Thesis for the Degree of Doctor of Philosophy

Efficient Semantic Reconciliation for Data Interoperability among Heterogeneous Health-care Systems

Wajahat Ali Khan

Department of Computer Engineering
Graduate School
Kyung Hee University
South Korea

February 2015

EFFICIENT SEMANTIC RECONCILIATION FOR DATA INTEROPERABILITY AMONG HETEROGENEOUS HEALTH-CARE SYSTEMS

Wajahat Ali Khan

Department of Computer Engineering
Graduate School
Kyung Hee University
South Korea

February 2015

EFFICIENT SEMANTIC RECONCILIATION FOR DATA INTEROPERABILITY AMONG HETEROGENEOUS HEALTH-CARE SYSTEMS

by

Wajahat Ali Khan

Supervised by

Prof. Sungyoung Lee

Submitted to the Department of Computer Engineering and the Faculty of Graduate School of Kyung Hee University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

Dissertation Committee:

Prof. Young-Koo Lee ____Young-Koo Lee____

Prof. Eui-Nam Huh ________

Prof. Kyungmo Park ________

Prof. Jinsung Cho ________

Prof. Sungyoung Lee ________

Dedicated to the memory of my father

Liaqat Ali Khan

During all the rough and tough stages of my PhD timespan, my father kept on encouraging me with his words of wisdom. Although he is not with me, but his teachings will always guide me in difficult situations I will face ahead.

This thesis is dedicated to my father who had always stood with me in my sorrows and achievements. He guided me all the way through in my educational career and finally met his demise on 11th March, 2014. I will always remember the affection, enthusiasm, and encouragement that he entrusted on me that gave me strength to face all the difficulties and challenges during my course of research. May Allah rest his soul in peace.

Abstract

Semantic interoperability among health-care systems complaint to heterogeneous standards is bottleneck for the vision of integrated Electronic Health Records (EHR). Health-care standards follow reference models that contain similarities and differences among them. Semantic reconciliation using ontology mappings provides the platform for utilizing the similarities and exposing the differences among health-care standards. This reconciliation process becomes more effective when organizational conformance issues are also incorporated in the mappings process. An effective and expressive mapping representation and storage technique is required to accommodate the mappings generated using ontology matching techniques and organizational conformance incorporation technique. These mappings also require a mechanism for transformation among different health-care standards heterogenous formats. This leads to a set of challenges required to be resolved for achieving data interoperability among health-care systems. The challenges for integrated EHR among the systems compliant to heterogeneous standards include: i) using ontology matching techniques to generate accurate mappings between health-care standards ontologies; ii) organization's conformance issues handling in the generated mappings; iii) a mechanism for efficient mappings representation and storage for effective retrieval; and iv) finally transformation of the source standard format to target standard format. These challenges require a comprehensive framework that provides gateway for true data interoperability using matching tools, personalized mappings approach, effective mapping representation techniques, and accurate standard format transformation schemes.

To address these challenges, this research introduces *Health-care Semantic Reconciliation Framework* with the main aim to achieve semantic data interoperability among health-care systems compliant to heterogeneous standards. The proposed framework resolves the challenges

described by using different approaches. The mappings between health-care standards ontologies is generated using *System for Parallel Heterogeneity Resolution* (SPHeRe) system and the generated mappings are categorized as generalized mappings. The *SPHeRe* system has deficiency of not handling the organizational conformance issues which are resolved by *Personalized-Detailed Clinical Model (P-DCM)* approach. Mappings evolution process mediates between the two categories of mappings by removing the stale mappings. The customized mappings compliments the generalized mappings for higher accuracy of mappings stored in mapping repository. The mappings from SPHeRe and P-DCM are stored in the mapping repository called the *Mediation Bridge Ontology* (MBO) which is effective and expressive ontology mappings representation and storage repository. The MBO is based on object oriented and ontology alignment design patterns that provides extendibility and reusability aspects to the system. The mappings are represented into logic form to be utilized for transformation purpose between heterogeneous health-care standards. Finally, the validity of the accuracy of mappings stored in the MBO is provided by the mappings execution that transforms one standard format to another. The level of accuracy of transformation defines the level of data interoperability achieved between different health-care systems.

The proposed approach produces high level of precision among biomedical ontologies matching evaluated by participating in the OAEI campaign. The MBO design factors are analyzed and compared with existing FALCON and Logmap systems. The MBO design factors show its design agility compared to the existing systems. The transformation is performed between HL7 CDA and vMR standards that reflected higher level of accuracy when SPHeRe and P-DCM approaches mappings are used together. This research provides mechanism for achieving true data interoperability among health-care systems compliant to heterogeneous health-care standards. It also contributes in achieving the goal of integrated EHR that can communicate information among the health-care systems easily reducing the health-care costs associated with the physicians, patients, pharmacies, and insurance companies.

Acknowledgement

First and foremost, I render my humble and sincere thanks to the ***Almighty Allah*** for showering HIS blessings upon me. The Almighty gave me the strength, courage, and patience during my doctoral study.

Special thanks to my advisor Prof. Sungyoung Lee who provided me guidance, strength, support, and courage in overcoming the difficult challenges throughout my time as his student. I have learned a lot from him in becoming a productive person in diverse situations. He inspired me with his dynamic personality and his unreserved help and guidance lead me to finish my thesis. He has great role in polishing my skills such as thinking, creativity and technical soundness which are key ingredients for high quality research.

I am grateful to my dissertation evaluation committee for their insight comments and valuable suggestions during the dissertation defense. Their valuable comments improved the presentation and contents of this dissertation.

I am extremely grateful to some of my colleagues who have always provided me time, expertise, and encouragement in my course of research. They were always present to guide me in difficult situations of my PhD duration. I would like to thank Dr Asad Masood Khattak, Dr Zeeshan Pervaiz, Dr Muhammad Fahim, Dr Adil Khan, Maqbool Hussain, Bilal Amin, Taqdir Ali, Shujaat Hussain, Rabia Batool, and Hameed Siddiqi. They have contributed enormously in successfully performing various academic and personal tasks that confronted me during my stay at South Korea.

I am very thankful to all of my current and former Ubiquitous computing lab fellows and colleagues for their kind support to my personal and academic life at Kyung Hee University. I am highly obliged to brilliant researchers Rahman Ali, Muhammad Afzal, Jamil Hussain, Idris, Dr

Weiwei Yuan, Byunggill Go, Tae Ho Hur, Jae Hun Bang, Muhammad Shoaib Siddiqui, Zulkarnain Rashid, Kifayat Ullah, Waqas Nawaz, Aamir Saleem, Omar Farooq, Mahmood Ahmad, Saeed Ullah, Ejaz Ahmed, Abdul Sattar, Ahsan Raza Kazmi, Zain, Kamran, Maqbool Ali, Noman, Shehryar and Bilal Ali. This journey would not have been possible without their support. They contributed a lot in to my personal and academic life to polish myself. Also, I appreciate all my Korean and international friends who worked as a team with me and developed my team work skills and provided me wonderful memories during my stay in South Korea.

Last but not the least, I would like to express my sincere gratitude to my mother, sisters, and brothers for their endless love, support, prayers, and encouragement. Their support and encouragement has made this dissertation possible. They provided me the strength to work hard after death of my father. I would like to thank my uncles for their support and guidance in every difficult situation I faced during all this time. Finally, I would like to leave the remaining space in memory of my father for his encouragement in all of my pursuits and inspiration to follow my dreams.

Wajahat Ali Khan

September, 2014

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Overview

Health-care industry envisions personalized treatment for individuals with integrated Electronic Health Record (EHR) infrastructure for timely medical care provisioning resulting in reduced health-care cost. This requires determining the optimal way to organize and present the patient record which is the most widely research area with focus on graphical representation of summary of the patient's information [3]. The biggest challenge for this is the integration of multiple heterogeneous and autonomous systems which is a complicated and resource consuming task, because of differences between service interfaces, business processes, data formats and underlying technologies [4]. The eventual cause is semantic heterogeneity in the underlying sources with respect to the data and storage structure; therefore, requiring data interoperability mechanism for data sharing among heterogeneous sources [5]. Data interoperability can be achieved by exploiting existing health-care standards and profiles [6] and adopting consistent clinical messaging and data standards for communicating shared meaning [7]. E-Health standardization is key to achieving data interoperability but diversity in these standards hampers the progress for integrated EHR due to huge number of available standards, with many of them competing and overlapping, and some even contradicting one another [8–10]. The various dimensions of health-care standards use increases the complexity in achieving the semantic data interoperability. Tools, technologies, approaches, methodologies are necessary for fulfilling the desire of interconnected remote medical systems infrastructure for easy availability of medical records saving the overly burden of health-care cost on individuals, patients, organizations and governments. Ontology matching is one of such approaches to resolve semantic heterogeneity problem among heterogeneous health-care standards.

Ontologies encode meaning to the information and ontology matching uses it in reducing the semantic gap between overlapping representations of same domain [11]. The tasks of ontology matching includes ontology merging, query answering, or data translation by finding correspondences between semantically related entities [12]. The main task of ontology matching is resolving heterogeneity that is divided into types such as syntactic (two ontologies not expressed in same ontology language), terminological (variation in names of entity in different ontologies), conceptual (differences in modeling the same domain of interest), and semiotic heterogeneity (entities interpretation by people) [13]. These heterogeneities are resolved by using ontology matching techniques that are classified into element level and structure level techniques. The element level techniques consider entities in isolation from relations with other entities and is divided into string based, language based, constraint based, formal resource based, and informal resource based technique categories [14]. Whereas structure level techniques covers entities with their instances, and children and these types of techniques are divided into graph based, taxonomy based, model based, and instance based technique categories [14]. The significance of these techniques in health-care domain can be observed from a special track in Ontology Alignment Evaluation Initiative (OAEI)[1] (OAEI, an annual campaign for systematic evaluation of ontology matching systems) called Large Biomed track involving matching FMA, SNOMED and NCI ontologies [15]. Many medical projects with the objective of interoperability are designed and developed in recent years by utilizing ontology matching techniques. These faced problems because of their approach directed to only providing solutions based on specic standards and technologies in order to satisfy the needs of a particular scenario [16]. The mapping information used for interoperability between different health-care standards or diverse medical ontologies of different organizations tries to resolve semantic heterogeneities at different levels, we categorizes as generalized mappings. The semiotic heterogeneity is not catered by most of these systems as it is a very challenging task, categorized as customized mappings. These mappings are organization specific and therefore play important role in data level interoperability.

An ongoing international collaboration effort is working on improving interoperability among medical systems through shared implementable clinical information models, called the Clinical

[1] http://oaei.ontologymatching.org/

Information Modeling Initiative (CIMI) [17]. The main purpose of this effort is to develop Detailed Clinical Models (DCM) for creation of semantically interoperable information to be shared in Electronic Health Records (EHR). The data exchange in EHR requires common clinical knowledge modeling to prevent loss of information, which is the goal of CIMI [18]. Existing approaches use clinical standards for interoperability in modeling clinical knowledge, but their exists difference in approach of conceptual, logical, and technical expressions incorporation [18]. One of the approaches for DCM's was Clinical Element Model that was based on models to define structure for representing instances (Abstract Instance Model), and model to define constraints on values of Abstract Instance Model (Abstract Constraint Model) [19]. The CEM model is represented in Clinical Element Modeling Language (CEML), but due to lack of expressing semantics and formal definitions of structure and semantics, CEM meta-ontology is developed to cover basic structures, properties and their relationships and constraints for expressing semantics [20]. This shows the importance of representing DCM in ontological form and then relating it with the standards mappings discussed previously as generalized mappings. The challenge for both type of mappings is in representations in the ontological format. Therefore, a bridge ontology that stores both standards based mappings as well as clinical model based mappings and its relationship with standards is required for handling semantic expressiveness. Ontological design patterns can play vital role in these semantic expressiveness for finding and storing these mappings in a centralized repository.

The generalized and conformance based mappings highlighted previously requires an expressive storage mechanism that can be used for transformation among medical records. A requirement of a mediation ontology that can do mapping reconciliations with expressive mappings storage and utilization scheme is solution for providing such type of data interoperability among medical systems compliant to heterogeneous health-care standards. Continuity of mappings provides the mappings reconciliations aspects for data interoperability. The ontology design patterns just like software design patterns decrease the incidence of poor modelling, makes ontology more readable, more maintainable and more reusable [21]. This mediation ontology drives the overall resolution for semantic heterogneity and provision of semantic data interoperability.

1.1 Motivation

Medical data heterogeneity, complexity and its continuously increasing size has acquired the attention of many research fields. Medical systems face semantic heterogeneity problem for exchange of data because health-care data is typically stored in heterogeneous and often autonomous IT systems [22]. Health-care standards for data exchange helps in reducing the semantic heterogeneity to a certain level [22]. Two organizations are interoperable if they are compliant with the same standard. The problem occurs when communicating parties are compliant to different health-care standards and want to communicate with each other. An openEHR[2] compliant system cannot directly communicate with an HL7 compliant system. These standards have information overlap and at the same time differences. The information overlap is good enough to bring the standards close to each other and make them interoperable. Ontology matching is one of the methods which can effectively utilize the data overlap for achieving semantic data interoperability among different health-care standards.

The increase in heterogenous medical data sources demands mediation for exchange of information between Health Information System (HIS). These HISs are compliant to conflicting standards which result in lack of communication among them due to gaps in medical standards [23]. So the bottleneck for these HISs is their compliancy with different health-care standards. The solution to handle complex medical data exchange is achieving interoperability. Semantic interoperability can only be achieved when existing standards are harmonized and bridged [24]. Therefore, mediation systems can play critical role in interoperability among HISs compliant to heterogeneous HISs.

Mediation systems provide common platform that understands the sender and receiver's HIS compliancy with heterogeneous standards and interprets information accordingly. This can be observed from the Clinical Information Modeling Initiative (CIMI) international collaboration initiative taken by health-care standardization bodies with the motive "dedicated to provide a common format for detailed specifications for the representation of health information content so that semantically interoperable information may be created and shared in health records, messages and documents" [25]. The dependency of these mediation systems is on the mappings to be used

[2] http://www.openehr.org/

for transformation. To quantify the efficiency of a mediation system, accuracy and evolution of mappings tend to be the two substantial factors for achieving data interoperability. To measure the effectiveness of accuracy and continuity of the mappings, a mediation system must have a transformation process for conversion among different standard formats.

Approaches

The literature review related to semantic interoperability based on semantic reconciliations can be divided into three main categories considering health-care domain. Firstly, generic ontology matching techniques and approaches that are used for matching the different sizes of ontologies. These ontology matching systems mainly focus on resolving semantic heterogeneity problem by matching entities to determine an alignment, and using data translation schemes for interpreting an alignment according to application needs [12]. Secondly, medical ontologies matching systems that only focuses on the matching medical ontologies for semantic interoperability. Their main focus is on the integration of biomedical information by matching large scale biomedical ontologies [26]. Lastly, data translation mechanisms that can use different mapping schemes for converting different data formats for interoperability. The main objective is to improve the exchange of meaningful clinical information among different medical systems [27]. Although these categories for achieving semantic interoperability have their impact on the progress of more and more expert systems development, yet some metrics should be defined for analyzing comprehensive semantic interoperability framework.

The factors that influence the comprehensiveness of semantic interoperability framework are Mappings Scheme, Data Translation, Mapping Representation, Flexibility in Mapping Representation, Accuracy of Mappings, and Continuity of Mappings. Mapping schemes are the techniques that are used for matching different ontologies, while data translation is the use of the mappings stored in the repository for conversion between different structural formats. Mapping representation is related to the way generated mappings are stored in the mappings repository, and flexibility in mappings representation is the ability to incorporate changes in the mapping schemes and mappings structures easily. Accuracy of mappings influences the extent of success of data translation and data integration, while continuity of mappings compliments the accuracy of mappings by reflecting the changes in the mapping repository. The categories for semantic interoperability as well as the factors that are required for comprehensiveness of the framework need to be analyzed with existing systems in the literature.

The categories of systems that are based on ontology matching techniques are selected based on their participation in the OAEI competition, and also their availability of publications and code for use. The systems that covers these criteria includes GOMMA [28], LogMap [29], Agreement Maker and Agreement Maker Light [30], and FALCON [31]. These systems focus on mapping schemes and mapping representation, with GOMMA and LogMap working on the accuracy of mapping as well. GOMMA is the only system among them that focuses on the continuity of mappings with the other factors already explained. Medical ontologies matching systems includes SAMBO [32] and ASMOV [33] as the most accomplished systems with their focus on mapping schemes, data translation, and mapping representation factors only. Data translation category consists of LinkEHR [34] and PPEPR [35] systems, with their focus on data translation from one health-care standard format to another. These were plenty of potential systems that can be categorized into these categories but we selected a few of them for explaining our point of view. These categories, factors, and the existing systems with the objective of semantic interoperability consists of limitations that hampers the progress in this area.

The existing systems only focus on mapping schemes but not on effective mapping representations. This is also the case because of limited design patterns incorporation in the mapping process. One of the limitation of the system is, there is no room for organization specific mapping handling, and dealing with accuracy of mappings for specific size of ontologies. Also, lack of appropriate transformation among EHR standards and not applicability of continuity of mappings are the bottleneck for comprehensive semantic interoperability frameworks. The proposed *Health-care Semantic Reconciliation Framework* resolves all these issues for a comprehensive semantic interoperability framework between medical systems compliant to heterogeneous standards. The two terms efficiency and effectiveness used in this thesis have different meanings for evaluating the system. Efficiency is the matching techniques used to generate mappings among the ontologies. Also, the ability to incorporate new matching techniques with the existing ones easily. Effectiveness is the use of the generated mappings for effective transformation from one health-care standard format to other.

Problem Statement

The problem with medical systems' compliant to heterogeneous standards is their inability to effectively communicate with each other due to non interoperable solutions and lack of semantic reconciliations applicability. Health-care standards follows reference models that are represented in the form of ontologies and can be mapped using ontology matching techniques. These issues related to this scheme is maintaining the accuracy of the mappings and handling organizations conformance by integrating it in the mappings. The challenges with these issues are to effectively represent the mappings in a manner easily understandable for verification and also to store these mappings in logic form that can be efficiently interpreted for transformation between various standard formats. To resolve the main semantic heterogeneity problem, there is requirement to set certain goals for achieving data interoperability.

The goals for the problem statement includes accurate generalized mappings, customized mappings, effective mapping representation, and standard formats transformation. Accurate generalized mappings can be achieved by utilizing various combination of ontology matching techniques and its storage in the mapping repository. Customized mappings are organization's specific mappings based on conformance with different concepts of the different standards. These compliments the generalized mappings in improving the accuracy of mappings stored in the repository. The mappings generated in the form of generalized and customized categories are required to be stored in effective mapping representation that helps in flexibility, extendibility, and reusability aspects. Standard formats transformation are based on the expressivity of the mappings representations and its stored logic, for transformation among different standards.

The main challenges for successfully achieving the goals for data interoperability is i) to maintain higher accuracy in generating mappings between heterogeneous standards with integrated organization conformance information; and ii) to take into consideration the design factors necessary to effectively store, represent, and transform the generated mappings. An ontology matching system that can incorporate the different ontology matching techniques is the effective solution for resolving the first challenge. An expressive ontology that represents the categories of the approaches in finding the mappings between different standards and also storing the logic based on these mappings used for transformation can resolve the second challenge.

The solution to the problems and challenges in the way of data interoperability is a framework that incorporates ontology matching for generating generalized mappings; handles organization conformance issues; provides expressive mappings representation ontology; resolves inconsistences with evolution of mappings; and finally effectively transform the standards formats for data exchange among medical systems.

1.2 Contributions

The goal of this research work is to provide a methodological framework for semantic reconciliations among different medical ontologies for semantic data interoperability among medical systems. In particular, the objectives of this research work is to: develop an ontology matching system used for matching ontologies and generating semantic mappings, formulate an approach to cater organizational conformance issues and generating organization specific customized mappings, design and develop a comprehensive ontological structure for ontology matching storage where semantic reconciliations takes place, prepare methodology for ensuring continuity of mappings by handling change detection, change collection, and change formulation, and finally execute transformation scheme for transformation between different health-care standards for data exchange between heterogeneous medical systems. Fig. 1.1 shows the abstract proposed framework called *Health-care Semantic Reconciliation Framework* with the main purpose to achieve the above mentioned objectives for semantic data interoperability between medical systems. The main contributions of the thesis in the proposed framework is described as follows in the subsequent sections.

Mapping Authoring

Medical systems interoperable exchange of data requires mappings between the the compliant standards of the communicating systems. Ontology matching provides the platform for matching different standard ontologies and storing the mappings in the mapping repository. We developed

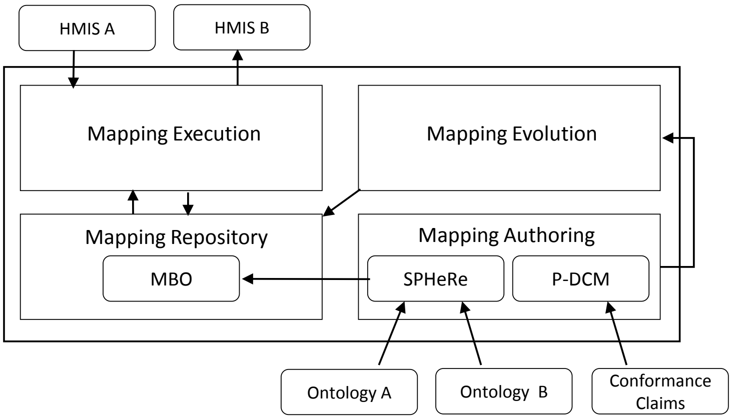


Figure 1.1: Abstract Health-care Semantic Reconciliation Framework

ontology matching system called System for Parallel Heterogeneity Resolution (SPHeRe), that is used for matching ontologies using ontology matching techniques such as string based, language based, taxonomy based, and informal resource based matching techniques. This matching is used for the generating the generalized mappings based on matching the generic ontologies of different health-care standards. We participated in the OAEI 2013 campaign with SPHeRe system to demonstrate its results in Large Biomed track.

Personalized-Detailed Clinical Model (P-DCM) approach uses ontologies for handling the conformance issues of the organizations. In the design and development of the P-DCM ontology, the annotations of the concepts of organizations are represented with different standard concepts. These annotations are then used for the generation of the customized mappings. Our contribution in Mapping Authoring component is SPHeRe ontology matching system for generalized mappings and P-DCM approach for customized mappings generation and storage with main focus on ontology matching. These define the accuracy of mappings stored in the mapping repository.

Mapping Repository

Ontology alignment patterns plays vital role in finding the alignments and also storing them in more expressive manner. The Mediation Bridge Ontology (MBO) provides the platform for storing the generalized and customized mappings with logic interpreting the transformation schemes among different standard formats. It uses object oriented design patterns as well as ontology alignment design patterns for addressing the design factors such as coupling, polymorphism, and rate of change. Based on these design factors we compared our proposed MBO with FALCON and LogMap ontology matching systems and found our results better than these systems to show the strength of design of our system.

Mapping Evolution

Continuity of Mapping drives the evolution of mappings by removing stale and redundancy in the mappings repository. Continuity of mapping module in the proposed framework caters evolution of mappings by obtaining the organization conformance information from P-DCM ontology. The change detection, change collection, and change formulation techniques are used for ensuring the continuity of mappings in the mapping repository.

