore, M., & Gabbay, J. (2000). Reactions to the use of evidence-based performance indicators in primary care: a qualitative study. *BMJ Quality & Safety, 9*(3), 166-174.
Wright, J. G., Swiontkowski, M. F., & Heckman, J. D. (2003). Introducing levels of evidence to the journal. In: LWW.

Zhao, J., Kan, M.-Y., Procter, P. M., Zubaidah, S., Yip, W. K., & Li, G. M. (2010). *Improving search for evidence-based practice using information extraction.* Paper presented at the AMIA Annual Symposium Proceedings.

Zheng, J., Harris, M. R., Masci, A. M., Lin, Y., Hero, A., Smith, B., & He, Y. (2014). OBCS: The Ontology of Biological and Clinical Statistics.

Zheng, J., Harris, M. R., Masci, A. M., Yu, L., Hero, A., Smith, B., & He, Y. (2016). The Ontology of Biological and Clinical Statistics (OBCS) for standardized and reproducible statistical analysis. *Journal of Biomedical Semantics, 7*(53).

Zheng, J., Li, H., Liu, Q., & He, Y. (2017). The Ontology of Biological and Clinical Statistics (OBCS)-based statistical method standardization and meta-analysis of host responses to yellow fever vaccines. *Quantitative Biology, 5*(4), 291-301.

Zillner, S., Oberkampf, H., Bretschneider, C., Zaveri, A., Faix, W., & Neururer, S. (2014). *Towards a technology roadmap for big data applications in the healthcare domain.* Paper presented at the Proceedings of the 2014 IEEE 15th International Conference on Information Reuse and Integration (IEEE IRI 2014).

Zúñiga, G. L. (2001). *Ontology: its transformation from philosophy to information systems.* Paper presented at the Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001.

CURRICULUM VITAE

Kanitha Phalakornkule

Education

Doctor of Philosophy	Fall 2006 – Fall 2020
Informatics and Computing	
Indiana University, Indianapolis, IN	
Master of Science	Fall 1997 – Fall 1999
Agricultural Economics	
Purdue University, West Lafayette, IN	
Master of Science	Fall 1995 – Spring 1997
Management	
Purdue University, West Lafayette, IN	
Bachelor of Business Administration	Fall 1989 – Fall 1992
Marketing	
Kasetsart University, Bangkok, Thailand	

Certification

Association of Clinical Research Professionals (ACRP) Spring 2017

Publications

1. Phalakornkule, K., Duncan, W., Jones, J.. (2020). ACLR-Rehabilitation Domain-Specific Ontological Model for Evidence-Based Practice.
2. Phalakornkule, K., Jones, J., Duncan W. (2020). Formal Ontology for Evidence-Based Practice in ACLR Knee Rehabilitation.
3. Phalakornkule, K., Jones, J., Duncan, W.,(2020). Development of a Patient-Focused Ontology for Knee Rehabilitation with Top-Down Approach.

4. Sawesi, S., Rashrash, M., Phalakornkule, K., Carpenter, J.S., & Jones, J.F. (2016). The Impact of Information Technology on Patient Engagement and Health Behavior Change: A Systematic Review of the Literature. *JMIR Medical Informatics*, *4*(1), e1.

Conference, Presentations, and Posters

1 Meta-Analysis of Ontology Applications in Healthcare. *Kanitha Phalakornkule, Suhila A. Sawesi, Josette F. Jones. AMIA 2015 annual symposium.*

2 Meta-analysis: Impact of Health Information Technology on Patient Engagement and Health Behavior Change. *Suhila A. Sawesi, Kanitha Phalakornkule, Josette F. Jones. AMIA 2015 annual symposium.*

3 Ontological model for CDSS in knee injury management. *Phalakornkule, K., Jones, J. F., & Finnell, J. T.* (2013, July). In *International Conference on Universal Access in Human-Computer Interaction (pp. 526-535). Springer, Berlin, Heidelberg.*

4 Developing protégé to structure medical report. *Jones, J., Phalakornkule, K., Fitzpatrick, T., Iyer, S., & Ombac, C. Z. (2011, July). In International Conference on Universal Access in Human-Computer Interaction (pp. 356-365). Springer, Berlin, Heidelberg.*

5 Leveraging user search behavior to design personalized browsing interfaces for healthcare web sites. *Mahoui, M., Jones, J. F., Zollinger, D., & Andersen, K. (2009, July). In International Conference on Human Centered Design (pp. 511-520). Springer, Berlin, Heidelberg.*