Mapping Execution

Logic based transformation among different standard formats is the role of Mapping Execution module. The accuracy of the mappings stored is verified by the level of transformation that takes place between different standards. We have performed experiments and evaluation of HL7 CDA and vMR standards, by transforming their instances and validating the accuracy of the transformation process. The accuracy level with SPHeRe based transformation is approximatel 80% while SPHeRe and P-DCM based transformation resulted in approximately 94% of conversion.

1.3 Thesis Organization

This dissertation is organized into chapters as following.

- **Chapter 1: Introduction**. Chapter 1 provides brief introduction of the research work on semantic heterogeneity resolution for medical systems and in particular the role of semantic reconciliations in achieving data interoperability among medical systems. It focuses on the problems in the areas, the goals to achieve these problems, and finally the objectives achieved in this research work.

- **Chapter 2: Related Work**. A background detail is provided in this chapter about the ontology matching techniques and approaches, biomedical ontologies systems, mediation systems with the objective of achieving semantic data interoperability. Finally, it provides comparison of these systems with the proposed system of the research thesis to reflect the limitations of current systems addressed by the proposed system.

- **Chapter 3: Health-care Semantic Reconciliations Framework**. A proposed solution in the form of a framework for achieving data interoperability is presented in this chapter to overcome the limitations of current approaches. This chapter also provides overview of the concepts used in the thesis related to the proposed approach. It defines the scope of the thesis in achieving the semantic data interoperability among medical systems.

- **Chapter 4: Generalized Mappings and Mediation Bridge Ontology (MBO)**. An ontology matching system is presented in this chapter that is used for generating the generalized mappings between different medical ontologies. The details of the SPHeRe system and the algorithms used for mappings generation is described for storage in the MBO. Furthermore, this chapter describes the ontology alignment patterns used in the proposed MBO ontology. The examples and bridge patterns are defined and explained in detail to show its functioning for storage and representation of mappings in expressive manner.

- **Chapter 5: Customized Mappings, Mappings Evolution and Mapping Execution**. An approach is described that is used for generating the customized mappings by taking into account the conformance issues of the organizations. P-DCM approach takes organization specific conformed concepts and generates the customized mappings to be stored in the MBO. Also, this chapter reflects the changes in the MBO mappings that have been invoked by the customized mappings in the generalized mappings. Algorithms are described that provides continuity of mappings in the MBO by removing stale and redundant mappings from the mapping repository.

- **Chapter 6: Results and Evaluation**. The results and evaluation of different techniques used in the proposed framework are highlighted in this chapter. SPHeRe, MBO, and transformations between different standards and its accuracy is presented in this chapter.

- **Chapter 7: Conclusion and Future Directions**. This chapter concludes the thesis and also provides future directions in this research area. The main contribution of the thesis is also highlighted in this chapter.

Chapter 2

Related Work

Data and process interoperability demands mediation systems that can interpret communicating systems information easily. These mediation systems mostly are dependent on common mappings that behave as bridge among communicating systems. One such type of mediation ontology commonly known as bridge ontology is part of the literature, initially introduced in [36, 37], to the best of our knowledge. Semantic bridges were proposed by describing a structure of the bridge ontology but the work has limitations of providing limited set of bridges and no realization. The proposed system takes that concept as a baseline for development of Bridge Ontology for data interoperability among medical systems. The systems with the objective of interoperability can be expressed into different categories. The ontology matching systems help interoperable systems for understanding the translation provided after the matching process. There are ontology matching systems that specifically work on biomedical ontologies to integrate the commonalities and differences between these ontologies for the objective of interoperability. Another category is mediation systems, purpose of these are to interoperate between heterogeneous systems. These systems also depend on ontology matching for resolving the heterogeneity and trend of the use of ontology design patterns is increasing for development of these systems. There also exists mediation systems in literature that have worked on interoperability aspects systems compliant to heterogeneous health-care standards. These categories defines the flow of this chapter for identifying the steps necessary for achieving interoperability among health-care systems compliant to heterogeneous health-care standards. Some of the existing systems with the objective of interoperability are discussed in this chapter.

2.1 Ontology Matching Techniques and Approaches

We selected some ontology matching tools for discussion in this section based on their participation and adaption in OAEI, and also some of the existing state of the art systems. Falcon is one of the ontology matching systems that has shown best results in the first few years of OAEI campaign [31]. It provides fundamental technologies for finding, aligning and learning ontologies [31] by using divide and conquer approach to target large ontologies generating 1:n alignments as output [12]. Although this system is still effective in generating alignments between ontologies due to its matching techniques and also user interface; extendibility and reusability are its two major disadvantages. It is extremely difficult to add new matching matching techniques and algorithms in the system. Agreement Maker is another ontology matching tool that resolves extendibility issue by displaying the ontologies, supporting several mapping layers visually, and presenting automatically generated mappings for producing the alignments [38] [39]. This system is not scalable for large scale ontologies matching, but provides flexible and extensible framework with a comprehensive user interface. The scalability issue is resolved in its new framework AgreementMakerLight, that preserves original Agreement Maker framework with main focus on computational efficiency and handling very large ontologies [30]. AgreementMakerLight competes with the recent OAEI performers, GOMMA and LogMap in large bio-med track, but lacks approach for expressive mapping representation. GOMMA [28] provides infrastructure for managing matching and evolution of ontologies and its impact on mappings. On the other hand, LogMap is an ontology matching tool that address scalability issue for large ontologies matching and produces almost clean set of output mappings [29]. GOMMA and LogMap demonstrates better accuracy as compare to other systems and were equally matched by another matching tool YAM++. YAM++ matching tool supports self-configuration, flexibility, and extensibility in combining individual matchers [40]. It discover mappings using information retrieval techniques and also deals with multi-lingual ontologies matching problems [41]. Semantic Information Layer (SIL) [42] is an ontology mediation approach for data interoperability among Enterprise Information Systems (EIS). SIL is a part of the framework that extracts data from the data sources, queries the mapping information and provides the information to the upper layers for providing the mapping services to the enterprises.

2.2 Biomedical Ontologies Matching Systems and Approaches

Ontology matching systems that focus on resolving heterogeneities between biomedical ontologies assist in achieving interoperability. Two approaches based on matching biomedical ontologies are System for Aligning and Merging Biomedical Ontologies (SAMBO) and Automated Semantic Matching of Ontologies with Verification (ASMOV). SAMBO [32] is an ontology merging system based on matching biomedical ontologies. The process focuses on aligning relations and concepts. User involvement is necessary step while handling biomedical ontologies, SAMBO involves users for creation of every alignment that becomes hurdle in automatic generation of new ontology. These systems lack flexibility for adjusting with ontology evolution. Another ontology matching approach for information integration in field of bio-informatics is ASMOV [43]. Similarity calculation and semantic verification are the two main steps of this approach and uses WordNet and UMLS for increasing the accuracy level in similarity calculation. The primary focus of these two systems is on biomedical applications [12] and doesn't consider ontology mapping representation. A system called virtual Clinical Data Repository (vCDR) [44] is based on hybrid ontology integration driven approach, in which a DebugIT Core Ontology (DCO) behaves as a semantically mediation ontology for multiple semantically flat data description ontology (DDO). This work is extended to an architecture that caters real-time antimicrobial resistance monitoring to support transnational resistance surveillance with the use of semantic web based model for interoperability of inter-institutional and cross-border microbiology laboratory databases [45].

2.3 Patterns based Mediation Systems

Design patterns provide solution to the common occurring problem, and ontologies domain utilized ontology design patterns to facilitate the ontology development process. One of the semantic technologies potential areas that is focusing on incorporating design patterns as the solution to semantic heterogeneity problem is ontology matching that finds the similarities between concepts. Peigang Xu et al [46] proposed a differentor based similarity matrix creation technique used to integrate different similarities measures. Weights are assigned to various entities of the matching ontologies for aggregation tasks after finding the similarity measures. Another approach proposed

Tree Structure Based Ontology Integration (TSBOI) [47] methodology used to integrate ontologies with Document Type Definition (DTD) based tree structure development for ontology mappings. This is further utilized by ontology applications for data sharing purpose. These approaches leads to the development of ontology matching tools/ systems. OAEI provides a platform to introduce state of the art ontology matching tools, but their adaption for limited years and difficulty in extendibility, reusability, and expressive mapping representations defines the future directions for ontology matching tools. Some of these tools and approaches for ontology alignment patterns are discussed in this section.

Most of the ontology matching systems focus on automation and accuracy of results and not on expressive alignment representation using ontology alignment patterns. Zamazal et al. [48] presented a generic framework for ontology pattern detection, generation of instructions and ontology transformation from source ontology to target ontology. Scharffe et al. [49] took a step forward by introducing ontology alignment design patterns representation method and then create a pattern library to be extended with new patterns. The work also explains transformation of ontologies using ontology alignment patterns.

2.4 Health-care Standards based Mediation System

There are different medical systems that uses health-care standards for interoperability purpose. Some of the prominent work in literature with interoperability as objective among different health-care standards is described in this section.

Artemis [50] is a semantic mediation system between different Health Information Systems (HIS). It uses OWL-S as the approach for implementing semantic web services and uses HL7 as a standard for communication. Artemis uses OWL mapping tool (OWLmt) for the communication between sender and receiver providing semantic interoperability. OWLmt works as a mediator between sender and receiver by comparing sender ontology instances and receiver ontology instances with each other for ensuring the communication [51]. The primary focus of Artemis project is on data interoperability aspect by resolving heterogeneities between HL7 standards V2 and V3.

Plug and Play Electronic Patient Records (PPEPR) is a semantic SOA based platform with the objective of integrating heterogeneous Electronic Patient Records (EPRs). The integration in

PPEPR project is based on SOA, web services and semantics. The initial prototype of PPEPR tackles heterogeneity between two types of HL7 standard HL7 V2 to HL7 V3 [52]. They also propose mappings approach based on ontologies of heterogeneous HL7 standards [35].

Integrating the Health-care Enterprise (IHE)[1] is an initiative by health-care professionals and industry to coordinate implementation of standards for health-care systems integration [53]. IHE Integration profiles addresses interoperability assessment criteria based on interoperability at interface level, semantic level, legal and organizational level, and, security level [54]. IHE profiles makes significant contribution towards interoperability among health-care systems using different standards, still posses limitations mainly related to conformance of organizations to specific aspects based on their needs [54]. Therefore, process mediation is required when specifications lacks definition of these aspects of interoperability.

Jini Health Interoperability Framework (HIF-J) [55] uses Jini technology which is based on SOA. The main purpose of HIF-J is to exchange semantically interoperable messages. It provides translation services, that behaves as a mediator between standards. These translation services convert message instances HL7 V2 and V3 and also HL7 and openEHR message instances. It is based on XSLT transformations between message instances of different standards. Since standards are growing with new domains, so managing XSLT becomes very difficult. Moreover, XSLT is just transforming syntactic structure and semantic transformation is not achieved.

Ortho-EPR [56] standard is a proposed standard that is based on the integration of HL7 and DICOM standards for electronic orthodontic patient records. The main purpose of this standard is storage and communication of orthodontic patient records. The message part is handled by HL7 while imaging is handled by DICOM and there integration results in Ortho-EPR standard. Its main purpose is the integration of two standards and not interoperability between standards.

Other prominent work that provides tools for transformation includes LinkEHR (tool used for transformations among standards such as HL7, openEHR and CEN 13606) [34], and Poseacle Converter (CEN 13606 and openEHR standards archetypes and extracts transformation and validation) [16]. Jose A. Maldonado et al. [34] developed a tool called LinkEHR-Ed that is used for building, processing and validating archetypes based on multiple reference models. This editor

[1] http://www.ihe.net/

supports HL7 CDA, openEHR and CEN 13606 standards reference models and build archetypes based on their reference models. Catalina Martinez Costa et al. [16] work focused on dual model architectures standard transformation. CEN 13606 and openEHR standard archetypes and extracts transformation is discussed and validated using the Poseacle Converter [16]. The transformations are performed at archetypes and data levels.

In addition to these, in [57], the authors focus on semantic process interoperability with the help of interaction ontology in HL7 V3. Interaction ontology is responsible for handling the heterogeneities between processes of different health-care organizations compliant to HL7 V3 standard. This work is only related to semantic process interoperability using standard HL7 V3 and semantic data interoperability is not discussed. Ozgur Kilic et al. [58] proposed mapping algorithms that are used for generating mapping definitions. These mapping definitions are used for transformations among HL7 clinical statements instances and EHRcom instances.

Existing systems in literature highlights the importance of interoperability for CDSS systems. The role of ontologies becomes more important in the path towards interoperable CDSS. This is highlighted in [59], by describing the role of biomedical ontologies based on health-care standards to manage knowledge management, data integration and interoperability aspects and their fusion for decision support systems. Another project is SAPHIRE, a multi agent system supported by intelligent decision support system to improve patient lifecare by monitoring their activities. It depends on semantically enriched web services for communicating information to tackle interoperability [60]. Other than decision support systems, there are other systems that work on interoperability aspects among different standards.

2.5 Comparison with Proposed System

In summary, some factors are observed in the discussion that are the building blocks for interoperability. These includes: Medical Systems, Biomedical Ontologies, health-care standards, and Matching Systems. The above mentioned systems lacked at least one of these blocks therefore are considered as path towards interoperability. The proposed systems unlike these systems provides complete package for mediation systems to achieve true data interoperability. Accuracy and continuity of mappings provides the umbrella for all the building blocks. Accuracy not only consist of matching system but also includes conformance issues of organizations that involves personalization aspect. Continuity of mapping relates to the change management perspective for incorporating any changes that occurs in any standard and reflecting these in the mappings. In addition, it also relates to the changes that conformance can bring in the mappings already generated.

To summarize, existing ontology matching tools and ontology alignment patterns based approaches are unable to reflect a comprehensive system that utilizes object oriented design pattern combined with ontology alignment design patterns and storing the correspondences between matched ontologies into a mapping storage and representation repository. Once the mappings are generated and stored, their validity requires the level of accuracy achieved in transformation among different health-care standards.

Chapter 3

Health-care Semantic Reconciliation Framework

The proposed *Health-care Semantic Reconciliation Framework* behaves as a mediation system to resolve heterogeneity among HISs compliant to heterogeneous health-care standards for achieving data interoperability. The proposed approach is adaptive as it can be used for data interoperability among different health-care standards. It provides data interoperability services by addressing accuracy and continuity of mapping aspects. **S**ystem for **P**arallel **He**terogeneity **Re**solution (SPHeRe)[1] [61] [62] and Personalized-Detailed Clinical Model (P-DCM) [63] approaches are used for accuracy of mappings. We developed SPHeRe as an ontology matching tool to generate mappings between medical ontologies and store them in the *Mediation Bridge Ontology (MBO)* (a mapping representation ontology). *MBO* adopts ontology alignment patterns-oriented approach for the storage of semantic mapping information between heterogeneous standards. P-DCM approach handles the organization conformance details and results in evolution for continuity of mappings.

3.1 Semantic Interoperability and Health-care Standards

Information on the web and contained in knowledge bases of autonomous organizations is present in organizations required formats i.e., heterogeneous representations. The heterogeneity at conceptual level in data of two organizations is a complex problem and needs to be resolved for the purpose to share information. The heterogeneity resolution becomes critically important when the information sharing organizations are from the health-care domain because health-care is a complex domain in terms of complex concepts. The heterogeneity in health-care domain is at two levels: data and process. Health-care standards play an important role in achieving interoper-

[1] http://uclab.khu.ac.kr/sphere

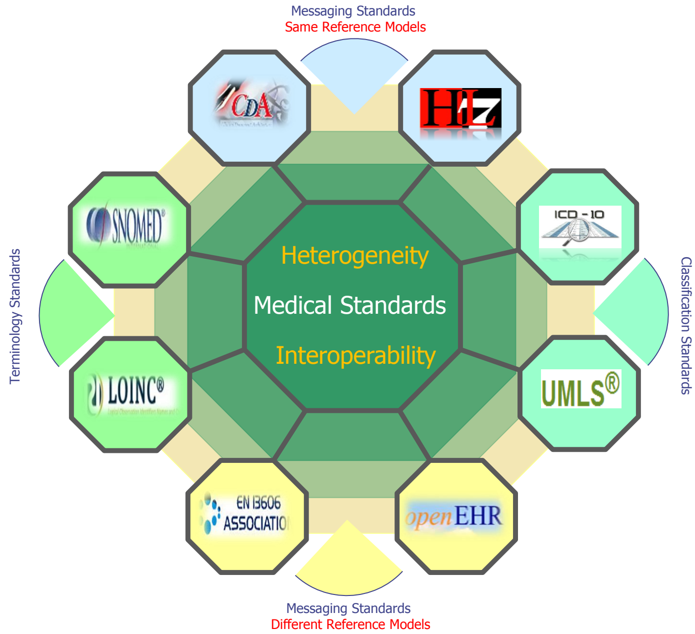


Figure 3.1: Heterogeneity among various Standards

ability between EHR systems [64]. Each health-care standard is based on its own objectives and goals. These include standards related to messaging (HL7)[2], terminologies (SNOMED CT), clinical information and patient records (openEHR and HL7 CDA), and imaging (Digital Imaging and Communications in Medicine (DICOM)). Moreover, some initiative like Integrating the Health-care Enterprise (IHE)[3] provide process level interoperability using existing standards. The main purpose of these standards is to provide interoperability between different health-care systems. Some of these standards are shown in Fig. 3.1.

3.2 Background

This section describes the background information necessary to understand the proposed approach. As a proof of concept, part of the proposed approach is integrated as an adapter [65] into Smart CDSS [1], an initiative taken in our lab [4] to develop a clinical decision support system that provides

[2]http://www.hl7.org/
[3]http://www.ihe.net/
[4]http://uclab.khu.ac.kr/

recommendations and guidelines to physicians and patients. The adapter is specifically build to resolve heterogeneities among different health-care standards so that legacy systems compliant to heterogenous standards can utilize Smart CDSS services. There are different terminologies and techniques that are used in the proposed approach in order to increase the understandability of the system. These includes our system SPHeRe, used for generalized mappings; our approach P-DCM, used for customized mappings; our ontology MBO, used to use patterns for generation and storage of mappings; and finally the health-care standards which we have used to evaluate the proposed approach. Therefore, we discuss Smart CDSS, *SPHeRe* and *P-DCM* approaches, and *MBO* as background for understanding the proposed approach.

3.2.1 An Overview of Smart CDSS

Smart CDSS is standard based clinical decision system that provides recommendations to physicians and patients based on heterogeneous data sources including clinical data, social media data, behavior modeling data, and activities and emotion recognition data [66] [67]. Among its different features, interoperability of HISs and smart homes compliant to different standards with Smart CDSS is a key challenge. This kind of interoperability is considered as data level interoperability which is the ability to communicate data among systems with the original semantics of the data retained irrespective of its point of access [68]. This challenge can be resolved by resolving heterogeneity between different heterogeneous health-care standards.

Smart CDSS consumers include systems that are compliant to different health-care standards. Smart CDSS can only process information in vMR standard. Therefore, an adapter is required to transform HMIS compliant health-care standard to Smart CDSS compliant health-care standard and vice versa. The proposed approach in the form of **A**dapte**R** **I**nteroperability **E**ngine (ARIEN) system facilitates Smart CDSS in achieving interoperability with different HISs. ARIEN performed experiments on HL7 CDA and vMR ontologies for mappings generation and then using these mappings for instance level transformation. The proposed system ARIEN is part of the Adaptability Engine as shown in Fig. 3.2.

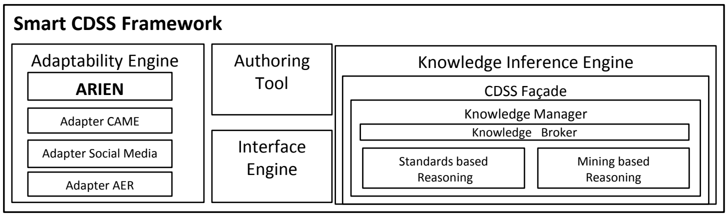


Figure 3.2: Architecture of Smart CDSS [1]

3.2.2 SPHeRe and P-DCM Approach

We have designed and developed *SPHeRe* system that is an ontology matching system used for matching health-care standard ontologies to generate generalized mappings. It uses different bridge algorithms to generate mappings and store them in a particular format addressed by the *MBO*. For example, *Overlap Bridge* matches concepts based on class, attributes, and their value by using *Overlap Pattern Relationship Model* and stores the mapped information between ontologies in the *MBO*. The problem with generalized mappings is its inability to accommodate organizational conformance information. Organizations can conform to particular concepts, and the non-conformed concepts lead to some of stale generalized mappings. These organizational conformance based mappings are categorized as customized mappings and our approach P-DCM manages these mappings.

3.2.3 Mediation Bridge Ontology (MBO)

MBO is a bridge ontology that is based on ontology design patterns and stores alignments between matching ontologies. *MBO* is categorized into two main classes *MediationBridge* and *PatternClass*. MediationBridge is divided into syntactic and structural bridge subclasses: *String Matching Bridge, Label Bridge, Synonym Bridge, Overlap Bridge, Customized Bridge, CBSB,* and *PBSB*. These bridge classes are used to represent the alignments generated from particular algorithms in their specified format. These are dependent on *PatternClass* for structuring the output of the alignment process. *PatternClass* include *MappedSequence, Standard1Class, Standard2Class, Match, MappedClass, ClassLabel, ListStandard1,* and *ListStandard2* subclasses. These are used to provide the structure for representation of the alignments in the *MBO*.

3.2.4 HL7 CDA and vMR Standards

Health-care standards play a vital role towards interoperability among HISs. HL7 provides a family of standards for achieving this goal. Some of the HL7 standards that we are currently using in our Smart CDSS version discussed in the previous section include HL7 CDA, vMR, Medical Logic Module (MLM), and Arden Syntax. HL7 CDA is used for medical document generation and exchange by the HMIS. HL7 vMR is the medical record that is used by the Smart CDSS system for the processing of medical information. MLM is the standard used for representing the clinical guidelines that can be used in the decision making process [69]; these are examples of sharing via a common knowledge representation format [70]. MLM's are written in Arden Syntax to represent clinical and scientific knowledge in an executable format [71].

Table 3.1: Structural Transformation Pattern

CDA Skeleton	vMR Skeleton
<?xml version="1.0"?> <ClinicalDocument xmlns="urn:hl7-org:v3" xmlns:voc="urn:hl7-org:v3/voc" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="urn:hl7-org:v3 CDA.xsd"> <typeId root="" extension=""/> <templateId root=""/> <id extension="" root=""/> <code code="" codeSystem="" codeSystemName="" displayName=""/> <title> </title> <effectiveTime value=""/> <confidentialityCode code="" codeSystem=""/> <languageCode code=""/> <setId extension="" root=""/> <recordTarget></recordTarget> <author></author> <custodian></custodian> <component> <structuredBody> .... </structuredBody> </component> </ClinicalDocument>	<?xml version="1.0" encoding="UTF-8"?> <cdsInput xmlns="urn:hl7-org:v3/vmr" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:dt="urn:hl7-org:v3/cdsdt" xsi:schemaLocation= "urn:hl7-org:v3/vmr ..\Schemas\cdsInput.xsd"> <templateId root=""/> <cdsContext> <cdsSystemUserPreferredLanguage code="" codeSystem="" displayName=""/> </cdsContext> <vmrInput> <templateId root=""/> <patient> <id root="" extension=""/> ..... </patient> </vmrInput> </cdsInput>

HL7 CDA and vMR are our case study standards for data interoperability between HMISs compliance with CDA and CDSSs compliance with vMR. Both of the standards are based on the HL7 Reference Information Model (RIM) [72] that is the root of all the information models and consists of backbone classes, and their specialization and structural attributes for further defining the roles of the classes. The core classes are Act, Entity, Role, Act Relationship and Participation. The RIM is specialized into Domain Message Information Model (DMIM) that specializes the RIM core classes and uses its sub classes based on a particular domain. The DMIM is specialized into Refined Message Information Model (RMIM) that forms a base for different messages in

the domain, and it will be applicable to one or more Hierarchical Message Definitions (HMD) [72]. HL7 CDA follows a CDA RMIM [73] that contains information about document creation and manipulation. A CDA document can be transferred within a message or independently [74]. The skeleton structure of CDA record is shown in Table 3.1 (as CDA Skeleton). Likewise, a vMR is also an HL7 standard that is based on the RIM and is used in clinical decision support (CDS). VMR is a data model for representing clinical data relevant to CDS by recording patient's demographics and clinical history data [75]. It is divided into two types of structures: vMR Input and vMR Output. VMR Input models the input information of the patient and, after processing and guidelines generation, vMR Output is used to model that information. VMR Input is shown in Table 3.1 (vMR Skeleton). Integration of HMIS compliant to CDA and CDSS compliant to vMR requires transformation between these standards; Table 3.1 shows the Structural transformation pattern when conversion from CDA to vMR Input is necessary. On the other hand, vMR Output is converted to CDA when guidelines are generated and are communicated with HMIS.

3.2.5 HL7 and openEHR Standards

HL7 V3 is a messaging standard used for communicating medical information between health information systems. On the other hand openEHR is patient records related content modeling standard related to management and storage, retrieval and exchange of health data in the form of electronic health records (EHR's). Both of these standards are based on a reference model and also follow layer of constrained models. As openEHR is based on EHR generation and manipulation, one of HL7 standard also fulfills the criteria called HL7 Clinical Document Architecture (CDA). HL7 CDA is based on Reference Information Model (RIM) for the generation of EHR document. HL7 models are explained in the previous section, therefore we explain openEHR models and their relationship.

3.2.5.1 openEHR Models

openEHR is based on two level modeling approach, a stable reference information model constitutes the first level of modeling, while formal definitions of clinical content in the form of archetypes and templates constitute the second [76]. openEHR follows layers of constrained model and consists of EHR Information Model, Demographic Information Model, Data Structure Information Model, Common Information Model, Data types Information Model, and Integration Information Model. On the other hand, the clinical content related information are handled by Template Object Model (TOM) and Archetype Object Model (AOM). Archetypes and templates are the formal models of domain concepts controlling data structure and content of data [77]. Archetypes are the constraints based models of domain content expressed in a formal language called Archetype Definition Language (ADL) [78].

3.2.5.2 Relationship of HL7 and openEHR

More specifically, one of HL7 standard HL7 CDA, which is also based on RIM is closely aligned with openEHR. Four different types of mappings conversions are of consideration between HL7 and openEHR communities [79]; 1) HL7 Model to an openEHR archetype, 2) openEHR Archetype to HL7 Model, 3) instance of HL7 Model to instance of an openEHR Archetype, 4) instance of openEHR Archetype to an instance of HL7 Model.

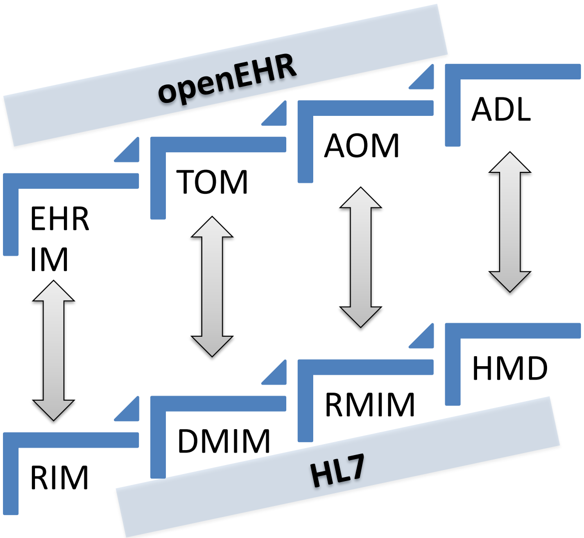


Figure 3.3: Relationship of HL7 and openEHR Models

Ontologies of both the standards are available openly covering the information related to their reference models. The different level models of HL7 and openEHR can be compared according to their scope as shown in Fig. 3.3. RIM of HL7 V3 can be mapped with the information models of openEHR (EHR IM [80], Demographic IM [81], Data Structure IM [82], EHR Extract IM, Integration IM, Common IM [83], Data Structure IM [84] and Support IM [85]. The TOM and AOM openEHR models resembles with HL7's DMIM and RMIM models respectively. In similar way, HMD represents the contents of the RMIM which resembles with ADL of openEHR that represents Archetypes and Templates.

3.3 Health-care Semantic Reconciliation Framework

The goal of the proposed approach, *Health-care Semantic Reconciliation Framework* is to achieve data interoperability which is dependent on the accuracy and continuity of mappings. The layered architecture for semantic reconciliation is shown in Fig. 3.4. The health-care standards provides specifications that is modeled in ontological form. Some of these standards are described in the previous section that includes HL7 CDA, vMR and openEHR. The four layers highlighted in Fig. 3.4 are the contributions of this thesis. These includes Mapping Authoring, Mapping Repository, Mapping Evolution, and Mapping Execution. There details are presented as follows:

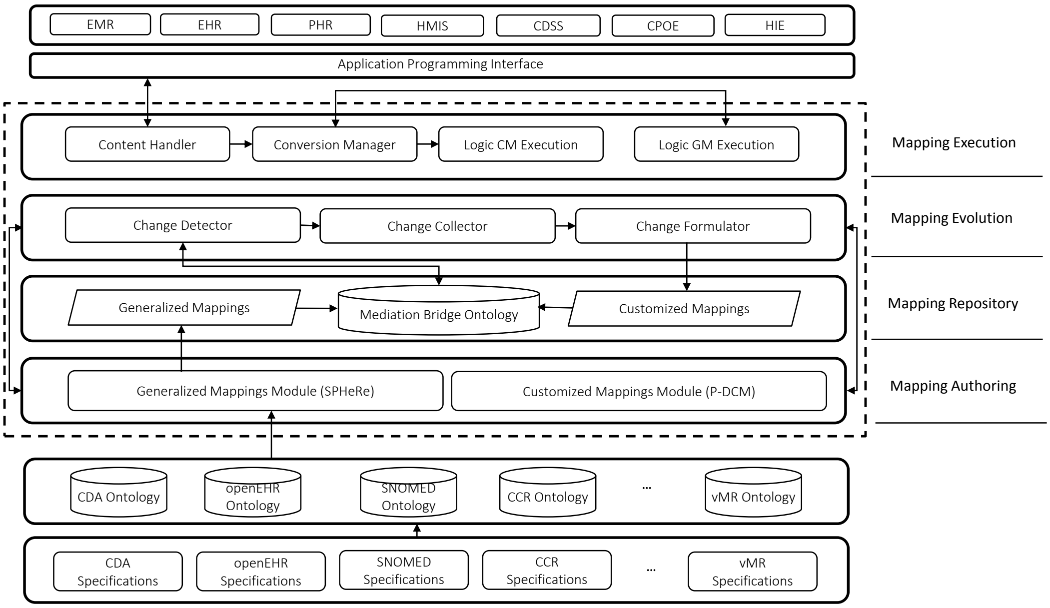


Figure 3.4: Semantic Reconciliation Layered Architecture

3.3.1 Mapping Authoring

This process is based on using different standards information for generating and storing mappings. The steps include the creation of ontologies from standards specification, validation of ontologies, loading of heterogeneous ontologies, matching ontologies, storing mappings, and expert verifications. Mapping generation process creates the foundation for accuracy and continuity of mappings for achieving data interoperability. The process consists of *Generalized Mapping* and *Customized Mapping* modules. *Generalized Mapping Module* module is responsible for accuracy of mappings by generating mappings between heterogeneous health-care standards and storing them in a mapping repository using ontology matching techniques. The *Customized Mapping Module* deals with the organizational conformance issues for increasing the accuracy of the mappings stored.

3.3.1.1 Generalized Mapping Module

This component deals with applying the bridge algorithms for the generation and storage of ontology mappings between two matching medical ontologies. It takes as input any two health-care standards ontologies for the matching process. The proposed system focused on CDA and vMR standards and developed ontologies for each using their standard specifications.

We developed an ontology matching tool called *SPHeRe* for generating generalized mappings between medical ontologies. SPHeRe is an effective ontology matching system that performs computationally intensive operations using optimized matching algorithms executed over matched medical ontologies. It is a high performance-based initiative that improves ontology matching performance by exploiting parallelism over multicore commodity hardware of Cloud Platform [86]. It consists of different bridge algorithms for generalized mappings generations. *String Matching Bridge* is used to identify similar concepts in the matched ontologies using string matching techniques. *Synonym Bridge* identifies similar meaning concepts by utilizing information from online dictionary available, such as WordNet [87]. *Label Bridge* match use labels of matching concepts stored in the ontologies to identify similar labels for concluding matched concepts as similar concepts. It also uses string matching techniques for matching labels of the concepts. *Overlap Bridge* finds concepts that contain overlapping information and are necessary for information exchange.

It uses structural hierarchies' information for finding overlapping concepts. *Child Based Structural Bridge* and *Property Based Structural Bridge* use concepts children similarity matching and property similarity matching techniques to find similar concepts in matching ontologies. Ontology alignment patterns are also used in these bridge algorithms for alignments in ontology mediation process. These algorithms are used for generating *Generalized Mappings* between medical ontologies and then storing them in the *MBO*.

3.3.1.2 Customized Mapping Module

Organizational conformance with concepts based on their requirements effects the accuracy of mappings and is handled by *P-DCM* approach. The information of conformance with specific concepts and non-conformance with particular concepts categorizes these mappings as *Customized Mappings*. *P-DCM* approach focuses on the generation of *Customized Mappings* that creates the necessary linkage between organization's conformed health-care standards concepts and clinical model concepts to ensure data interoperability among HISs [63]. *Customized Bridge* is used to represent P-DCM based mappings between health-care standard ontologies and storing these in *MBO*. *P-DCM ontology* annotates the clinical model concepts with different standard concepts and also represents the non-conformed concepts. The non-conformed concepts are responsible to invalidate some of the mappings in generalized mappings. Therefore, deprecating those mappings is necessary for smooth information exchange for a particular organization's HIS communication processes with other HISs. In summary, *Customized Mappings* in addition to *Generalized Mappings* help in improving the overall accuracy of the mappings.

3.3.2 Mapping Repository

The *MBO* is the mapping repository that stores the *Generalized* and *Customized* mappings generated by *SPHeRe* and *P-DCM* approach respectively. It's called *MBO* because it is the centralized mediation point for all the other components of the proposed approach. *Mapping Authoring* layer stores mappings in it, *Mapping Evolution* applies change management for evolving mappings, and *Mapping Execution* uses these mappings for transformation from one standard format to another. Automation of alignments generation and storage, scalability for adding more bridge algorithms and evolution to changes are the features of *MBO*. Also, it focuses on the mapping representation aspect for management and reuse of the stored alignments. Expert verification of the *MBO* mappings is required after the mapping generation process is completed for further strengthening the accuracy of mappings.

3.3.3 Mapping Evolution

Customized Mappings invalidates some of the *Generalized Mappings* in the *MBO* due to organization's conformance issues by bringing change to the existing mappings. These changes are necessary for mappings evolution to ensure continuity of mappings. *Mapping Evolution* layer executes change management functions to evolve the mappings for changes to take effect in the *MBO*. *P-DCM ontology* contains all the information about conformed and non-conformed concepts and relationship of clinical model concepts with different standard concepts. *Customized Mappings* can influence *MBO* by insertion, deletion, and modifications of mappings. The steps involved in this component are as follows:

- *Change Detector* listens for non-conformed concepts information from the *P-DCM ontology*.
  - Non-conformed concepts are accessed from *P-DCM ontology*.
  - Mappings based on non-conformed concepts are identified in the *MBO* and stored as stale mappings.
  - A single stale mapping is composed of a conformed concept aligned with a non-conformed concept.
- *Change Collector* accesses stale mappings from *Change Detector*.
  - Conformed concepts are collected from the stale mappings.
  - Annotated mappings based on the conformed concept are searched in the *P-DCM ontology*. A corresponding conformed concept of the target standard concept is identified from the annotated mapping information. For example, if the conformed concept A belongs to standard 1 and non-conformed concept B belongs to standard 2, then based

on A its corresponding conformed concept C of standard 2 is accessed from *P-DCM ontology* and an annotated mapping is generated.

  – *P-DCM ontology* annotated information are also searched to find out some new mappings that didn't existed before for storing in the *MBO*. For example, concept X and Y of standard 1 and 2 are related with clinical model concept Z in *P-DCM ontology* which is the annotated information. Concept X and Y are candidate alignment in this case, therefore if this alignment is not already present in the *MBO* then it is a new candidate mapping.

- *Change Formulator* collects the modified and new mappings identified by *Change Collector* and formulates it in *MBO* compatible format for storing.

The whole change management process discussed above is intended to accomplish continuity of mappings. Continuity of mappings complements accuracy of mappings to achieve data interoperability among HISs. This paper only highlights the procedure for continuity of mapping issue for data interoperability and thoroughly describes the standard transformation process using mappings stored in the *MBO*.

3.3.4 Mapping Execution

Measuring the level of effectiveness of accuracy and continuity of mappings determines the degree of data interoperability accomplished among HISs. *Mapping Execution* layer provides realization to a certain level of effectiveness of data interoperability among HISs. This includes structural, sequential and data transformation mechanism to convert and generate valid health-care standard formats. We consider mapping based transformation between HL7 CDA and vMR standards for proof of concept and evaluating the ability of the proposed system to achieve data interoperability. Legacy HISs can utilize the health-care standard transformation services of this step directly and communicate with other HISs.

This layer executes mediation strategy designed for enabling communication between HISs that are compliant with heterogeneous health-care standards. It consists of four sub-components:

- *Content Handler* that communicates with communicating HISs in their complaint standard formats.

- *Conversion Engine* that executes the actual transformation process between different standards.

- *Logic CM Execution* executes the customized logic stored in the *MBO*.

- *Logic GM Execution* executes the generalized logic stored in the *MBO*.

We assume that HMIS compliant with CDA standard wants to take benefits from CDSS compliant with vMR standard. Therefore, HMIS communicates with *Content Handler* and provides clinical document containing patient information in CDA format. The basic purpose is to obtain recommendation from CDSS that is compliant to vMR standard. *Content Handler* forwards information to *Conversion Engine* for applying transformation process, that uses *Logic CM Execution* to identify the matched pattern from the *MBO* based on the input information. The *Logic CM Execution* consists of the logic information based on organization conformance. If the logic in not found in the *Logic CM Execution*, the information is passed to *Logic GM Execution* that converts the input information of one standard format to output format of another standard format using pattern identified from the *MBO*. In our scenario, CDA standard format requires to be transformed to vMR format for CDSS processing. *Conversion Engine* converts CDA to vMR by applying transformation patterns modeled in *MBO*. CDSS processes information; generates recommendation in vMR output format that is converted back to CDA format by *Conversion Engine* using *Logic CM Execution*, and *Logic GM Execution* modules. The final recommendation information is provided to HMIS via *Content Handler* and thus data interoperability is achieved. The detail of every layer internal functionalities is described in the proceeding chapters.

3.4 Summary

This chapter presented the background information necessary for building the proposed framework. Also, it explained the health-care standards to and their structure to be used for matching purpose. Finally, it provided insight into the concepts that are used for performing different functionalities of the framework.

Chapter 4

Generalized Mappings and Mediation Bridge Ontology(MBO)

Ontology mappings enables accessibility of information by aligning the resources in ontologies belonging to diverse organizations [88]. These also resolves semantic heterogeneities among data sources. Mainly two steps are required to overcome semantic heterogeneity: Matching resources to determine alignments and interpreting those alignments according to application requirements [12].

SPHeRe is an ontology matching system that utilizes cloud infrastructure for matching large scale ontologies and focus on alignment representation to be stored in the Mediation Bridge Ontology (MBO). MBO is the mediation ontology that stores all the alignments generated between the matched ontologies and represents it in a manner that provides maximum metadata information.

The proposed MBO ontology is an ontology alignment representation scheme that enables expressiveness to formalize correspondence by utilizing object oriented and ontology alignment design patterns. Ontology matching has made measurable progress to resolve semantic heterogeneities among heterogeneous data repositories for ontology merging, query answering, or data translation [12]. Its existing schemes only focus on matching the ontologies and storing their alignments in a format which only describes source and target concepts. Ontology Alignment Evaluation Initiative[1] (OAEI), a benchmarking initiative, carries out annual campaigns for the evaluation of the ontology matching tools [89]. In the past few years, OAEI evaluated several ontology matching systems; some of these remained in spotlight for many years. Shvaiko et al. [12] presents a survey of some of the recent matching systems based on their operations and matching approaches. A common behavior among these matching tools is their duration of use that lasts

[1] http://oaei.ontologymatching.org/.

for few years and are replaced afterwards. The main reasons for their replacement is the difficulty in extendibility and reusability of these systems. The structure of matching systems should be extensible enough to accommodate new algorithms based on novel matching techniques, replace previous algorithms if they are non-effective, and utilize combination of existing techniques to build new technique. Therefore, incorporating object oriented design patterns [90] with ontology design patterns in ontology matching tools define the longer adaption of such systems.

There is need to find the alignments using design patterns for providing solutions to the common problems. Also, expressiveness in the storage of correspondence is necessary for multiple reasons. First, expert verifications become easier as the correspondence speaks for itself. The correspondence includes not only the source and target concepts, but the attributes involved in correspondence, the procedure of the alignments, and the confidence value of the alignment. Second, feedback about the matching process and alignment can be easily obtained, that helps in the overall improvement of the system and satisfaction of the users. We developed SPHeRe ontology matching system that incorporates bridge algorithms which are stored in expressive alignment representation format in the MBO. Strategy[2] and Mediator[3] design patterns are used from object oriented design patterns, combined with ontology alignment patterns called Pattern Relationship Models (PRM). The PRMs are the ontology alignment patterns that defines the expressive formal representation of the correspondences to be stored in the MBO. The proposed system supports collaborative ontology concepts by adding metadata information in alignments stored in the MBO. This helps in achieving extendibility and reusability metrics of the overall SPHeRe system.

4.1 SPHeRe

Our proposed MBO is part of the ontology matching system we developed called SPHeRe [61, 91]. SPHeRe system is based on different bridge algorithms that are represented in a mapping representation format provided by the MBO. Accuracy and performance are the two factors that helps in achieving the goals, and these are accomplished by Matcher Library and Performance Matching Framework[4] of SPHeRe working model as shown in Fig. 4.1.

[2]http://www.oodesign.com/strategy-pattern.html.
[3]http://www.oodesign.com/mediator-pattern.html.
[4]Performance Matching Framework is not in the scope of this thesis.

SPHeRe Execution Control manages the communication with external and internal entities. It is responsible for ontology loading and providing information about execution of bridge algorithms in the Matcher Library and to Performance Matching Framework for parallel execution of the algorithms. Matcher Library consists of bridge algorithms such as String Matching Bridge, Synonym Bridge, Label Bridge, Overlap Bridge, Customized Bridge, CBSB, and PBSB.

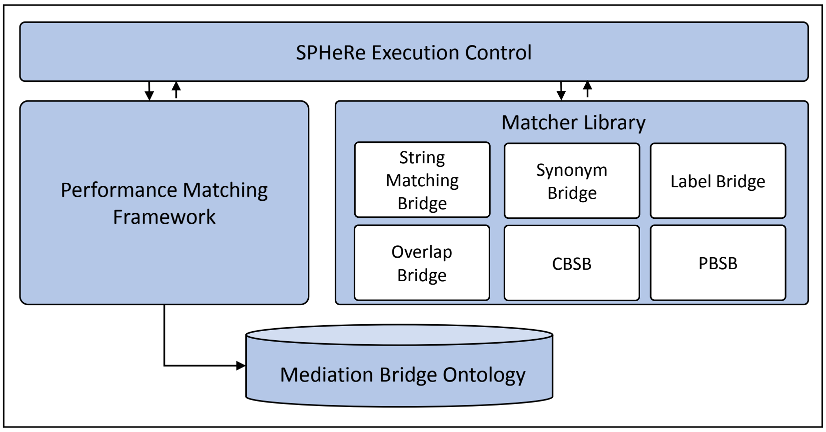


Figure 4.1: System for Parallel Heterogeneity Resolution (SPHeRe) Architecture

The process workflow for SPHeRe system is controlled by the SPHeRe Execution Control that initially loads the ontologies to be matched using ontology loader function. Bridge algorithms to be invoked from the Matching library are invoked by algorithms assessment function and provided to the Performance Matching Framework. The Performance Matching Framework executes the bridge algorithms in efficient manner and generates results. The generated results are compiled by the matched result functions and population of the mappings in the MBO starts with the use of design patterns. The mappings are represented and stored in the MBO based on the bridge algorithm and the pattern relationship model used for generating mappings.

4.2 SPHeRe Bridges Definition, Examples and Scenarios

The MBO provides the platform that represents alignments found by different bridge algorithms. These bridge algorithms are defined and explained with real world examples, and scenarios using medical standard ontologies. We use medical standard ontologies as scenarios for the proposed system. The alignments generated and represented in the MBO can be used for ontology translation, standard format transformation, and expert verification based on metadata availability about every alignment. One of the bridge algorithms is String Matching Bridge that is used for concepts matching to identify similar concepts in the matched ontologies. These are based on string based matching techniques by considering the sequence of letters of matching concepts. These sequence of letters consideration for matching are based on the intuition that the more similar the strings, the more likely is the chance of the concepts to be similar [92]. Edit distance is one of the technique used for string based matching techniques [93]. Table 4.1 show the examples and medical ontologies scenario of SPHeRe's bridge algorithms.

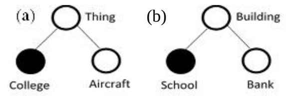
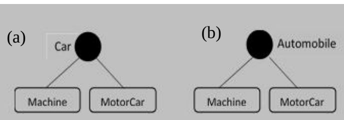
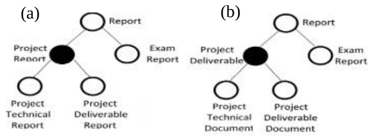
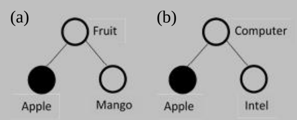
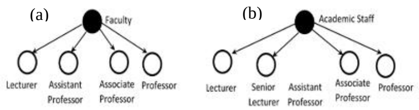
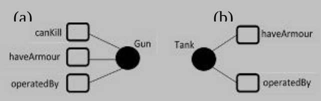
4.3 SPHeRe Matching Algorithms

SPHeRe system is based on bridge algorithms run in the parallel execution environment to generate alignments to be stored in the MBO. Matcher Library components stores all the bridge algorithms to be run on the matching environment represented by Performance Matching Framework. Communication between these two components is regulated by SPHeRe Execution Control module that behaves as a controller. The alignments are stored in the MBO; generated by the bridge algorithms stored in Matcher Library that are run by the Performance Matching Framework.

The working environment for SPHeRe system is presented in Fig. 4.2. Matching ontologies are input to the *SPHeRe Execution Control*, that extracts concepts from these ontologies for matching purpose. Object oriented design patterns define the execution flow of the matching libraries. MBO Mediation design pattern is used for the bridge algorithms that are dependent on the output of others. These bridge algorithms are executed from the matching library with mediation among them provided by the MBO Mediation design patterns. In the same way MBO Strategy design pattern is used to execute the bridge algorithms is a specified flow. The threshold value is matched after the bridge algorithms are executed and if the threshold value is achieved then the mappings are stored in the MOB otherwise the mapping is discarded.

String Matching Bridge provides matching results by finding similar concepts based on string matching techniques in the matching ontologies. This includes applying the string matching algorithms such as prefix, suffix, edit distance, n gram and others. The equation *f(x)* describes

Table 4.1: Mediation Bridge Ontology Concepts, Examples, Scenarios.

Bridge	Example	Medical Ontologies Scenario
Synonym Bridge		[SNOMED CT and Mesh ontology]: Concept DRUG of SNOMED CT ontology which is synonym of concept MEDICINE of Mesh ontology.
Label Bridge		[FMA and NCI ontology]: Concept CARTILAGE CELL of FMA ontology which is similar to concept CHONDROCYTE of NCI ontology based on common label CARTILAGE CELL.
Overlap Bridge		[HL7 and openEHR ontology]: OBSERVATION concept exists in both standard ontologies, and EVALUATION is the sub-concept of OBSERVATION concept in openEHR ontology. Therefore, EVALUATION concept of openEHR ontology can also be transformed to OBSERVATION concept of HL7 ontology while information exchange.
Polysemous Bridge		[SNOMED CT and HL7 ontology]: EVENT concept in SNOMED CT ontology includes concepts that represent occurrences of different events while in HL7 ontology it is any act that has taken place. EVENT concept of SNOMED CT ontology and HL7 ontology are Polysemous in nature.
Child Based Structural Bridge		[HL7 and openEHR ontology]: ENTITY concept of HL7 ontology is equivalent to PARTY concept in openEHR ontology based on their children matched. ENTITY concept has ORGANIZATION, PERSON and DEVICE sub concepts that are mapped with ORGANIZATION, PERSON and AGENT sub concepts of the PARTY concept.
Property Based Structural Bridge		[HL7 and vMR ontology]: OBSERVATION concept belongs to HL7 ontology while OBSERVATION RESULT concept is part of vMR ontology. Both the concepts are similar based on property match. CODE, CODE SYSTEM, and DISPLAY NAME are the common properties between the concepts that leads to the conclusion of property based match.

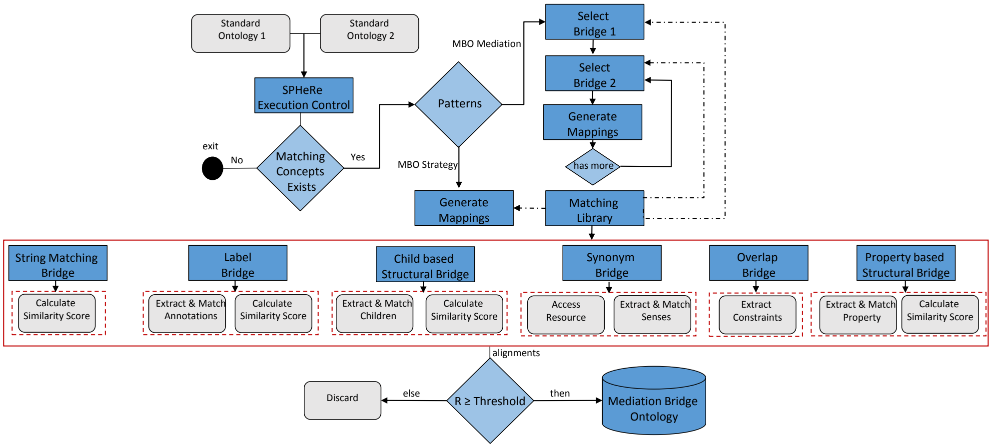


Figure 4.2: Flowchart of SPHeRe System

performing string matching tasks $\bigcap$ on any two concepts C_i and C_j of the ontologies O_s and O_t respectively. The string matching techniques are applied to find similarity score matching value. A threshold $Threshold$ value of n is set for matching in String Matching Bridge algorithm to limit the number of impure mappings. The equation for String Matching Algorithm is as follows:

$$f(x) \longleftarrow C_i \bigcap C_j, \quad where \quad \bigcap \in \{set \quad of \quad String \quad based \quad matching \quad techniques\},$$
$$and \quad \forall C_i \in O_s, \forall C_j \in O_t$$
$$\mathbb{R}_{i=1}^{C_n} \mathbb{R}_{j=1}^{C_n} f(x)$$

Label Bridge uses the labels of the source and target concepts for matching. Initially, concept labels are normalized e.g. using stop word elimination, then list of the source concept labels are matched with list of the target concept labels. If any label in the lists matches, the source and target concepts are stored in the MBO as mappings. The function *h(x)* is responsible for eliminating the stop words and special characters. The normalized labels of concepts C_i and C_j of the ontologies O_s and O_t respectively are mapped using String Matching Bridge methodology. This label matching of concepts is represented by function *g(x)*. The equations for label bridge are as follows:

$$h(x) \longleftarrow C_i.label_s, \quad where \quad \{i \in \{SourceOntology, TargetOntology\}\},$$
$$and \quad s = \quad labels \quad without \quad stop \quad words \quad and \quad special \quad characters$$
$$\Sigma_{i=1}^{C_n} h(x)$$

$$g(x) \longleftarrow C_i.labels \bigcap C_j.labels,$$
$$where \bigcap \in \{set \ of \ string \ based \ matching \ techniques\},$$
$$and \quad \forall C_i \in O_s, \forall C_j \in O_t$$
$$\mathbb{R}_{i=1}^{C_n} \mathbb{R}_{j=1}^{C_n} g(x)$$

Synonym Bridge is based on finding the similarity between concepts using wordnet [87]. The relationship is identified based on matching the synonyms of the concepts accessed using wordnet. The function *k(x)* represents synonyms extraction using wordnet dictionary for source C_i and target C_j concepts. The number of common synonyms is calculated using string matching bridge methodology for calculating the matching value. If its value is less than the threshold then this alignment is discarded, otherwise stored in the MBO. The equation for synonym bridge is as follows:

$$k(x) \longleftarrow Synonymsof(C_i) \bigcap Synonymsof(C_j),$$
$$where \quad \bigcap \in \{set \quad of \quad string \quad based \quad matching \quad techniques\},$$
$$and \quad \forall C_i \in O_s, \forall C_j \in O_t$$
$$\mathbb{R}_{i=1}^{C_n} \mathbb{R}_{j=1}^{C_n} k(x)$$

Child Based Structural Bridge (CBSB) bridge generates mappings between source and target ontologies based on matching children of the concepts. The function l(x) represents initially, children of source C_i and target C_j concepts are accessed as lists. The number of common children in the lists is identified and finally the matching value $SimScore$ is calculated and compared with the threshold $Threshold$ that is assigned value n. Property Based Structural Bridge (PBSB) uses String Matching Bridge techniques to match properties of source and target concepts for finding similar properties. This information is utilized as in PBSB for matching the source and target ontologies concepts based on their properties. The equation for CBSB is as follows:

$$l(x) \longleftarrow C_i.Children \bigcap C_j.Children/Avg(No\ of\ C_iChildren, No\ of\ C_j.Children),$$
$$where\ \ \bigcap \in \{set\ \ of\ \ string\ \ based\ \ matching\ \ techniques\},$$
$$and\ \ \forall C_i \in O_s, \forall C_j \in O_t$$
$$\mathbb{R}_{i=1}^{C_n}\mathbb{R}_{j=1}^{C_n} l(x)$$

Overlap Bridge uses the initial matching performed using other algorithms and finds the mandatory attributes associated with the matching concepts. If the matched concepts are associated with the mandatory attributes then this information is also stored in the MBO. This information is useful for transformation purpose from source to target format. The function *m(x)* represents the use of Overlap Bridge for source and target concepts, C_i and C_j respectively. The equation for Overlap Bridge is as follows:

$$m(x) = C_i \bigcap C_j \mid \{C_i \bigcap C_j\} \neq \emptyset$$
$$A \leftarrow \exists a \mid a \in C_y$$
$$wherey = i, j$$

These bridge algorithms use ontology alignment design patterns for generating and storing the mappings in the MBO.

4.4 Ontology Alignment Design Patterns

Ontology Design Patterns (ODP) support pattern based ontology design [94], and are used to capture common modelling situations, help facilitate ontology development and avoid common mistakes [94]. ODP have evolved from Content Ontology Design Patterns (reusable solutions to recurrent content modelling problems) [95] to Ontology Alignment Design Patterns (used to refine correspondences, by alignment designer or pattern detection algorithm) [49]. Ontology matching algorithms detect simple correspondences by following alignment format that lacks the expressiveness needed to formalize correspondence [96]. Therefore, an approach is necessary to design and develop an extendible and reusable system that provide expressive capability to formalize correspondences. The proposed Mediation Bridge Ontology (MBO) based approach incorporates object oriented design patterns combined with ontology alignment design patterns in our ontology matching system, System for Parallel Heterogeneity Resolution (SPHeRe) [61].

4.5 Mediation Bridge Ontology (MBO)

Ontology mediation techniques provide the platform to necessitate interoperability between heterogeneous ontological descriptions [49]. Mediation is based on the alignments generated between heterogeneous sources, and representation of these alignments play vital role in effective interoperability. Little focus is provided to alignment representation area by the semantic web community [97]. An effort towards representing correspondences as a centralized repository was introduced as bridge ontology [23] [24], but it lacked effectiveness, agility, and realization. Although it provided the base for alignment representation but was never the focus, mainly because accuracy of alignments is given higher priority. Effective alignment representation results in; (1) efficient ontology translation, (2) format transformations, (3) systems mediation, and (4) easy expert verification and modification. Therefore, the proposed MBO targets effective alignment representations in its design and development process. The design aspect utilizes object oriented as well as ontology alignment design patterns for effective mapping representation in the form of low coupling, high polymorphism, and low rate of change. The MBO not only benefits the ontology alignment storage but also its use in the transformation process between different heterogeneous formats. The proposed MBO provides the alignment representation scheme using ontology design patterns approach keeping in mind the goals expressed for achieving true interoperability.

MBO is categorized into two main classes MediationBridge and PatternClass. MediationBridge is divided into syntactic and structural bridge subclasses: String Matching Bridge, Label Bridge, Synonym Bridge, Overlap Bridge, Customized Bridge, Children Based Structural Bridge, and Property Based Structural Bridge. These bridge classes are used to represent the alignments generated from particular algorithms in their specified format. These are dependent on PatternClass for structuring the output of the alignment process. PatternClass include MappedSequence, Standard1Class, Standard2Class, Match, MappedClass, ListStandard1, and ListStandard2 subclasses. These are used to provide the structure for representation of the alignment in the MBO. The overall hierarchy of MBO is shown in Fig. 4.3.

These concepts are related to each other by using object properties, its triples are shown in Table 4.2. MediationBridge class is related through usesPattern object property with PatternClass. Every subclass of MediationBridge use some pattern classes from the PatternClass subclasses to define its alignment representation. OverlapBridge class is related through hasSourceClass and hasTargetClass object properties with Standard1Class and Standard2Class respectively. Standard1Class uses hasSameRelationship and consistsMandatoryAttributes object properties to connect with Standard2Class and MandatoryAttribute respectively. Based on the previous triples, OverlapBridge is related with the Match class using hasRestriction object property. This makes the complete alignment representation for OverlapPRM described in the later section. In the same way, other MediationBridge classes defines their pattern to represent alignment in the MBO.

4.6 MBO Design Patterns

MBO utilizes Strategy Design Pattern and Mediator Pattern to incorporate object oriented design approach for agility and reusability of the system. It also used PRM to define mapping representation format that can be used for easy expert verifications, format transformation, and ontology translation purposes. Fig. 4.4 represents class diagram that shows MBO Strategy Design Pattern, MBO Mediator Pattern and Pattern Relationship Models (PRM) as realization of the MBO in the SPHeRe system. MBO Strategy and Mediator design patterns explains the implementation

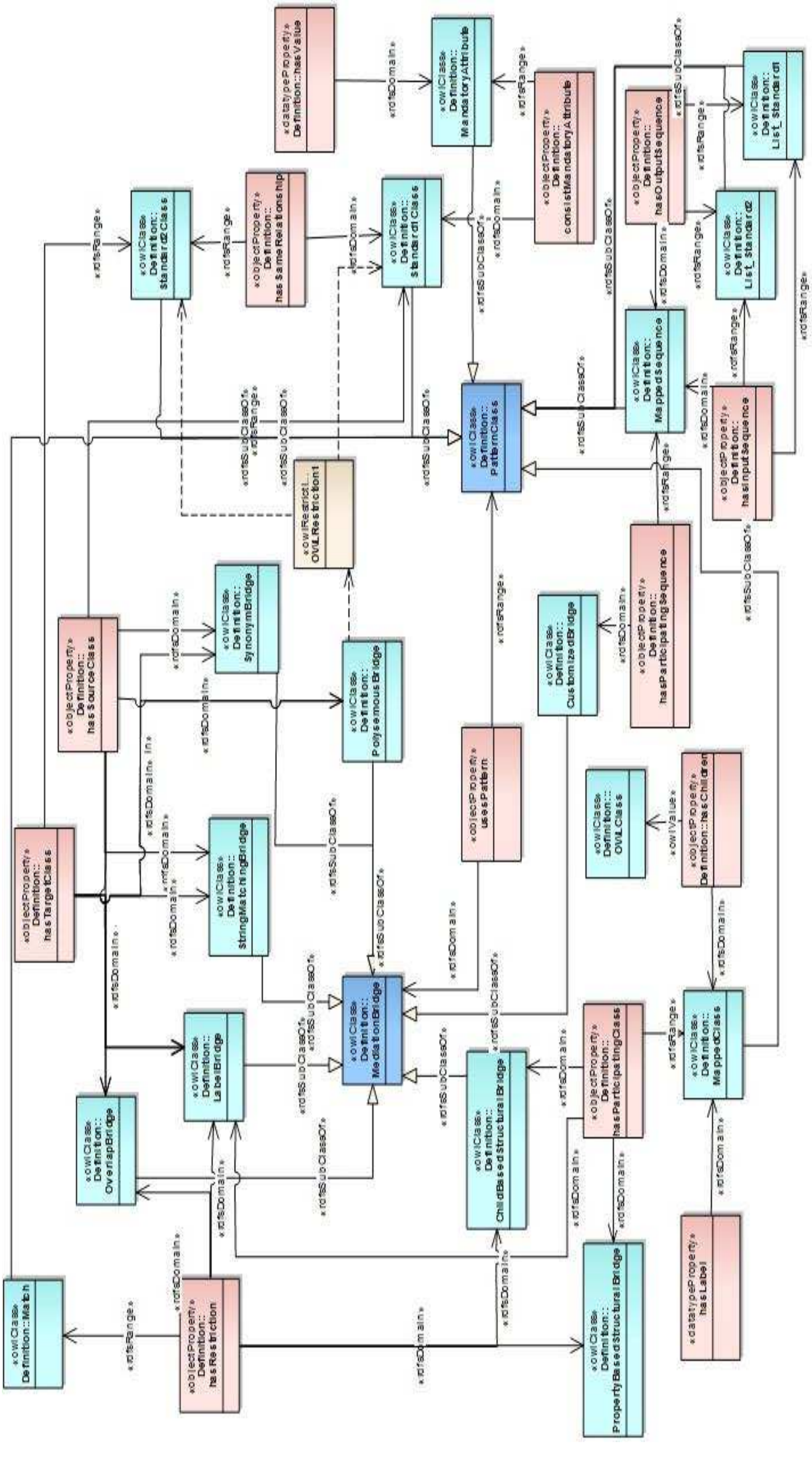


Figure 4.3: Mediation Bridge Ontology

Table 4.2: Mediation Bridge Ontology (MBO) triples

Domain	Property	Range
Mediation Bridge	usesPattern	PatternClass
Standard1Class	exactMatch	Standard2Class
Customized Bridge	hasParticipatingSequence	MappedSequence
Label Bridge Overlap Bridge Polysemous Bridge String Matching Bridge Synonym Bridge	hasSourceClass	Standard1Class
Label Bridge Overlap Bridge Polysemous Bridge String Matching Bridge Synonym Bridge	hasTargetClass	Standard2Class
Mapped Class	hasChildren	owl:Class
Mapped Sequence	hasInputSequence	ListStandard1 ListStandard2
Mapped Sequence	hasOutputSequence	ListStandard1 ListStandard2
Label Bridge Overlap Bridge PBSB CBSB	hasRestriction	Match
Standard1Class	hasSameRelationship	Standard2Class
PBSB CBSB	hasParticipatingClass	MappedClass
Standard1Class	consistsMandatoryAttribute	MandatoryAttribute

view of the system design, while PRMs describes MBO ontology patterns as representation of the alignments. We have adopted the concept of Strategy and Mediator design patterns from the object oriented design community and proposed PRM in this research by interrelating them for extendible, flexible and agile system.

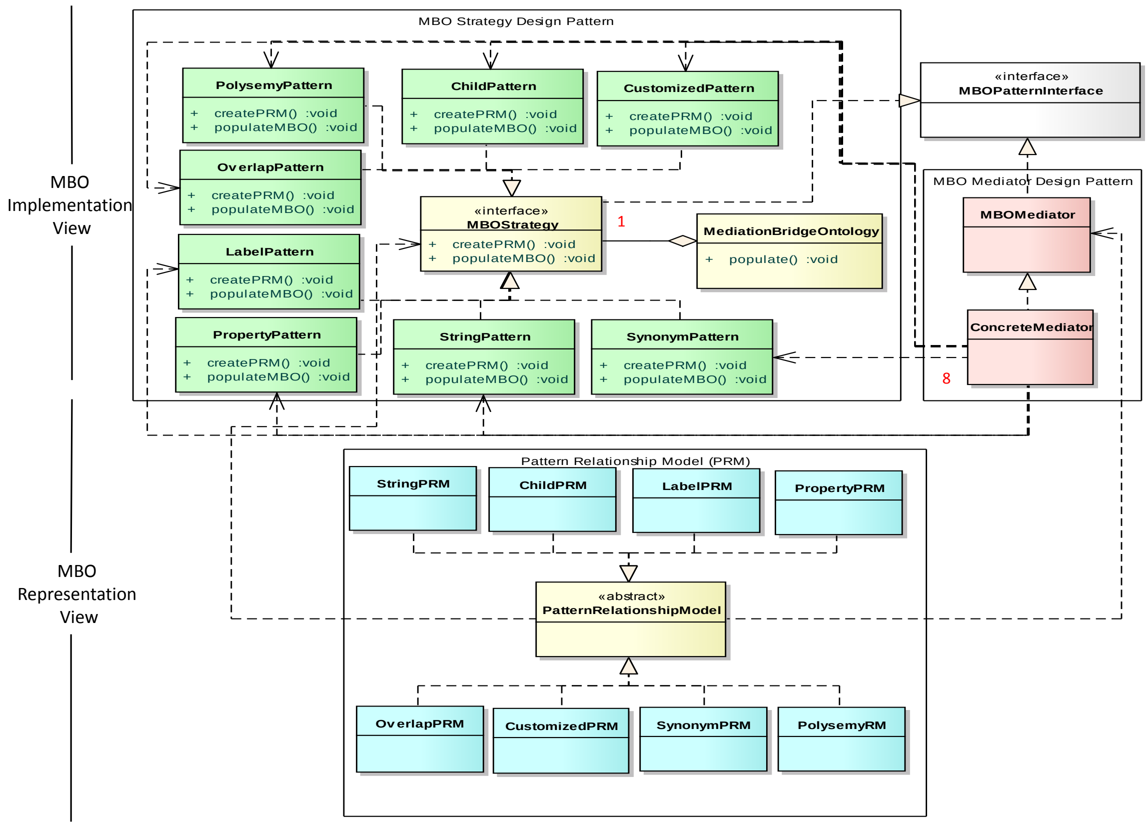


Figure 4.4: MBO Design Patterns Oriented Implementation and Representation Views

4.6.1 MBO Implementation View

4.6.1.1 MBO Strategy Design Pattern

Motivation:	MBO is based on classes that differ only in their behavior, therefore algorithms needs to be isolated to provide the ability to select different algorithms at runtime.
Intent:	Define a family of algorithms, encapsulate each one, and make them interchangeable. MBOStrategy lets the algorithm vary independently from clients that use it.
Applicability:	*MBOStrategy* - an interface that defines the behavior of a MediationBridgeOntology. Concrete Strategies: *ChildPattern, PropertyPattern, StringPattern, SynonymPattern, PolysemyPattern, OverlapPattern, LabelPattern, CustomizedPattern*; each of these pattern classes calls specific PRM for execution and then populate that information in the MediationBridgeOntology. *MediationBridgeOntology* - This class is the context class that gets alignments information from each pattern and store it in specified format.

4.6.1.2 MBO Mediation Design Pattern

Motivation:	MBO also provides classes that can use the services of other classes, therefore mediation is necessary between classes for reusability purpose.
Intent:	Define an interface for communicating with related objects for understanding interdependencies among them. *MBOMediator* provides that interface to other objects for communicating with related objects.
Applicability:	*MBOMediator* - An interface class used for communicating with other objects in well-defined and complex ways. ConcreteMediator - This class keeps reference of all the colleague objects and is used to transfer the messages between colleague classes such as *ChildPattern, PropertyPattern, StringPattern, SynonymPattern, PolysemyPattern, OverlapPattern, LabelPattern, and CustomizedPattern*.

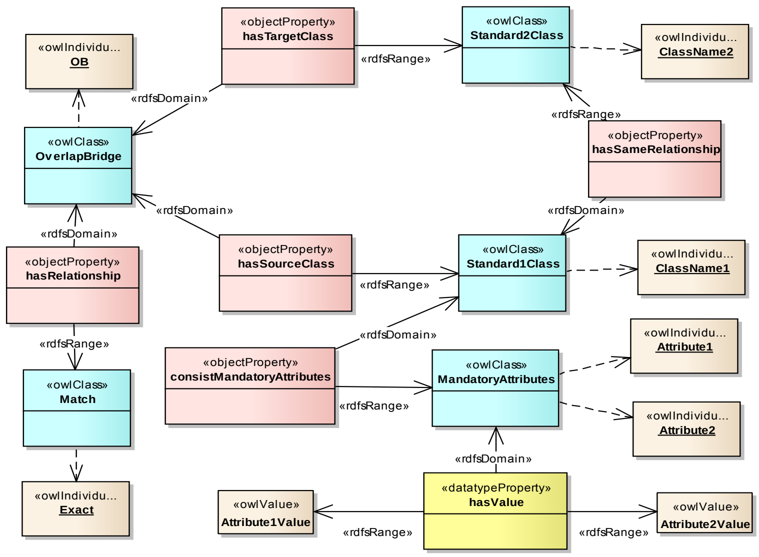
4.6.2 MBO Representation View

Each pattern class in the Strategy Design Pattern uses particular PRM class e.g. StringPattern class invokes StringPRM class for execution. All the PRM classes are derived from PatternRelationshipModel abstract class. Medical ontologies are used for matching purposes and performing experiments, therefore, medical standards are used as scenarios for understanding these PRMs. These PRMs realization is shown with Virtual Medical Record (vMR) and HL7 Clinical Document Architecture (CDA) standards ontologies. Both of the standards are based on the HL7 Reference Information Model (RIM) [72] that is the root of all the information models and consists of backbone classes, and their specialization and structural attributes for further defining the roles of the classes. HL7 CDA follows a CDA Refined Message Information Model (RMIM) [73] that contains information about document creation and manipulation. VMR is a data model for representing clinical data relevant to CDS by recording patient's demographics and clinical history data [75]. The generic pattern structure followed by its realization in vMR and CDA standard ontologies is described in this section. Some of these PRMs are explained in object oriented design template as follows:

4.6.2.1 Overlap Pattern Relationship Model (OverlapPRM)

Motivation:	*OverlapPRM* deals with the type of alignment patterns where source ontology concept with its mandatory attributes and values is mapped with target ontology concept.
Intent:	Define a mechanism to transform source and target concepts by taking into account mandatory attributes as well. The mapping representation targets *Overlap Bridge* of MBO.

Implementation: *OverlapBridge* class has relationship with *Standard1Class* and *Standard2Class* through *hasSourceClass* and *hasTargetClass* object properties respectively. *OverlapBridge* class is related with *Match* class using *hasRelationship* object property with individuals *Exact* or *Subsume*. There are cases in which mandatory properties of both the standards are exactly matched while in some cased source concept has subsumption relationship with target concept. *Standard1Class* and *Standard2Class* are also related with each other using *hasSameRelationship* object property. *Standard1Class* consists of *MandatoryAttribute* connected by *consistMandatoryAttributes* object property and these *MandatoryAttribute* contains some values represented by *hasValue* data type property. The pattern for *Overlap Bridge* is shown as:



Applicability: HL7 CDA consists of classes in the form of triplet ”class-attribute-value”. Attributes are divided into mandatory and optional categories. Therefore, while transformation of concepts between vMR and CDA these mandatory attributes transformation is necessary for correct parsing of the document. Class Mandatory Attributes Value to Class Matching Pattern deals with such type of patterns where source standard concept with its mandatory attributes and values is converted into target concept. In this type of pattern an ontology O_i consist of class C_i with mandatory attributes MA_i having values V_i is mapped with class C_j of another ontology O_j.

We explain *OverlapPRM* with *EntryRelationship* concept of CDA standard with *RelatedClinicalStatement* concept of vMR concept. *EntryRelationship* class of HL7 CDA has mandatory attributes such as *typeCode* and *contextConductionInd* with values *CAUS* and *true* respectively. This information is mapped with *RelatedClinicalStatement* class of vMR, therefore translation of *RelatedClinicalStatement* class is performed with *EntryRelationship* class and its mandatory attributes and values. This realization of CDA and vMR concepts alignment using OverlapPRM is represented in OWL format as:

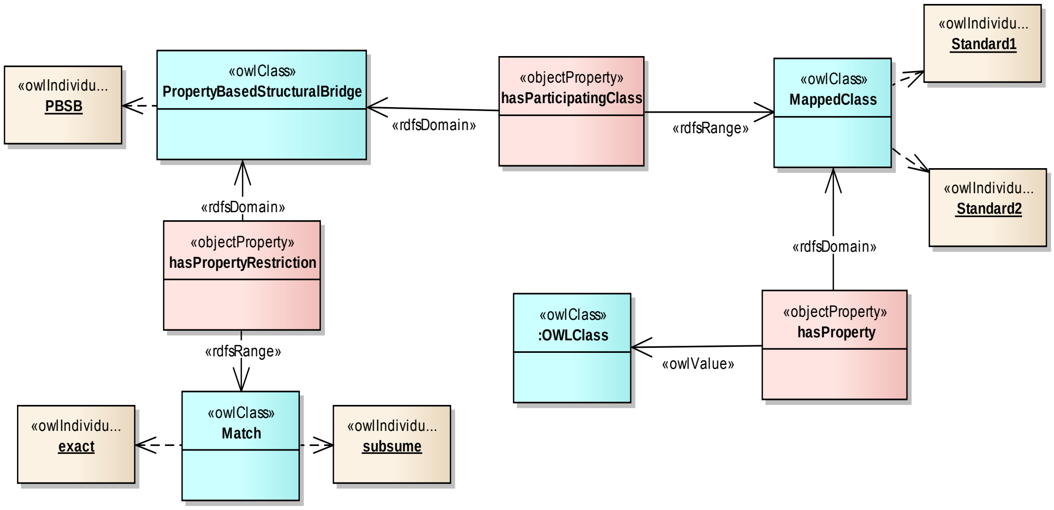
```
<rdf:RDF
 xmlns:vmr="http://www.owl-ontologies.com/VMR.owl#"
 xmlns:cda="http://www.owl-ontologies.com/CDA.owl#">
 <owl:Ontology rdf:about="BridgeOntology"/>
 <!-- Defining Classes for Class Mandatory Attributes Value
 to Class Matching Pattern -->
  <owl:Class rdf:ID="OverlapBridge"/>
  <owl:Class rdf:ID="MandatoryAttribute"/>
  <owl:Class rdf:ID="Match"/>
  <owl:Class rdf:ID="Standard1Class"/>
  <owl:Class rdf:ID="Standard2Class"/>
 <!-- Object Properties -->
   <owl:ObjectProperty rdf:ID="consistMandatoryAttributes">
    <rdfs:domain rdf:resource="#Standard1Class"/>
    <rdfs:range rdf:resource="#MandatoryAttribute"/>
   </owl:ObjectProperty>
   <owl:ObjectProperty rdf:ID="hasRelationship">
    <rdfs:domain rdf:resource="#OverlapBridge"/>
    <rdfs:range rdf:resource="#Match"/>
   </owl:ObjectProperty>
   <owl:ObjectProperty rdf:ID="hasSameRelationship">
    <rdfs:domain rdf:resource="#Standard1Class"/>
    <rdfs:range rdf:resource="#Standard2Class"/>
   </owl:ObjectProperty>
   <owl:ObjectProperty rdf:ID="hasSourceClass">
    <rdfs:domain rdf:resource="#OverlapBridge"/>
    <rdfs:range rdf:resource="#Standard1Class"/>
   </owl:ObjectProperty>
   <owl:ObjectProperty rdf:ID="hasTargetClass">
      <rdfs:domain rdf:resource="#OverlapBridge"/>
      <rdfs:range rdf:resource="#Standard2Class"/>
   </owl:ObjectProperty>
 <!-- CDA Class with its mandatory attributes and values -->
   <Standard1Class rdf:ID="CDA_EntryRelationship">
    <hasSameRelationship rdf:resource=
    "#RelatedClinicalStatement"/>
    <consistMandatoryAttributes rdf:resource="#TypeCode"/>
    <consistMandatoryAttributes rdf:resource=
    "#ContextConductionInd"/>
   </Standard1Class>
   <owl:DatatypeProperty rdf:ID="hasValue">
    <rdfs:domain rdf:resource="#MandatoryAttribute"/>
    <rdfs:range rdf:resource="&xsd;string"/>
   </owl:DatatypeProperty>
   <MandatoryAttributes rdf:ID="ContextConductionInd">
    <hasValue rdf:datatype="&xsd;string">true</hasValue>
   </MandatoryAttributes>
   <MandatoryAttributes rdf:ID="TypeCode">
    <hasValue rdf:datatype="&xsd;string">CAUS</hasValue>
   </MandatoryAttributes>
   <Match rdf:ID="Exact"/>
  <!-- VMR Class -->
   <Standard2Class rdf:ID="VMR_RelatedClinicalStatement"/>
  <!-- Overlap Bridge Relationship -->
   <OverlapBridge rdf:ID="OverlapBridgeInd">
    <hasSourceClass rdf:resource="#EntryRelationship"/>
    <hasRelationship rdf:resource="#Exact"/>
    <hasTargetClass rdf:resource="#RelatedClinicalStatement"/>
   </OverlapBridge>
 </rdf:RDF>
```

4.6.2.2 Property Pattern Relationship Model (PropertyPRM)

Motivation: *PropertyPRM* deals with the type of alignment patterns where properties of the source ontology concept matches with the properties of the target ontology concept.

Intent: Define a mechanism to compare properties of source and target concepts and represent them as alignment if particular threshold is reached. This pattern reflects the mappings for Property Based Structural Bridge (PBSB).

Implementation: Three main classes *PropertyBasedStrcuturalBridge, MappedClass* and *Match* are related to each other by object properties *hasParticipatingClass, hasProperty* and *hasPropertyRestriction*. Each individual of *PBSB* class is related with *MappedClass* individuals from different standards by *hasParticipatingClass* object property. Each individual of *MappedClass* consists of properties in the form of *OWL:Class* related by *hasProperty* object property. These properties should be having exact or subsumes relationship with each other. Therefore, *PBSB* class individual is related with any of the *Match* class individuals using *hasPropertyRestriction* object property. This information identifies the nature of relationship between the matched classes. *PropertyPRM* is described in the MBO as:



Applicability: *Observation* class of CDA standard is equivalent to *ObservationResult* class of vMR standard based on their matching properties using *PropertyPRM*. *Observation* class has *Code, EffectiveTime, and Value* as its properties and *ObservationResult* class has *ObservationFocus, ObservationEventTime, ObservationValue* properties. Observation's class property *Code* is related with *ObservationFocus* property of *ObservationResult* class using *LabelPRM* and categorized under *Label Bridge*. In the same way, *EffectiveTime* and *Value* properties of *Observation* class are related to *ObservationEventTime* and *ObservationValue* properties of vMR class respectively. *SynonymPRM* that categorizes mapping information under *Synonym Bridge* is used for *EffectiveTime* and *ObservationEventTime* properties, while *StringPRM* is used for *Value* and *ObservationValue* matching by categorizing it under *String Matching Bridge*. An instantiation example for *PropertyPRM* is described as:

```
<rdf:RDF
 xmlns:vmr="http://www.owl-ontologies.com/VMR.owl#"
 xmlns:cda="http://www.owl-ontologies.com/CDA.owl#">
 <owl:Ontology rdf:about="BridgeOntology"/>
 <!-- Defining Classes for Property Match Pattern -->
 <owl:Class rdf:ID="PBSB"/>
 <owl:Class rdf:ID="MappedClass"/>
 <owl:Class rdf:ID="Match"/>
 <!-- Properties of Observation Class in CDA -->
 <owl:Class rdf:about="&cda;Code"/>
 <owl:Class rdf:about="&cda;EffectiveTime"/>
 <owl:Class rdf:about="&cda;Value"/>
 <!-- Observation Class associated with its properties -->
 <MappedClass rdf:ID="CDA_Observation">
   <hasProperty rdf:resource="&cda;Code"/>
   <hasProperty rdf:resource="&cda;EffectiveTime"/>
   <hasProperty rdf:resource="&cda;Value"/>
 </MappedClass>
 <!-- Properties of ObservationResult class in VMR -->
 <owl:Class rdf:about="&vmr;ObservationEventTime"/>
 <owl:Class rdf:about="&vmr;ObservationFocus"/>
 <owl:Class rdf:about="&vmr;ObservationValue"/>
 <!-- ObservationResult class associated with its properties -->
 <MappedClass rdf:ID="VMR_ObservationResult">
   <hasProperty rdf:resource="&vmr;ObservationFocus"/>
   <hasProperty rdf:resource="&vmr;ObservationEventTime"/>
   <hasProperty rdf:resource="&vmr;ObservationValue"/>
 </MappedClass>
 <!-- Indiviual of Match class -->
 <Match rdf:ID="exact"/>
 <!-- Indiviual of PBSB class -->
 <PBSB rdf:ID="PBSB_INS_CDA_VMR">
   <hasPropertyRestricvtion rdf:resource="#exact"/>
   <hasParticipatingClass rdf:resource="#CDA_Observation"/>
   <hasParticipatingClass rdf:resource="#VMR_ObservationResult"/>
 </PBSB>
 <!-- Relationship between PBSB and MappedClass -->
 <owl:ObjectProperty rdf:ID="hasParticipatingClass">
   <rdfs:domain rdf:resource="#PBSB"/>
   <rdfs:range rdf:resource="#MappedClass"/>
 </owl:ObjectProperty>
 <!-- Relationship between MappedClass and OWL:Class -->
 <owl:ObjectProperty rdf:ID="hasProperty">
   <rdfs:domain rdf:resource="#MappedClass"/>
   <rdfs:range rdf:resource="&owl;Class"/>
 </owl:ObjectProperty>
 <!-- Relationship between PBSB and Match class -->
 <owl:ObjectProperty rdf:ID="hasPropertyRestricvtion">
   <rdfs:domain rdf:resource="#PBSB"/>
   <rdfs:range rdf:resource="#Match"/>
 </owl:ObjectProperty>
</rdf:RDF>
```

4.7 MBO Artifacts

The scope of MBO is categorized into three aspects; Generalized Mappings, Customized Mappings, and Transformation Logic. Generalized Mappings are the alignments that are generated by matching two ontologies using PRMs. Customized Mappings are the alignments that are based on the conformance issues handling of organizations. Organizations conformance issues are handled through these mappings by detecting the stale mappings initially in the Generalized Mappings and then replacing them with the new modified mappings. The generalized as well as the customized mappings are converted into transformation logic that is used for conversion among different standard formats. This whole process is called as MBO Alignment Linkset, as MBO stores not only the alignments but also performs transformation as shown in Fig. 4.5.

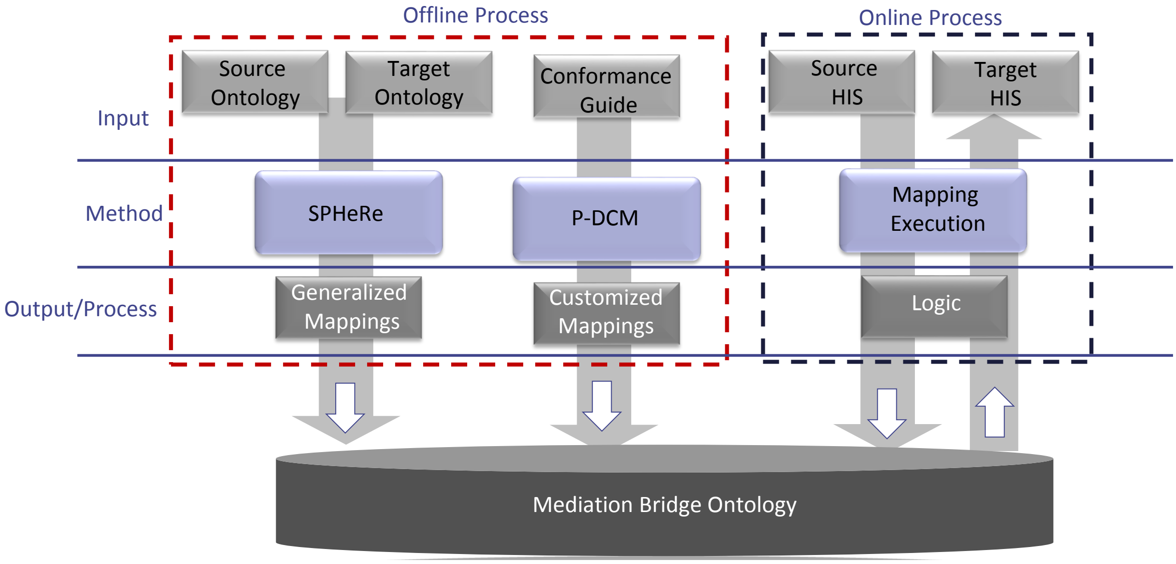


Figure 4.5: MBO Artifacts

4.7.1 Formal Modeling and Representation of MBO

MBO formal modelling using Backus-Naur Form (BNF) is described in this section. MBO constructs are defined by the generalized and customized mappings which are then represented in logic format for transformation among different standards. The generalized mappings are the focus of this paper and it includes the alignment information with the ontology alignment design pattern used for the creation of generalized mappings logic to be used for transformation. The formal definitions of all these concepts as well as the transformation logic based on generalized mappings is presented as follows:

```
<MBO>::= "Generalized Mappings:" <GM>
         "Customized Mappings:" <CM>
         "Transformation Logic:" <Logic>
```

```
<GM>::= "AlignmentInfo:" <AlignInfo>
        "Pattern Relationship Model:" <PRM>
        "Logic GM:" <LogicGM>

<AlignInfo>::= "Source Entity:" <SE>
               "Target Entity:" <PRM>
               "Measure Threshold Value:" <PRM>
               "Relationship:" <R>
```

$$< SE >::= \{x \mid O_1 \cap x \in < S_\Delta, x_i >\}$$

$$< S_\Delta, x_i >::= \{(x_i \in S_\Delta) \wedge (S_\Delta \in O_1)\}$$

$$< TE >::= \{x \mid O_2 \cap x \in < T_\Delta, x_i >\}$$

$$< T_\Delta, x_i >::= \{(x_i \in T_\Delta) \wedge (T_\Delta \in O_2)\}$$

$$< MTV >::= \{(\exists SE_\Delta \leftarrow O_1) \leftrightarrow (\exists TE_\Delta \leftarrow O_2) \cap (x \mid \text{x is a threshold value})\}$$

$$< R >::= \{(\exists SE_\Delta \leftarrow O_1) \leftrightarrow (\exists TE_\Delta \leftarrow O_2) \cap (x \mid \text{x is a relationship between SE and TE})\}$$

```
<PRM>::= StringPRM | ChildPRM | LabelPRM | PropertyPRM |
         OverlapPRM | CustomizedPRM | SynonymPRM

<LogicGM>::= <Logic1> <Logic2> ... <LogicN>
```

$$< Logic1 >::= TE \leftarrow SE$$

$$< Logic2 >::= \{\{TE \cap \{\exists TE.attribute \wedge (TE.attribute \geq 1)\}\} \leftarrow SE\}$$

$$< Logic3 >::= \{TE \leftarrow \{SE \cap \{\exists SE.attribute \wedge (SE.attribute \geq 1)\}\}\}$$

$$< Logic4 >::= \{\{TE \cap \{\exists TEChild \subseteq TE \wedge (TEChild \geq 1)\}\} \leftarrow SE\}$$

$$< Logic5 >::= \{TE \leftarrow SE \cap \{\exists SEChild \subseteq SE \wedge (SEChild \geq 1)\}\}$$

$$< Logic6 >::= \{TE \leftarrow SE \cap \{(\exists SEChild \subseteq SE) \vee (\exists SE.attribute)\}\}$$

$$< Logic7 >::= \{TE \cap \{(\exists TEChild \subseteq TE) \vee (\exists TE.attribute)\} \leftarrow SE\}$$

$$< Logic8 >::= \{TE \cap \{(\exists TEChild \subseteq TE) \vee (\exists TE.attribute)\}\} \leftarrow \{SE \cap \{(\exists SEChild \subseteq SE) \vee (\exists SE.attribute)\}\}$$

$$< Logic9 >::= \{TE \cap \{\exists TEChild \subseteq TE \wedge (TEChild \geq 1)\}\} \leftarrow \{SE \cap \{\exists SE.attribute\}\}$$

$$< Logic10 >::= \{TE \cap \{\exists TE.attribute \wedge (TE.attribute \geq 1)\}\} \leftarrow \{SE \cap \{\exists SEChild \subseteq SE\}\}$$

```
<Logic>::= <LogicGM> <LogicCM>
```

The constructs $< CM >$ and $< LogicCM >$ are related with the customized mappings and are not covered in the scope of this paper therefore its BNF are not presented. The rules in $< LogicCM >$ are the same as that in the $< LogicGM >$ construct.

4.8 Summary

This chapter provided insight into the SPHeRe ontology matching system with its internal methodologies. Also, it introduces and elaborates the MBO, that is used to represent and store mappings generated in the form of generalized and customized mappings. Also, logic based on these mappings to be used for transformation is shown in this chapter. The role of object oriented and ontology design patterns are also explained for demonstrating the effective design of the system.

Chapter 5

Customized Mappings, Mappings Evolution and Mappings Execution

Personalized Medicine envisions bringing health-care stakeholders under a single platform to individualize prevention, diagnosis and treatment of a person's disease [98]. The goal of Health Information Exchange (HIE) system is to facilitate physicians to access and retrieve clinical data to provide safer, timely, efficient, effective, equitable and patient-centered care [99]. HIE refers to the technological network infrastructure, that has the chief purpose of assuring accurate medical information exchange" [100]. HIE systems tries to ensure realizing concept of personalized medicine; however, face major challenges of interoperability because of difficulty in integrating data from heterogeneous data sources. This objective of personalized medicine using HIE systems can be achieved by using Detailed Clinical Models (DCM). DCMs provides methodology for structuring medical information by combining expert knowledge, data specification and terminology and enables various technical applications [101]. DCM allows its use in multiple standards by making clinical data explicit [102]. Common DCM gives precise semantics and makes the task of mapping between models manageable [103].

DCM's effectiveness can be observed from the international collaboration initiative, *Clinical Information Modeling Initiative (CIMI)* that is "dedicated to provide a common format for detailed specifications for the representation of health information content so that semantically interoperable information may be created and shared in health records, messages and documents" [104]. CIMI team has so far agreed to create and use a single logical representation called the CIMI core reference model, comprising one or more models as the basis for interoperability across formalisms [105].

The envisioned CIMI core reference model considers the use of already existing HL7 V3 Reference Information Model (RIM) and openEHR information models. openEHR is a health-care standard based on two level modeling approach: a reference information model and clinical content representation in archetypes and templates form [106]. Archetypes and templates are the formal models of domain concepts controlling data structure and content of data [77]. Archetypes are the constraints based models of domain content expressed in a formal language called Archetype Definition Language (ADL) [78]. Another clinical standard that specifies the structure and semantics of clinical documents for the purpose of exchange is called HL7 CDA [107]. HL7 CDA is based on RIM to generate clinical documents. RIM is the critical component of the development process of HL7 messages and is root of all the information models. It consists of backbone classes, their specialization and structural attributes for further defining the roles of the classes [108]. The core classes includes: Act, Entity, Role, Act Relationship and Participation. These core classes are further divided into sub-classes used for the generation of HL7 messages. Both openEHR and HL7 standards are based on reference models that contains similarities and differences between them. This would require consideration of concepts of both the models to be made part of a single generic model, easily adaptable for both standards. CIMI group's vision of a single model is handled by using ontology mappings in our work [109]. In the previous chapter, we discussed Generalized Mappings generated and stored using pattern relationship model in the MBO. Although these mappings achieve accuracy to a certain level, they still lack completeness at instance level transformations. This is because health-care organizations conform to concepts of different standards based on their information requirement. The problem with Generalized Mappings is not handling the organization conformance issues and therefore lacks in achieving interoperability. P-DCM encouraged us to extend generalized mapping solution to customized mappings in order to enhance efficiency of the system and provide more accurate mapping and minimize the role of human experts. Therefore, the concept of P-DCM is useful in achieving true semantic data interoperability among HIE's.

Generalized mappings generated using SPHeRe and customized mappings generated using P-DCM can have conflict due to organizations conformance to particular concepts. Mechanism is required to ensure the organization conformance issues are handled and the changes are properly propagated in the MBO to evolve generalized mappings with customized mappings. Continuity of mappings make sure that the conflicts between generalized and customized mappings are resolved by maintaining the accuracy of the mappings. Also, these mappings are stored in the logic part of the MBO for transformation. We have done preliminary evaluation for continuity of mappings by introducing a self evolutionary rule based scheme [110].

Transformation is based on the generalized and customized logic for conversion of one standard format to another. Initially, the customized logic in invoked and if the logic is not available then generalized logic is processed. The detailed description of all this process is provided in this chapter.

5.1 Customized Mappings Generation Process

5.1.1 P-DCM and Semantic Stack

P-DCM behaves as a centralized entity, easily used by health-care standards according to their structure and format. The association of P-DCM concepts with different standard concepts creates mapping relationship and results in customized mapping generation. Mapping relationships based on P-DCM are used when service is unable to transform particular concepts of source instance into target instance. We are using P-DCM's to achieve true semantic data interoperability among HIE's compliant to heterogeneous standards. For the proof of concept, we developed P-DCM's for diabetes and alzheimer's patient data, which are presented in the Result section. Diabetes P-DCM is based on data of 100 patients from Hospital Management Information System (HMIS) of local hospital in Korea. Whereas, Alzheimer's P-DCM is based on alzheimer's patient data gathered from a Home Health-care Monitoring System (HHMS). Our work [111] describes alzheimer's patient data transformation into heterogeneous standards. Instances of openEHR (called extracts) and CDA are derived from *P-DCM*. This mapping information is stored in the mapping file.

The proposed approach provides data interoperability between HIE's that are compliant to clinical standards. Data interoperability is based on mappings that requires higher level of accuracy. Initially, we focussed on accuracy using ontology matching techniques to generate generalized mappings by using clinical standards information. The level of accuracy was not appropriate for information to be exchanged seamlessly among HIE's, therefore we propose P-DCM approach as our second step. Our P-DCM approach achieves high level of accuracy by considering organizational information in mapping process in the form of customized mappings. This leads to resolving heterogeneities among clinical standards and ensures seamless communication between HIE's.

5.1.1.1 Semantic Interoperability Artifacts

Semantic interoperability is one of the vital challenges faced by health-care community. The concept of Semantic Stack, elaborated in [2], deals with the semantic interoperability artifacts. The stack deals with data, information and clinical pathways necessary for documentation of patient record [2]. The stack defines different models that include Model of Knowledge (ontologies), Model of Meaning (vocabularies), Model of Use (DCM or Archetypes) and Model of Syntaxes (Archetype Object Model (AOM)) and Model of Documentation, Archiving and Exchange. Although the semantic stack covers most of the artifacts necessary for semantic interoperability, it still lacks a model that can ensure complete, accurate and true semantic interoperability that considers organization clinical model and its representation. Conformance claim plays an important role for semantic interoperability, where health-care organizations represent clinical concepts by using specific set of concepts of different health-care standard formats based on their requirements. For example, HL7 provides artifact *Interactions*, that are used for communicating HL7 messages between medical systems. Each domain (laboratory, patient administration and others) in HL7 specifications provides specific set of interactions for exchange of messages. Laboratory domain interactions includes *Order, Promise and Result* interactions, each having further set of interactions. Organizations can conform based on their requirements to these interactions, such as, conforming only to set of Order and Result interactions. Therefore, a model is necessary that focuses on organization conformance claims to complete the semantic stack. We call this model, "Model of Purpose" and it deals with the concept of Personalized-DCM. To increase the effectiveness of using all the models, personalization concept is necessary. Model of Purpose is related with all the other models in Fig. 5.1.

In summary, health-care standards are striving hard and provide the base for interoperability; however, achieving interoperability at data level is still a bottleneck for integration of health-care

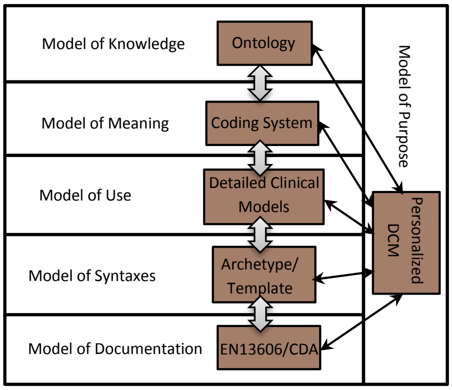


Figure 5.1: Modified Semantic Stack (of [2]) with P-DCM

systems compliant to heterogeneous health-care standards. Each standard defines clinical concepts that are not understandable outside the scope of that standard [112]. Ontology mappings is one of the method to integrate these heterogeneous standards but lacks application because of organizations conformance to artifacts based on their requirements. Model of Knowledge uses ontology for resolving heterogeneities but limited to handling generalized mappings. For example, HL7 RIM ontology[1], openEHR ontology[2], SNOMED CT ontology[3] are some of the health-care standard ontologies, that can be mapped to one another resulting in generation of generalized mappings. Generalized mappings include mappings between standard ontologies based on their reference models. HL7 RIM ontology and openEHR ontology matching results in generation of generalized mappings (Entity concept of HL7 RIM ontology mapped with Actor concept of openEHR ontology). Model of Purpose is related to Model of Knowledge as it is based on P-DCM to generate customized mappings for increase mappings accuracy level with generalized mappings. Generalized mappings are based on generic approach of matching multiple standard ontologies using ontology matching techniques, while customized mappings are based on specific approach of matching particular information model reflecting organization's requirement by relating them with concepts of multiple standards. Therefore, distinction between generalized and customized mappings is the addition of P-DCM concept in customized mappings for higher level of accuracy.

[1]http://www.w3.org/wiki/images/2/2e/HCLS$$ClinicalObservationsInteroperability$$HL7CDA2OWL.html$notes.html
[2]http://trajano.us.es/~isabel/EHR/
[3]http://krono.act.uji.es/people/Ernesto/umlsassessment/SNOMED-ontology.zip/view

For example, mapping of Observation concept of openEHR standard with *Observation* concept of HL7 standard is generalized mapping, while *Complication* (P-DCM concept) concept used to identify the problem of patient in health-care organization can be mapped with *Observation* concept of HL7 and *Section* concept of openEHR standard. Therefore, communication between two HIE's compliant to openEHR and HL7 standard would be accurate if customized mappings are used to interpret Complication concept in these standards rather than generalized mappings. Model of Meaning provides standard based clinical terminologies information;few of the standard vocabularies includes SNOMED CT[4], Mesh[5] and ICD-10[6]. Model of Use provides the initial platform for formulating the clinical information to Model of Purpose. It provides generic structure to clinical concepts in the form of detailed clinical models (DCM). Blood Pressure DCM and Cholesterol DCM are few of the examples of Model of Use. Model of Syntaxes is based on specific standard based structure/ format for representing clinical concepts. openEHR standard archetypes is one of the example of representing clinical concepts. The specifications of how the clinical information should be modeled in standard format is represented by Model of Documentation. Examples includes HL7 CDA and CEN 13606 standards specification. Model of Purpose is also related to Model of Syntaxes and Model of Documentation for conformance purpose. *Conformance* to health-care standard artifacts by organizations plays key role in data interoperability. For example, HL7 Laboratory domain provides set of interactions divided into three categories: *Order, Promise and Result*. Consider an organization conform to *Order* and *Result* interactions and not *Promise* interactions, requiring these interactions binding with clinical concepts in some model. Model of Purpose relates clinical concepts with each interaction in *Order* and *Result* interactions for communication. Each time the interaction takes place, the associated messages containing specified data and concepts would be communicated. *Test Order* message is conformed to one of *Order* interactions modeled in Model of Purpose, ensures that test order message is always communicated by order interaction. On the other hand in the current scenario the organization is not conformed to Promise interactions as per its requirements, therefore these interactions will not take place. In this way, Model of Purpose relates with other models to ensure data interoperability.

[4]http://www.ihtsdo.org/snomed-ct/
[5]http://www.nlm.nih.gov/mesh/
[6]http://apps.who.int/classifications/icd10/browse/2010/en

5.1.2 P-DCM Methodology

DCM represents knowledge about the clinical concepts, and when they are designed based on interpretation of an organization, it becomes Personalized-DCM. The information of P-DCM requires representation in standard formats for sharing, paving way for standards like HL7 and openEHR. The bonding of P-DCM with these standards defines the level of data interoperability.

5.1.2.1 Customized Mapping Module

Building Personalized-Detailed Clinical Model (P-DCM)

Building P-DCM requires clinical concepts (such as diabetes) understanding and also how a clinical concept is formulated for data capturing by HIE's (such as Subjective, Objective, Assessment, Plan (SOAP) method of documentation). Due to these constraints, organizations can conform to particular concepts of different standards for creation of Electronic Health Record (EHR) document. This adds the personalization aspect and thus requires customized mappings as a solution. P-DCM development follows three steps: identifying the clinical concept, identifying the structure of clinical concepts recording by HIE and lastly, building P-DCM based on the clinical concept and the way its information is structured. Firstly, each clinical concept is made up of some constructs, therefore identification of clinical concept is necessary such as: diabetes, alzheimer or cancer. This analysis would result in identification of all the related observations, history, medication and recommendations in the domain of particular clinical concept. Secondly, clinical concept information documentation is identified that provides the baseline for modeling P-DCM. SOAP, Data, Assessment and Plan (DAP), Functional Outcomes Reporting (FOR) and narrative notes are few of the examples for structuring the notes about medical records of a patient. Lastly, modeling P-DCM based on the information obtained from the first two steps is performed. This is performed by modeling the clinical concept information into classes, attributes and data-types. Modeling P-DCM allows easy association of its concepts with clinical standards concept, thus resulting in generation of customized mappings.

In order to analyze the development of P-DCM, we gathered about 100 diabetic patients' data from local hospital in, Korea, which includes 50 patients each of type 1 and type 2 diabetes. Also we gathered information about Alzheimer patients by monitoring their daily life activities using sensory devices. We developed *P-DCM* for type 2 diabetic patients portion, shown in Table 5.1, and for Alzheimer patient, shown in Table 5.4.

Table 5.1: HMIS based P-DCM and Clinical Information

HMIS based P-DCM	**Clinical Information**
Package DiabetesMellitus_Type2_V1 imports ClinicalInformation imports Datatypes class Subjective Complaint: Observations Allergies: Observations Medications: Medication ... class Objective ClinicalObservations: Observation ... class Assessment Complications: Observation ... class Plan Exercise: Observation Medication: Observation ... EndPackage	Package ClinicalInformation imports Datatypes class Observation Title: SET Value: Any Text: ED Device: SC CWE CodingSystem: CD ObservationValue: PQ ... Class Medication Title: SET DoseQuantity: IVA<PQ> Notes: ED ... EndPackage

Conformance Manager

Conformance Manager handles the association of P-DCM concepts with openEHR and HL7 CDA concepts. The conformance is based on HIE requirement to represent P-DCM concepts with specific concepts of both standards. With each and every entity, attribute, data-type or unit of openEHR and HL7 CDA template, a concept of *P-DCM* is associated. As an example, P-DCM's concept RELATIONSHIP conforms to PARTY_RELATIONSHIP and RELATIONSHIP_LINK concepts of openEHR and HL7 CDA concept respectively.

Annotated Info Extractor

Annotated Info Extractor generates rules in the form of *Customized Mappings* in addition to *Generalized Mappings*. *Customized Mappings* are the mappings based on P-DCM concepts and its relationship with standards concepts while *Generalized Mappings* are generated by matching ontologies of both the standards. Each concept of *P-DCM* is conformed to particular concepts of CDA and openEHR. Therefore concepts of CDA and openEHR bridged by *P-DCM* concept are similar in relationship. For example, x is RELATIONSHIP con-

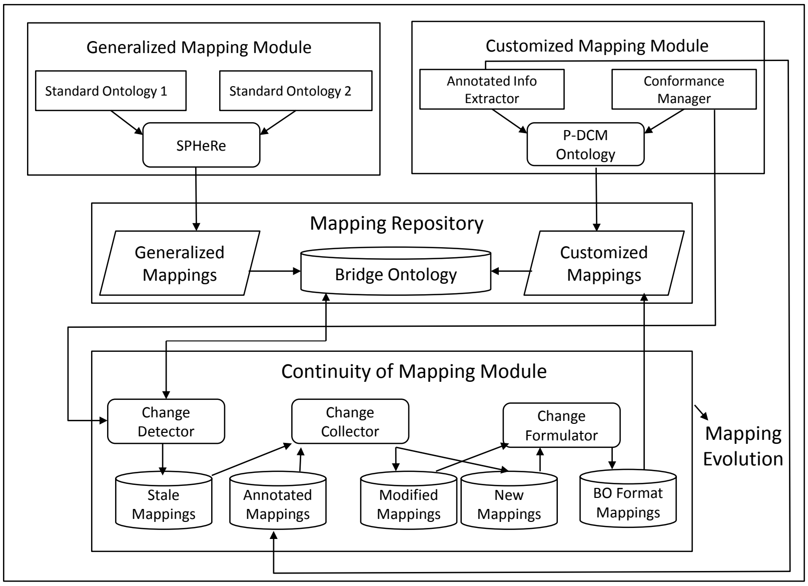


Figure 5.2: Personalized-Detailed Clinical Model Architecture

cept of P-DCM, y is PARTY_RELATIONSHIP concept of openEHR demographics model, while z is RELATIONSHIP_LINK concept of HL7 CDA Reference Information Model (RIM). PARTY_RELATIONSHIP and RELATIONSHIP_LINK concepts are derived from RELATIONSHIP concept, therefore both standard concepts are similar and can be mapped in instance transformation. To derive concepts from P-DCM means to associate P-DCM concepts with different standard concepts for its representation in clinical document generation for exchange of information. This information is further stored in the MBO.

The annotations are extracted based on all the conformed concepts CC_i from the P-DCM ontology. These annotations are stored in the Annotated Mappings repository as shown in Fig. 5.2. The CC_i concepts are related to the source standard ontology and for every CC_i their is annotated information in P-DCM ontology of target standard ontology concepts CC_j. These are stored in the repository for further processing by the change management modules. Algorithm 5.1.1 explains the whole procedure of finding the annotated mappings from P-DCM ontology.

Algorithm 5.1.1: ANNOTATED MAPPINGS(CC_i)

for $CC_i \ in \ PDCMO$

$/* where \ PDCMO \ is \ the \ Personalized \ Detailed \ Clinical \ Model \ Ontology */$

$$\textbf{do} \begin{cases} Annotations \equiv (PDCMC \longleftarrow CC_i \wedge CC_j) \\ /* where \ PDCMC \ is \ Personalized \ Detailed \ Clinical \ Model \ Concept, \\ CC_i \ is \ Conformed \ Concept \ of \ Source \ Standard \ Ontology, \\ CC_j \ is \ Conformed \ Concept \ of \ Target \ Standard \ Ontology. */ \\ AnnotatedMappings \longleftarrow getAnnotatedMappings(PDCMO) \end{cases}$$

$STORE \ AnnotatedMappings$

return $(AnnotatedMappings)$

5.2 Mappings Evolution

The customized mappings generated can have conflict with the existing generalized mappings stored therefore, a mechanism is required to remove redundancy and resolve conflicting mappings. Continuity of mappings provide such type of resolution and change management strategies.

Continuity of Mapping Module

Continuity of mappings extracts the annotation information from the P-DCM ontology and then utilizes it for the change detection, change collection and change formulation process.

5.2.1 Change Detection

Change Detector listens for non-conformed concepts information from the *Conformance Manager*. These non conformed concepts are identified in the MBO's generalized mappings and are stored in the Stale Mapping repository. These are then processed by the change collector module. The process is as follows:

- Non-conformed concepts NCC_j are accessed from *Conformance Manager* CM.
- The change detector CD identifies mappings based on non-conformed concepts in the *MBO* and stored as stale mappings SM.
- A single stale mapping is composed of a conformed concept aligned with a non-conformed concept.

The whole procedure of change detection is explained in Algorithm 1. Change Detection relates the customized with the generalized mappings by finding out the redundancy caused by the organization conformance.

Algorithm 1: Change Detector Algorithm

Input: CC_i, NCC_j /*where CC_i, NCC_j are Conformed Concepts and Non-Conformed Concepts respectively*/

Output: $M_{(CC_i \longleftrightarrow NCC_j)}$ /*where $M_{(CC_i \longleftrightarrow NCC_j)}$ are the mappings in the Mediation Bridge Ontology between Conformed and Non-Conformed Concepts*/

begin
    Conformance Manager: CM
    $CM \longleftarrow CC_i \wedge NCC_j$
    Change Detector: CD
    Mediation Bridge Ontology: MBO
    Stale Mappings: SM
    $CD \longleftarrow$ getNonConformedConcepts(CM)
    while NCC_j *in* CD **do**
        $CD \longleftarrow$ Find($M_{(CC_i \longleftrightarrow NCC_j)}$ in MBO)
        if $\exists\ M_{(CC_i \longleftrightarrow NCC_j)}$ **then**
            $SM \longleftarrow$ LABEL $M_{(CC_i \longleftrightarrow NCC_j)} \equiv$ Deferred
            STORE SM
        end
        else
            Goto 8
        end
    end
end

5.2.2 Change Collection

Change Collector accesses stale mappings identified by *Change Detector*. Organization conformance information can lead to two kinds of evolution of mappings: Modification and Addition. Both of these mappings are handled by the Change Collector module. The process for both categories is described as follows:

- Conformed concepts are collected from the stale mappings.
- Annotated mappings based on the conformed concept are searched in the *P-DCM ontology*. A corresponding conformed concept of the target standard concept is identified from the annotated mapping information. For example, if the conformed concept A belongs to standard 1 and non-conformed concept B belongs to standard 2, then based on A its corresponding conformed concept C of standard 2 is accessed from *P-DCM ontology* and an annotated mapping is generated.
- The change collector modification algorithm is described by Algorithm 2. Initially, the conformed concept CC_i is obtained from stale mappings SM using getConformedConcepts(SM) function.
- Based on the conformed concept CC_i, the change collector ChC collects the annotated mappings information from the repository using annotatedMappings(CC_i) function.
- The corresponding conformed concept CC_j is found in the annotated mapping using FindCorrespondingAnnotatedMappings(ChC).
- The corresponding conformed concept is obtained and stored as modified mappings $\triangle M_{(CC_i \longleftrightarrow NCC_j)}$ in the modified mappings repository.

Algorithm 2: Change Collector (Modification) Algorithm

Input: SM /*where SM reflects Stale Mappings from Change Detector*/

Output: $\triangle M_{(CC_i \longleftrightarrow NCC_j)}$ /*where $\triangle M_{(CC_i \longleftrightarrow NCC_j)}$ are the new modified mappings added in the Mediation Bridge Ontology*/

```
begin
    Change Collector: ChC
    Target Annotated Mapping: TAM
    Target Standard Ontology Conformed Concept: CC_j
    Modified Mappings: MM
    CC_i ⟵ getConformedConcepts(SM)
    ChC ⟵ annotatedMappings(CC_i)
    for CC_i in SM do
        TAM ⟵ FindCorrespondingAnnotatedMappings(ChC)
        CC_j ⟵ TAM.getTargetConformedConcept(CC_j)
        GENERATE △M_(CC_i ⟷ NCC_j)
    end
    MM ⟵ STORE △M_(CC_i ⟷ NCC_j)
end
```

- *P-DCM ontology* annotated information are also searched to find out some new mappings that didn't existed before for storing in the *MBO*. For example, concept X and Y of standard 1 and 2 are related with clinical model concept Z in *P-DCM ontology* which is the annotated information. Concept X and Y are candidate alignment in this case, therefore if this alignment is not already present in the *MBO* then it is a new candidate mapping. Algorithm 3 describes the overall procedure for the change collector addition.

- The annotated mappings AM stored in the annotated mappings repository consists of source conformed concept SCC_i and target conformed concept TCC_j.

- MBO is searched using Search(MBO) to find out whether this annotated mappings exists or not.

- If the annotated mappings is missing in the MBO then this new mapping is is generated uusing GENERATE $NAM_{(SCC_i \longleftrightarrow TCC_j)}$ and stored in the new mappings repository.

Algorithm 3: Change Collector (Addition) Algorithm

Input: $PDCMO$ /*where $PDCMO$ represents Personalized Detailed Clinical Model Ontology*/

Output: $NAM_{(CC_i \longleftrightarrow CC_j)}$ /*where $NAM_{(CC_i \longleftrightarrow CC_j)}$ are the New Annotated Mappings added in the Mediation Bridge Ontology*/

```
begin
    Annotated Mappings: AM
    New Annotated Mappings: NAM
    Source Conformed Concept: SCC_i
    Target Conformed Concept: TCC_j
    Personalized Detailed Clinical Model Ontology: PDCMO
    Mediation Bridge Ontology: MBO
    AM_i ⟵ SCC_i ≡ TCC_j
    while AM_i ∈ PDCMO do
        Search(MBO)
        if AM_i ≠ M_(SCC_i ⟷ TCC_j) then
            GENERATE NAM_(SCC_i ⟷ TCC_j)
        end
        else
            Goto 9
        end
    end
    NAM ⟵ STORE NAM_(SCC_i ⟷ TCC_j)
end
```

5.2.3 Change Formulation

Change Formulator collects the modified and new mappings identified by *Change Collector* and formulates it in *MBO* compatible format for storing.

- Modified mappings MM and new annotated mappings NAM are stored in the MBO with PopulateMBO(C_i, C_j, R, $Thresh$, $bridge$) method. These mappings are categorized as customized mappings and their logic is created in $LogicCM$ to be used for transformation from one standard format to another.

Algorithm 4: Change Formulator Algorithm

Input: MM, NAM /*where MM and NAM represents Modified Mappings and New Annotated Mappings respectively*/

Output: /*Customized Mapping Result*/

```
begin
    for MM_i ∈ MM do
        bridge := "Customized Matching Bridge"
        C_i := Source Ontology Concept that belongs to MM_i
        C_j := Target Ontology Concept that belongs to MM_i
        Thresh := Threshold Value
        R := Relationship between C_i and C_j
        PopulateMBO(C_i, C_j, R, Thresh, bridge)
    end
    for NAM_i ∈ NAM do
        bridge := "Customized Matching Bridge"
        C_i := Source Ontology Concept that belongs to NAM_i
        C_j := Target Ontology Concept that belongs to NAM_i
        Thresh := Threshold Value
        R := Relationship between C_i and C_j
        PopulateMBO(C_i, C_j, R, Thresh, bridge)
    end
end
```

The customized and generalized mappings in the MBO are used for transformation purpose among different standards. The priority for customized mappings is more than generalized mappings as it is more conformance based, therefore, $LogicCM$ is processed before $LogicGM$ for transformation purpose.

5.3 Mappings Execution

The process of instance level transformation among different standards is described using algorithms in this section. Algorithm 5 explains the procedure of source standard instance SMR_SA conversion to target standard instance TMR_SB. The conversion is based on loading the patterns stored in the *MBO* using *LoadMBO()* method. Initially, the target instance is empty TMR_{SB} = $\emptyset$ and the concepts from source instance is retrieved SMR_{SA}.*RetreiveConceptsList()* for instance transformation. Mapping of these retrieved concepts with target standard concept is performed by *RetrieveMappings(*$ListConceptsSMR_{SA}$*, Logic)* method that is described in Algorithm 6.

Algorithm 5: Heterogeneous Medical Standards Transformation

Input: SMR_{SA} /*MR $\rightarrow$ Source Medical Record & SA$\rightarrow$ $Standard'A'$*/

Output: TMR_{SB} /*MR $\rightarrow$ Target Medical Record & SB$\rightarrow$ $Standard'B'$*/

begin
- Let MBO := LoadMBO() /* where MBO is Mediation Bridge Ontology */
- Let TMR_{SB} = $\emptyset$
- $ListConceptsSMR_{SA} \leftarrow SMR_{SA}$.RetreiveConceptsList()
- $ListMappedConceptsTMR_{SB} \leftarrow$ RetrieveMappings($ListConceptsSMR_{SA}$, $Logic$)
- $ListConceptsTMR_{SB} \leftarrow TMR_{SB}$.InsertConceptList($ListMappedConceptsTMR_{SB}$)
- Let $SC_t \leftarrow ListMappedConceptsTMR_{SB}$.first()
- **for** $SC_i \epsilon ListConceptsSMR_{SA}$ **do**
  - SC_t.Value $\equiv$ SC_i.Value
  - $SC_t \leftarrow ListMappedConceptsTMR_{SB}$.next()
- **end**
- **return** TMR_{SB}

end

Algorithm 6 executes until all the corresponding concepts and properties of the source standard are identified from the *MBO* using $Logic$. Initially the SC_i counterpart should be obtained from $LogicCM$ using Execute($LogicCM$,SC_i) and the target concept TCP_j is to be added to $ListConceptsTMR_{SB}$. On the other hand if the source concept SC_i is not available in $LogicCM$ then $LogicGM$ should be processed using Execute($LogicGM$,SC_i) to find TCP_j for addition into $ListConceptsTMR_{SB}$.

Algorithm 5 utilizes the mapped concepts $ListConceptsTMR_{SB}$ from Algorithm 6 and insert it into TMR_{SB} using TMR_{SB}.*InsertConceptList(*$ListMappedConceptsTMR_{SB}$*)* method. The values are finally assigned to all the matched concepts using SC_t.Value $\equiv$ SC_i. Value statement and the target standard instance TMR_{SB} is generated. The transformation patterns with the help of CDA and vMR standards case study is demonstrated in the next section.

Algorithm 6: Logic Execution for Standards Structure Mapping

Input: SC_i, $LogicGM$, $LogicCM$
Output: $ListConceptsTMR_{SB}$

begin
  $Logic \leftarrow LogicGM \wedge LogicCM$
  Let SC_i is Source Ontology Concept
  $LogicGM$ is the Logic for generalized mappings
  $LogicCM$ is the Logic for customized mappings
  while $SC_i \leftarrow ListConceptsSMR_{SA}$*.next()* **do**
    if $\exists$ $SC_i \in$ *(*SC_i*.SourceEntity* $\vee$ SC_i*.SourceEntity_attributes* $\vee$ SC_i*.SourceEntity_child)* $\in$ $LogicCM$ **then**
      $TCP_j \leftarrow$ Execute($LogicCM$,SC_i)
    end
    else if $\exists$ $SC_i \in$ *(*SC_i*.SourceEntity* $\vee$ SC_i*.SourceEntity_attributes* $\vee$ SC_i*.SourceEntity_child)*$\in$ $LogicGM$ **then**
      $TCP_j \leftarrow$ Execute($LogicGM$,SC_i)
    end
  end
  $ListConceptsTMR_{SB}$.add(TC_j)
  return $ListConceptsTMR_{SB}$
end

5.3.1 HL7 CDA and vMR Standards Conversion Process

Standards follow structured format based on reference model that provides clearly defined concepts for interoperable communication among HISs. Their transformation mechanism among each other using *MBO* is described in detail in this section. We take the case of CDA and vMR standards as discussed previously.

VMR is a medical record therefore most of the linkage of information in this format is with the structured body portion of CDA. Table 5.2 shows CDA and vMR snippet code based on patient's systolic and diastolic blood pressure record in standard based structured format. The proposed system is responsible for providing gateway to HISs for exchanging meaningful information with each other by standard format transformation. It converts HMIS compatible CDA format to CDSS compatible vMR format and vice versa for processing of information.

Table 5.2: HL7 CDA and vMR Systolic and Diastolic Blood Pressure Observation

CDA Snippet (a)

```xml
<component typeCode="COMP" contextConductionInd="true">
 <section classCode="DOCSECT" moodCode="EVN">
  <entry typeCode="COMP" contextConductionInd="true">
   <observation classCode="OBS" moodCode="EVN">
    <code code="ObservationType"
    codeSystem="getCodeSystem"/>
    <entryRelationship typeCode="CAUS"
    contextConductionInd="true">
     <observation classCode="OBS" moodCode="EVN">
      <code code="271649006" displayName="Systolic BP"
      codeSystem="2.16.840.1.113883.6.96"
      codeSystemName="SNOMED CT"/>
      <effectiveTime xsi:type="IVL_TS">
    <low inclusive="true" value="20080220102200+0300"/>
    <high inclusive="true" value="20080220102200+0300"/>
      </effectiveTime>
     <value xsi:type="PQ" value="127" unit="mm[Hg]"/>
     </observation>
    </entryRelationship>
    <entryRelationship typeCode="COMP"
    contextConductionInd="true">
     <observation classCode="OBS" moodCode="EVN">
  <code code="271650006" displayName="Diastolic BP"
  codeSystem="2.16.840.1.113883.6.96"
      codeSystemName="SNOMED CT"/>
      <effectiveTime xsi:type="IVL_TS">
   <low inclusive="true" value="20080220102200+0300"/>
   <high inclusive="true" value="20080220102200+0300"/>
      </effectiveTime>
    <value xsi:type="PQ" value="79" unit="mm[Hg]"/>
     </observation>
    </entryRelationship>
    </observation>
  </entry>
 </section>
</component>
```

vMR Snippet (b)

```xml
<observationResults>
 <observationResult>
  <templateId root="2.16.840.1.113883.10.20.1.32"/>
  <id root="107c2dc0-67a5-11db-bd13-0800200c9a66"/>
    <relatedClinicalStatement>
 <observationResult>
      <templateId root="2.16.840.1.113883.10.20.1.31"/>
      <id root="33d27880-eb74-11e0-9572-0800200c9a66"/>
      <observationFocus displayName=" Systolic BP"
      codeSystem="2.16.840.1.113883.6.96"
      code=" 271649006"/>
      <observationEventTime low=" 20080220102200+0300"
      high=" 20080220102200+0300"/>
      <observationValue>
       <physicalQuantity value=" 127 " unit=mm[Hg]"/>
      </observationValue>
 </observationResult>
    </relatedClinicalStatement>
    <relatedClinicalStatement>
  <observationResult>
       <templateId root="2.16.840.1.113883.10.20.1.31"/>
       <id root="33d27880-eb74-11e0-9572-0800200c9a66"/>
       <observationFocus displayName=" Diastolic BP
       codeSystem="2.16.840.1.113883.6.96"
       code=" 271650006"/>
       <observationEventTime low=" 20080220102200+0300"
       high=" 20080220102200+0300"/>
       <observationValue>
       <physicalQuantity value=" 79 " unit=mm[Hg] "/>
       </observationValue>
  </observationResult>
    </relatedClinicalStatement>
 </observationResult>
</observationResults>
```

Table 5.2 (a) shows structured body portion of CDA by using **Observation** class to record patient blood pressure. The main classes used for recording this observation are **Component, Section, Entry, Observation, and Entry Relationship**. **Observation** class records systolic and diastolic blood pressure and relates with each other through **Entry Relationship** class. In the same way, to model the same information in vMR requires use of **ObservationResult** class. The main classes therefore used in vMR format are **ObservationResults, ObservationResult, and RelatedClinicalStatement** as shown in Table 5.2 (b). **ObservationResult** class maps to **Observation** class as it records systolic and diastolic blood pressure results and **RelatedClinicalStatement** class behaves as **Entry Relationship** class to relate results information. We describe the realization of alignment and transformation patterns required for conversion between these standards by carrying forward the scenario described in this section.

Ontology alignment patterns behave as reusable templates of recurring correspondences and plays vital role in achieving interoperability [49]. The patterns for conversion are divided into three main categories: Structural, Sequence and Data Transformations. Initially, output structure is defined for the conversion format based on the input format. Structural transformation is necessary because every standard has its own structural format derived from the reference model by applying development rules; therefore it is necessary for correct parsing of the document.

5.3.2 HL7 and openEHR Standards Case Studies

HL7 CDA and vMR are both HL7 standards, therefore these follow the same reference model. This section explains HL7 and openEHR standards transformation using SPHeRe and P-DCM mappings. HL7 and openEHR standards follows different reference models and their association with P-DCM concepts is also explained in this section. To demonstrate the approach, we present two case studies of HMIS (related to Diabetes) and HHMS (related to Alzheimer) based P-DCM. Initially, P-DCM is defined, that models the clinical concept in a format that satisfies the information and structure an organization intends for representation. The next step is association of P-DCM concepts with standard concepts, hence generating customized mappings.

We showed these customized mappings and how these can help with generalized mappings in generating different standard formats. A simple example is health-care organization using these mappings to share or exchange the same information with two different health-care organizations, having one compliant to HL7 and other to openEHR standard. To further illustrate, we discuss two case studies in the following sub-sections:

5.3.2.1 Case Study A: HMIS based Personalized DCM

For proof of concept, we created instances from the *P-DCM* and Clinical Information as described in Table 5.1. *P-DCM* is based on Subjective, Objective, Assessment, and Plan (SOAP) pattern, adopted by Saint Mary's Hospital staff to record diabetes patients' data at different encounters. SOAP is a method of documenting patient related data in sections. The instances in HL7 CDA and openEHR standard are thus derived from *P-DCM*. Fig. 5.3 and Fig. 5.4 show portions of CDA and openEHR standards instances respectively, recording the blood pressure of the diabetic patient.

```
<section>
  <code code="170558000" codeSystem="2.16.840.1.113883.6.96"
  codeSystemName="SNOMED-CT06" displayName="Chronic disease -
  follow-up assessment (finding)">
     <entry typeCode="COMP">
        <observation classCode="OBS" moodCode="EVN">
           <code code="44054006"
           codeSystem="2.16.840.1.113883.6.96"
           codeSystemName="SNOMED-CT06"
           displayName="Non-insulin dependent diabetes mellitus"/>
           <statusCode code="completed"/>
           <entryRelationship typeCode="COMP">
               <observation classCode="OBS" moodCode="EVN">
                  <code code="363237008" codeSystem="2.16.840.1.113883.6.96"
                  codeSystemName="SNOMED-CT06"
                  displayName="Neurological complication of procedure (disorder)"/>
                  <statusCode code="completed"/>
               </observation>
        </observation>
     </entry>
</section>
<section>
   <code code="260224007" codeSystem="2.16.840.1.113883.6.96"
   codeSystemName="SNOMED-CT06"
   displayName="Objective observation (qualifier value)">
      <entry typeCode="COMP">
          <organizer classCode="BATTERY" moodCode="EVN">
              <component typeCode="COMP">
                 <observation classCode="OBS" moodCode="EVN">
                    <code code="163020007"
                    codeSystem="2.16.840.1.113883.6.96"
                    codeSystemName="SNOMED-CT06"
                    displayName="On examination - blood pressure reading (finding)"/>
                    <statusCode code="completed"/>
                    <value xsi:type="PQ" value="" unit="mm[Hg]"/>
                 </observation>
              </component>
          </organizer>
          ......
```

Figure 5.3: CDA Instance of Type 2 Diabetes Mellitus Patient

A sample Ambulatory Encounter data of non-insulin-dependent diabetes mellitus with neurological complications patient is converted to both standards format based on *P-DCM*. SNOMED-CT terminologies are used as vocabularies for coding the data. SECTION class in both standards is used for categorizing related information. SOAP pattern creates section based on information like Objective section, used for categorizing observations. Therefore, *P-DCM* concept Objective and its attributes are categorized in sections in both openEHR and CDA standards. On the other hand, attributes of Objective concept of *P-DCM* are categorized under OBSERVATION class. Clinical information is represented by ClinicalInformation package while the contents of the clinical information are represented in *P-DCM*. ISO 20190 standard datatypes are used for modeling *P-DCM* and ClinicalInformation models. Interesting mapping information is obtained at attributes level, as shown in Fig. 5.3 and Fig. 5.4.

```
<data xsi:type="SECTION" archetype_node_id="at0001">
      <name>
            <value> Non-insulin-dependent diabetes mellitus
            </value>
      </name>
       <defining_code>
                    <terminology_id>
                    <value>SNOMED-CT06</value>
                    </terminology_id>
                    <code_string>44054006</code_string>
            </defining_code>
      <data xsi:type="SECTION" archetype_node_id="at0002">
            <name>
                  <value> Objective Observation </value>
            </name>
            <defining_code>
                    <terminology_id>
                    <value>SNOMED-CT06</value>
                    </terminology_id>
                    <code_string>260224007</code_string>
            </defining_code>
            <items xsi:type="OBSERVATION" archetype_node_id="at0003">
             <name>
                   <value>Blood Pressure</value>
             </name>
             <value xsi:type="DV_CODED_TEXT">
                    <value>On examination - blood pressure
                    reading (finding)</value>
                    <defining_code>
                                   <terminology_id>
                                    <value>SNOMED-CT06</value>
                                   </terminology_id>
                                   <code_string>163020007</code_string>
                    </defining_code>
             </value>
             <value xsi:type="DV_QUANTITY">
                    <value>mg</value>
             </value>
             </items>
...
</data>
```

Figure 5.4: openEHR Instance (Extract) of Type 2 Diabetes Mellitus Patient

CodingSystem attribute of Observation class in *P-DCM* is transformed with code attribute in CDA and defining code attribute in openEHR. Therefore, in CDA, code attribute of OBSERVATION class is mapped to defining code tag of openEHR extract. For blood pressure observation, both use SNOMED-CT vocabulary, information about it is stored in codeSystemName attribute of CDA, mapped to terminology_id attribute of openEHR. Some of the mappings of CDA and openEHR concepts with P-DCM concepts are shown in Table 5.3. In the same way, blood pressure is measured in ”mg” represented in P-DCM by *ObservationValue*, recorded by datatype Physical Quantity (PQ) in CDA and DV_Quantity datatype in openEHR, resulting in the *DCM Rule Engine* to generate rule discussed in Equation (1) earlier. The unit ”mg” from *P-DCM* annotates PQ and DV_Quantity by applying *DCM Rule Engine* transitiveOf rule and creates mapping between these data types. These mappings are then made part of the mapping file created in our previous work, resulting in interoperability service performing transformations. The resulting mapping instance

(3) is derived and stored in the mapping file. In mapping instance hasSourceEntity refers to HL7 CDA concept and hasTargetEntity refers to openEHR concept.

```
map:Mapping_Instance_3321
a                    map:Mapping ;
map:hasSourceEntity map:PQ;
map:hasTargetEntity map:DV_Quantity;
map:pdcmConcept map:ObservationValue;
map:hasSimilarityType   "=";
map:hasSimilarityValue  1.0;        ...(2)
```

Table 5.3: Content Mapping of Clinical Information to HL7 CDA and openEHR

P-DCM Concept	**Value**	**openEHR Class[Attribute]**	**HL7 RIM Core Class**	**HL7 Class[Attribute]**	**Terminology[Vocabulary Name]/Value**
Assessment	Chronic Disease Assessment	Composition	Act	Organizer	Chronic Disease-Follow Up Assessment [SNOMED CT]
Complication	Non Insulin Dependent Diabetes Mellitus	Section	Act	Observation	Non Insulin Dependent Diabetes Mellitus [SNOMED CT]
Objective	(objective)	Section	ActRelationship	Entry	Objective Observation [SNOMED CT]
Clinical Observation	Blood Pressure	Observation	Act	Observation	On Examination-Blood Pressure Reading (finding) [SNOMED CT]

5.3.2.2 Case Study B: Home Health-care Monitoring System (HHMS) based Personalized DCM

We performed experiments on Alzheimer's patient daily life activities monitoring as P-DCM case study in Home Health-care Monitoring Systems (HHMS). The baseline of this work is Human Activity Recognition Engine (HARE) [113], designed and developed by our lab for monitoring the activities of Alzheimer disease patients. HARE focuses on monitoring human activities (Alzheimer's patient as case study) using heterogeneous sensor technology. We extended this work in [111] by introducing health-care standards based sensory data exchange among HMIS and HHMS. P-DCM developed for this case study is shown in Table 5.4.

Table 5.4: HHMS based P-DCM

HHMS based Alzheimer's Patient P-DCM

```
Package Alzheimer_Patient_Activities_Monitoring_V1
imports Datatypes

class Activities
Activity: String
Type: String
SensorID: Integer
DetectedBy: String
ActivityName: String
Time: DateTime

EndPackage
```

```
<entry>
  <organizer classCode="OBS" moodCode="EVN">
    <code code="161108005" codeSystem="2.16.840.1.113883.6.96"
     codeSystemName="SNOMED CT" displayName="Alzheimer's disease
     society member (finding)">
      <statusCode code="completed"/>
        <component typeCode="COMP">
          <observation classCode="OBS" moodCode="EVN">
             <code code="184090004"
             codeSystem="2.16.840.1.113883.6.96"
             codeSystemName="SNOMED CT" displayName="Patient works away"/>
             <statusCode code="completed"/>
             <effectiveTime value="20110516010000"/>
             <value xsi:type="ST" value="Leaving Room"/>
               <performer typeCode = "PRF">
                 <assignedEntity classCode="ASSIGNED">
                   <id extension="1" root="2.16.840.1.113883.2.99.6.1"/>
                   <code code="408746007" codeSystem="2.16.840.1.113883.6.96"
                   codeSystemName="SNOMED CT" displayName="Sensor Device"/>
                     <Device classCode="DEV" determinerCode="INSTANCE">
                        <manufacturerModelName> Motion Sensor
                        </manufacturerModelName>
                     </Device>
                 </assignedEntity>
               </performer>
               <participant typeCode="LOC" contextControlCode="OP">
                 <participationRole classCode="ROL">
                   <code code="homeHealth" codeSystem="2.16.840.1.113883.5.111"
                   codeSystemName="HL7" displayName="Home Health"/>
                     <playingEntity classCode="ENT" determinerCode="INSTANCE">
                       <code code="224700005" codeSystem="2.16.840.1.113883.6.96"
                         codeSystemName="SNOMED CT" displayName="Bedroom"/>
                     </playingEntity>
                 </participationRole>
               </participant>
          </observation>
        </component>
       ......
    </organizer>
</entry>
```

Figure 5.5: Sample CDA of Patient Activity

The information about the patient is collected using sensor application. Each activity consists of its type that shows whether patient is moving, sleeping, eating or walking. These activities are identified by particular sensor or camera (example shows Motion sensors, Wearable sensors and 2D Camera). The sensors and cameras are provided with unique ID's. Date and time of the activity performed is also maintained in the repository. A threshold of 1 hour is set for the data to be accumulated and stored in HARE repository. This information is then transformed into CDA document and openEHR extract for communication with corresponding HMIS.


```
<content xsi:type="OBSERVATION" archetype_node_id="at0002">
        <name xsi:type="DV_CODED_TEXT">
                <value> Patient Works Away </value>
                <defining_code>
                            <terminology_id>
                            <value> SNOMED-CT06 </value>
                            </terminology_id>
                            <code_string> 184090004 </code_string>
                </defining_code>
        </name>
        <data archetype_node_id="at0001">
                <name>
                    <value> Leaving Bedroom </value>
                </name>
                <data xsi:type="ITEM_TREE" archetype_node_id="at0003">
                    <name>
                            <value> Activity </value>
                    </name>
                    <items xsi:type="ELEMENT" archetype_node_id="at0005">
                            <name>
                                <value> Motion Sensor </value>
                            </name>
                            <time>
                                    <value>20110516010000</value>
                            </time>
                            <value xsi:type="DV_CODED_TEXT">
                                    <value> Sensor Device </value>
                                    <defining_code>
                                        <terminology_id>
                                        <value> SNOMED-CT06 </value>
                                        </terminology_id>
                                        <code_string> 408746007 </code_string>
                                    </defining_code>
                            </value>
                    </items>
                    ......
        </data>
</content>
```


Figure 5.6: Sample openEHR Extract of Patient Activity

Mapping Alzheimer's Patient Activities to HL7 CDA and openEHR

The activities of Alzheimer's patients are stored in HARE repository. These activities are monitored in a home environment with the help of sensors and cameras, are part of the HHMS. We generated HL7 CDA document and openEHR extract from the activities information. HL7 CDA is generated from standard CDA RMIM and the coded values are derived from standard vocabularies like SNOMED CT and HL7. Sample CDA document generated is shown in Fig. 5.5. The same way, Fig. 5.6 shows sample openEHR extract from the activities. Table 5.5 summarizes the mappings of patient's activities information with HL7 CDA and openEHR. It shows the tags and values of activities mapping to HL7 CDA and openEHR classes and attributes. Patient leaving his bedroom is categorized as *Motion* in XML file, mapped with *Observation* class of both the standards. The same way *Motion Sensor* is mapped to *Device* class in HL7 CDA while *Element* class in openEHR.

Table 5.5: Content Mapping of Activities to HL7 CDA and openEHR

P-DCM Concept	**Value**	**openEHR Class[Attribute]**	**HL7 RIM Core Class**	**HL7 Class[Attribute]**	**Terminology[Vocabulary Name]/Value**
Activities	(all activities)	Composition	Act	Organizer	Alzheimer's disease society member [SNOMED CT]
Activity	(single activity)	Item_Tree	Act	Component	
Type	Motion	Observation	Act	Observation	Patient Works Away [SNOMED CT]
DetectedBy	Motion Sensor	Element	Participation	Performer	
			Role	AssignedEntity	Sensor Device [SNOMED CT]
			Entity	Device	Motion Sensor
ActivityName	Leaving Bedroom	Contact	Participation	Participant	
			Role	ParticipationRole	Home Health [HL7]
			Entity	PlayingEntity	Bedroom [SNOMED CT]
SensorID	1		Entity	AssignedEntity[id]	1
Time	2011:05:16:01:00:00	Item[time]	Act	Observation[effectiveTime]	20110516010000

5.4 Summary

This chapter explained the approach P-DCM used for generating the customized mappings. The mappings evolution process is described that resolves customized and generalized mappings differences to ensure continuity of mappings. Lastly, these mappings in the MBO are used for mapping transformation process using mapping execution layer.

Chapter 6

Results and Evaluation

The proposed Health-care Semantic Reconciliations Framework's result and evaluation is divided into three stages, presented in this chapter. Firstly, we participated in the OAEI 2013 campaign for matching different tracks ontologies. The system was indexed with 15 other ontology matching systems. Secondly, the design aspects of MBO ontology are evaluated with the existing systems. Finally, the mappings stored in the MBO are validated with transformations of standard formats, to demonstrate higher level of accuracy. CDA and vMR standards are used for validating the accuracy of the mappings stored for mapping execution process. Explanation of all these stages for evaluation is described in the following sections.

6.1 SPHeRe Participation in OAEI 2013

SPHeRe system participated in the large bio medical track of OAEI 2013 campaign. The statistics in the proceeding sections are taken from OAEI website [114] based on SPHeRe participation in different tracks.

6.1.1 Large Biomed Track: FMA and NCI Ontologies

There are two tasks to match FMA and NCI ontologies based on the fragments (small or whole) of these ontologies. The small fragments of FMA and NCI ontologies contains 3,696 classes (5% of FMA) and 6,488 classes (10% of NCI). On the other hand, whole FMA and NCI ontologies contains 78,989 and 66,724 classes respectively. The results of SPHeRe system related to these tasks is shown in Table 6.1. The systems are categorized based on F-Measure score, although differences exists between the score of precision and recall of these systems.

The detail of the experiments include 2,890 mappings available as benchmark for evaluation of our system. Small FMA and NCI ontologies matching resulted in total 2,349 mappings generation. These includes 2,234 True Positive (TP), 115 False Positive (FP), 656 False Negative (FN). On the other hand for Whole FMA and NCI ontologies matching, total of 2,736 mappings were generated, that includes 2,224 TP, 512 FP, and 666 FN mappings.

Table 6.1: Large Biomed Track: FMA and NCI Ontologies

Small FMA and NCI Ontologies Matching

S.No	System	Precision	Recall	F-Measure
1	LogMap-BK	0.95	0.88	0.91
2	YAM++	0.98	0.85	0.91
3	GOMMA	0.96	0.86	0.91
4	AML-BK-R	0.96	0.86	0.90
5	AML-BK	0.94	0.87	0.90
6	LogMap	0.95	0.85	0.90
7	AML-R	0.96	0.82	0.89
8	ODGOMS-v1.2	0.95	0.83	0.89
9	AML	0.95	0.83	0.89
10	LogMapLt	0.96	0.81	0.88
11	ODGOMS-v1.1	0.96	0.81	0.88
12	ServOMap	0.95	0.81	0.88
13	SPHeRe	0.96	0.77	0.86
14	Hotmatch	0.96	0.75	0.84
15	IAMA	0.98	0.58	0.73

Whole FMA and NCI Ontologies Matching

S.No	System	Precision	Recall	F-Measure
1	YAM++	0.90	0.85	0.87
2	GOMMA	0.86	0.83	0.85
3	LogMap	0.87	0.79	0.83
4	LogMap-BK	0.87	0.79	0.83
5	AML-BK	0.82	0.79	0.80
6	AML-BK-R	0.83	0.78	0.80
7	AML-R	0.89	0.72	0.80
8	AML	0.88	0.73	0.80
9	SPHeRe	0.85	0.75	0.80
10	ServOMap	0.73	0.80	0.76
11	LogMapLt	0.69	0.81	0.74
12	IAMA	0.90	0.58	0.71

6.1.2 Large Biomed Track: FMA and SNOMED Ontologies

The matching problem between FMA and SNOMED ontologies also consists of two tasks based on small and whole fragments of these ontologies. For small fragments of these ontologies, FMA contains 10,157 classes (13% of FMA) and SNOMED contains 13,412 classes (5% of SNOMED). In the same way, for whole ontologies, FMA contains 78,989 classes and SNOMED contains 1,22,646 classes (40% of SNOMED). SPHeRe system is indexed in these tasks as shown in Table 6.2 but with low recall value of the system.

The detail of the experiments include 2,890 mappings available as benchmark for evaluation of our system. Small FMA and SNOMED ontologies matching resulted in total 1,688 mappings generation. These includes 293 TP, 1,395 FP, 2,597 FN mappings. On the other hand, for Whole FMA and SNOMED ontologies matching, total of 2,366 mappings were generated, that includes

303 TP, 2,063 FP, and 2,587 FN mappings.

Table 6.2: Large Biomed Track: FMA and SNOMED Ontologies

Small FMA and SNOMED Ontologies Matching

S.No	System	Precision	Recall	F-Measure
1	YAM++	0.98	0.73	0.84
2	AML-BK	0.94	0.73	0.82
3	AML	0.94	0.72	0.82
4	AML-BK-R	0.95	0.70	0.80
5	AML-R	0.95	0.69	0.80
6	LogMap-BK	0.96	0.67	0.79
7	LogMap	0.97	0.66	0.78
8	ServOMap	0.95	0.62	0.75
9	ODGOMS-v1.2	0.86	0.57	0.69
10	GOMMA	0.92	0.38	0.54
11	ODGOMS-v1.1	0.88	0.22	0.35
12	HotMatch	0.87	0.21	0.34
13	LogMapLt	0.97	0.18	0.30
14	Hertuda	0.57	0.20	0.29
15	SPHeRe	0.92	0.16	0.27

Whole FMA and SNOMED Ontologies Matching

S.No	System	Precision	Recall	F-Measure
1	YAM++	0.95	0.72	0.82
2	AML-BK	0.94	0.65	0.77
3	AML	0.96	0.62	0.76
4	AML-BK-R	0.94	0.62	0.74
5	AML-R	0.97	0.59	0.74
6	ServOMap	0.86	0.62	0.72
7	LogMap-BK	0.87	0.60	0.71
8	LogMap	0.89	0.59	0.71
9	GOMMA	0.41	0.26	0.31
10	LogMapLt	0.88	0.18	0.30
11	SPHeRe	0.61	0.16	0.25
12	IAMA	0.75	0.13	0.23

6.1.3 Large Biomed Track: NCI and SNOMED Ontologies

The small fragment of NCI and SNOMED ontologies consists of 23,958 classes (36% of NCI) and 51,128 classes (17% of SNOMED) for matching purpose. The whole fragments consists of NCI 66,724 classes and 1,22,464 classes (40% of SNOMED). The results of SPHeRe participation in these tracks are shown in Table 6.3.

The detail of the experiments include 2,890 mappings available as benchmark for evaluation of our system. Small NCI and SNOMED ontologies matching resulted in total 9,957 mappings generation. These includes 294 TP, 9,663 FP, 2,596 FN mappings. On the other hand for Whole NCI and SNOMED ontologies matching, total of 10,508 mappings were generated, that includes 297 TP, 10,211 FP, and 2,593 FN mappings.

Table 6.3: Large Biomed Track: NCI and SNOMED Ontologies

Small NCI and SNOMED Ontologies Matching

S.No	System	Precision	Recall	F-Measure
1	LogMap-BK	0.89	0.68	0.68
2	LogMap	0.90	0.67	0.67
3	ServOMap	0.93	0.64	0.64
4	AML-BK-R	0.92	0.65	0.65
5	AML-BK	0.89	0.66	0.66
6	AML-R	0.92	0.63	0.63
7	YAM++	0.97	0.61	0.61
8	AML	0.89	0.64	0.64
9	LogMapLt	0.94	0.56	0.56
10	GOMMA	0.94	0.54	0.54
11	SPHeRe	0.92	0.47	0.47
12	IAMA	0.96	0.44	0.44

Whole NCI and SNOMED Ontologies Matching

S.No	System	Precision	Recall	F-Measure
1	ServOMap	0.82	0.64	0.72
2	YAM++	0.88	0.60	0.71
3	AML-BK	0.92	0.56	0.70
4	AML-BK-R	0.93	0.55	0.69
5	LogMap-BK	0.87	0.58	0.69
6	LogMap	0.88	0.57	0.69
7	AML	0.93	0.55	0.69
8	AML-R	0.94	0.54	0.68
9	LogMapLt	0.80	0.56	0.66
10	GOMMA	0.79	0.53	0.63
11	SPHeRe	0.88	0.47	0.61
12	IAMA	0.92	0.44	0.59

6.1.4 Discussion

SPHeRe system is capable of performing ontology matching on different sizes and domain of ontologies. The initial purpose of this matching system was to build it for matching between large scale biomedical ontologies, but later experiments were performed on multiple domain ontologies so verify its applicability in other domain different sizes of ontologies. Experiments were performed on other tracks of OAEI 2013 campaign such as Anatomy Track, Conference Track, and Library Track as shown in Table 6.4.

Table 6.4: Small Ontologies Tracks

Anatomy Track					Conference Track					Library Track				
S.No	System	Precision	Recall	F-Measure	S.No	System	Precision	Recall	F-Measure	S.No	System	Precision	Recall	F-Measure
1	AML-bk	0.95	0.93	0.94	1	SPHeRe	0.81	0.80	0.80	1	ODGOMS	0.70	0.83	0.76
2	GOMMA-bk	0.92	0.93	0.92	2	YAM++	0.78	0.65	0.71	2	YAM++	0.69	0.81	0.74
3	YAM++	0.94	0.87	0.90	3	AML-bk	0.82	0.53	0.67	3	MatcherPref	0.91	0.63	0.74
4	AML	0.95	0.83	0.89	4	LogMap	0.76	0.54	0.65	4	ServOMap	0.70	0.78	0.74
5	LogMap	0.92	0.85	0.88	5	AML	0.82	0.51	0.66	5	AML	0.62	0.88	0.70
6	GOMMA	0.96	0.80	0.87	6	ODGOMS1_2	0.70	0.55	0.62	6	MatcherPrefDE	0.98	0.58	0.73
7	StringsAuto	0.90	0.78	0.83	7	StringsAuto	0.74	0.50	0.62	7	MatcherAllLabels	0.61	0.88	0.72
8	LogMapLite	0.96	0.73	0.83	8	ServOMap_v104	0.69	0.50	0.59	8	Logmap	0.78	0.64	0.70
9	MapSSS	0.90	0.77	0.83	9	MapSSS	0.77	0.46	0.61	9	LogmapLite	0.65	0.77	0.70
10	ODGOMS	0.98	0.71	0.82	10	ODGOMSI1_1	0.72	0.47	0.59	10	HerTUDA	0.52	0.92	0.67
11	WikiMatch	0.99	0.67	0.80	11	HerTUDA	0.70	0.46	0.58	11	SPHeRe	0.67	0.66	0.65
12	HotMatch	0.98	0.64	0.77	12	WikiMatch	0.70	0.45	0.57	12	HotMatch	0.73	0.58	0.65
13	StringEquiv	1.00	0.62	0.77	13	WeSeE-Match	0.79	0.42	0.60	13	MatcherPrefEN	0.88	0.42	0.57
14	SPHeRe	0.99	0.62	0.76	14	IAMA	0.74	0.44	0.59	14	XmapSig	0.80	0.32	0.45
15	XMapSig	0.86	0.67	0.75	15	HotMatch	0.67	0.47	0.57	15	StringAuto	0.77	0.19	0.30

In all these tracks, for small different domain ontologies, SPHeRe system managed to be listed in the top 15 indexed systems. In conference track, it showed better precision, recall, and f-measure as compared to other systems. SPHeRe system is based on using object oriented and ontology alignment design patterns for storing the mappings in the MBO. Therefore, the design factors associated with MBO are important for comparison with existing systems.

6.2 Ontology Alignment Patterns based MBO

Existing ontology matching systems mainly focus on the accuracy of mappings and lack assessment of the external quality factors from the measurement of the internal design properties. We evaluate our proposed system with Coupling Factor (COF), Number of Polymorphic methods (NOP), and Rate of Change (RoC) metrics by comparing it with existing systems, FALCON and LogMap. We selected FALCON and LogMap for comparison with the proposed system because of factors such as; participation in OAEI several years, corresponding publications availability to understand their approach thoroughly, its source code availability (to understand the design and implementation of the system), and also complete system availability (to run ontology matching tests for observing its output). These systems class diagrams are generated from their source code using Intellij Idea tool4, that support a wide array of refactoring for Java, cross language refactoring and other advanced features [31]. We use Quality Model (QMOOD) approach [10] to quantitatively assess the external factors such as extendibility and reusability as measures of software maintainability.

6.2.1 Coupling Factor (COF)

Coupling Factor (COF) is a metric to determine dependencies between the classes. Therefore, the formula to calculate COF is given in Equation 6.1.

$$COF = df/(tc^2 - tc) \tag{6.1}$$

where *df* = Total Dependency Factor

and *tc* = Total No. of Classes

SPHeRe system is based on the MBO using object oriented and ontology design patterns. Therefore, COF value of SPHeRe is less as compare to FALCON and LogMap systems. Fig. 2 shows the df and tc of the proposed system and the COFSPHeRe is calculated as shown in Equation 6.2.

$$COF_{SPHeRe} = 9/(12^2 - 12) = 0.068 \tag{6.2}$$

We compared our system with FALCON ontology matching system and used its Matcher package to calculate COF of its different sub-packages as shown in Fig. 8(a). We observed that FALCON has high coupling as compare to the proposed system. Class diagram of FALCON system's Package PBM is shown in Fig. 8(b) and Equation 6.3 calculates its COF value as 0.127, which is very high as compare to proposed system.

$$COF_{FALCON_{PackagePBM}} = 14/(11^2 - 11) = 0.127 \tag{6.3}$$

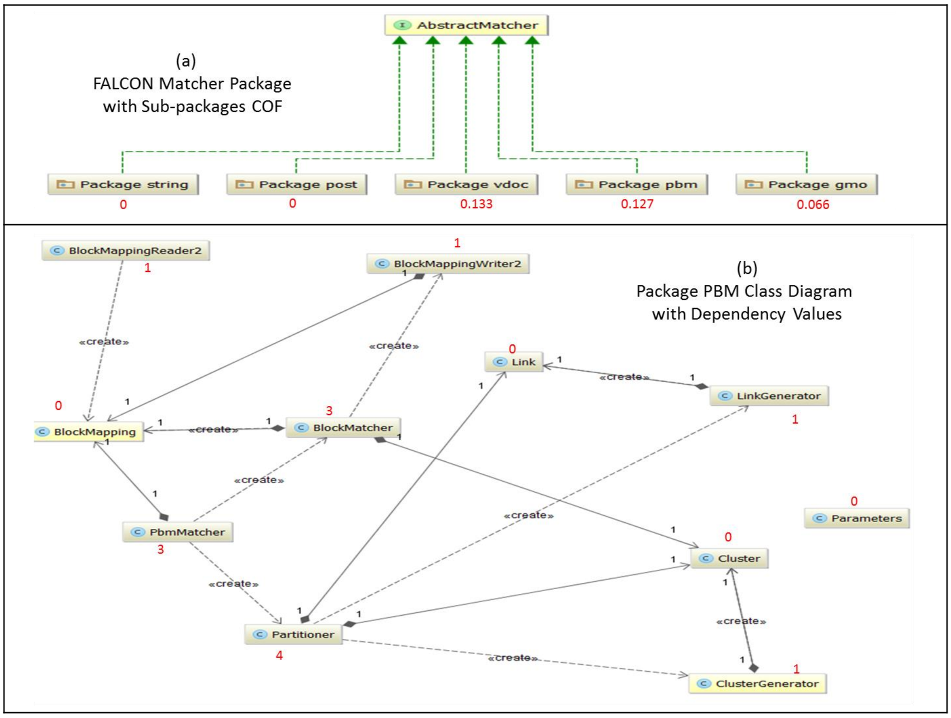


Figure 6.1: FALCON Packages and Coupling Factor

LogMap system overall class diagram consists of approximately 26 packages and classes having too much dependencies with each other, resulting in highly coupled system. We selected two packages (Stemming and Reasoning) for comparison with the proposed system. These packages class diagrams are shown in Fig. 9. Fig. 9 (a) and (b) illustrates class diagrams of LogMap system's Stemming and Reasoning packages respectively. Stemming package has more COF as compare to proposed system while Reasoning package has less COF value as shown in Equation 6.4 and 6.5 respectively.

$$COF_{LogMap_{PackageStemming}} = 20/(14^2 - 14) = 0.11 \tag{6.4}$$

$$COF_{LogMap_{PackageReasoning}} = 13/(17^2 - 17) = 0.047 \tag{6.5}$$

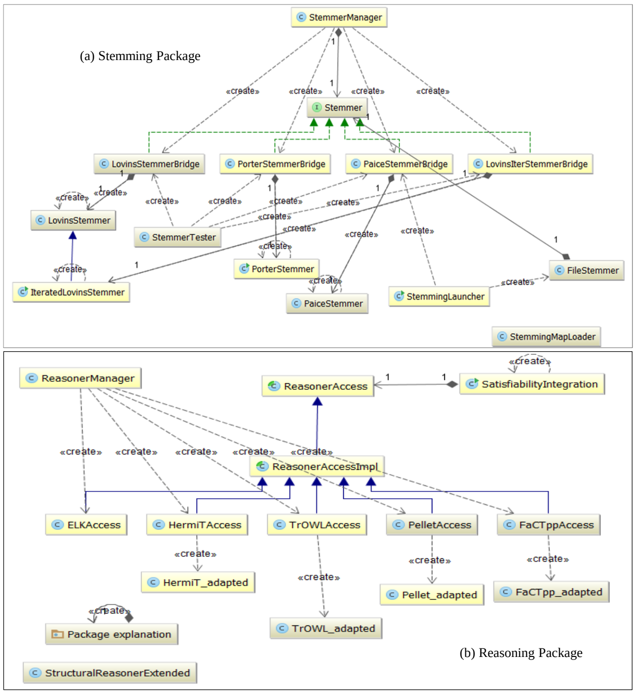


Figure 6.2: LogMap Class Diagrams: (a) Stemming Package, (b) Reasoning Package.

6.2.2 Number of Polymorphic Methods (NOP)

The number of polymorphic methods (NOP) in a class diagram determines the value for polymorphism. Therefore, in Fig. 2, it can be observed that *populateMBO()* is the polymorphic method that returns *MBOStrategy* instance. So, the NOP in a class diagram is the level of polymorphism which is 7 in the proposed system as shown in Equation 6.6. This suggests that the system has more extendibility by implementing only the *populateMBO()* polymorphic method.

$$NOP_{SPHeRe} = 7 \tag{6.6}$$

The increase in composition and association of a class diagram results in high coupling and less polymorphism. FALCON class diagram shows more composition and association relationships whereas the proposed system contains more polymorphic methods in the class diagram. Fig. 8(a) shows extend relationship to AbstractMatcher class, which suggests that there may be a polymorphic method in class diagram of FALCON system's Package PBM, shown in Fig. 8(b). Therefore, maximum polymorphism value for FALCON system is 1 as shown in Equation 6.7, which is less as compare to the proposed system. A new bridge algorithm must have to implement populateMBO() polymorphic method, thus increasing the polymorphism value. LogMap system two packages polymorphism value is 5 as shown in Equation 6.8, which is also less than the proposed system.

$$NOP_{FALCON} = 1 \tag{6.7}$$

$$NOP_{LogMap} = 5 \tag{6.8}$$

6.2.3 Rate of Change (RoC)

The key factor for successful ontology matching system is flexibility and extendibility based on new requirements. As new techniques and methodologies continuously evolve in ontology matching domain, measurement of Rate of Change (RoC) based on COF becomes necessary for evaluating the extendibility of the system. Therefore, RoC can be measured by Equation 6.9, based on change in the COF due to addition of new classes and dependencies.

$$RoC = \Delta COF \tag{6.9}$$

For testing rate of change, we introduced unidirectional dependency of +1 in *df* and *tc*, so equations 6.2 and 6.3 are transformed to equations 6.10 and 6.11 respectively. In the same way LogMap's Equations 6.4 and 6.5 are transformed to Equations 6.12 and 6.13 respectively.

$$COF_{SPHeRe'} = 10/(13^2 - 13) = 0.064 \tag{6.10}$$

$$COF_{FALCON'_{PackagePBM}} = 15/(12^2 - 12) = 0.114 \tag{6.11}$$

$$COF_{LogMap'_{PackageStemming}} = 21/(15^2 - 15) = 0.1 \tag{6.12}$$

$$COF_{LogMap'_{PackageReasoning}} = 14/(18^2 - 18) = 0.045 \tag{6.13}$$

The proposed system RoC is considerably less than FALCON and LogMap system, which shows the extendibility and reusability features of our system and easy adaptation of new changes. Equation 6.14 and Equation 6.15 shows that the proposed system has the better capacity to accommodate any changes in the system design as compare to FALCON system. LogMap's Stemming package has higher RoC while Reasoning package has less RoC value as compare to proposed system RoC value. These packages RoC values are shown in Equations 6.16 and 6.17.

$$\Delta COF_{SPHeRe} = COF_{SPHeRe} - COF_{SPHeRe'} = 0.068 - 0.064 = 0.004 \tag{6.14}$$

$$\begin{aligned} \Delta COF_{FALCON_{PackagePBM}} = COF_{FALCON_{PackagePBM}} - COF_{FALCON'_{PackagePBM}} \\ = 0.127 - 0.114 = 0.013 \end{aligned} \tag{6.15}$$

$$\begin{aligned} \Delta COF_{LogMap_{PackageStemming}} = COF_{LogMap_{PackageStemming}} - COF_{LogMap'_{PackageStemming}} \\ = 0.11 - 0.1 = 0.01 \end{aligned} \tag{6.16}$$

$$\Delta COF_{LogMap_{PackageReasoning}} = COF_{LogMap_{PackageReasoning}} - COF_{LogMap'_{PackageReasoning}} \\ = 0.047 - 0.045 = 0.002 \tag{6.17}$$

6.2.4 Discussion

Extendibility and reusability are the two main metrics for evaluation of the proposed system. These are discussed in relation to polymorphism and coupling of the proposed system measured in the previous sub-sections.

6.2.4.1 Extendibility

Extendibility is one of the evaluation metric of the proposed system. A new bridge algorithm can easily be accommodated in the system design with low coupling, high polymorphism and less rate of change as explained in previous section. This is achieved by using strategy design pattern with the PRMs. The new bridge algorithm only requires to implement the interface. We consider as a scenario that a new bridge is introduced that is based on instance based matching, called Instance Matching Bridge. *InstancePRM* is connected to the PRM in the MBO representation view that deals with actual representation of the alignment. A class *InstancePattern* will implement the *MBOStrategy* interface class and provide its reference information to *ConcreteMediator* class. Therefore, its tuple metrics information is as follows:

- I_F: An algorithm to match source and target concepts based on instances comparison.
- I_N: *InstancePRM* and *InstancePattern* classes to be added in the class diagram to support extendibility. This algorithm resolves specific problem and only requires to implement an interface.
- Q: ⟨polymorphism, increased⟩
- S: Source Concept that belongs to the matching source ontology.

- *T*: Target Concept that belongs to the matching target ontology.
- *A*: Instances of source and target concepts.
- *EC*: Specific number of instances matches than source and target concepts are similar. A threshold value n should be achieved by the number of instance matched.
- *MV*: A value between 0 and 1 that is based on instances matched.

6.2.4.2 Reusability

New bridge algorithm can be added to the system that can utilize existing bridge algorithms. Mediation between new and existing bridges is performed using mediator design pattern and PRMs. For example, a new bridge called Hyponym Bridge is introduced that uses *CBSB* and *PBSB* together to find matching concepts. *HyponymPRM* is connected to *PRM* in *MBO* representation view, and *HyponymPattern* class is also introduced to implement *MBOStrategy* interface class and provide reference to *ConcreteMediator* class. Tuple information is as follows:

- I_F: An algorithm to match source and target concepts based on existing CBSB and PBSB algorithms.
- I_N: HyponymPRM and HyponymPattern classes to be added in the class diagram for reusability.
- *Q*: $\langle$coupling, decrease$\rangle$
- *S*: Source Concept that belongs to the matching source ontology.
- *T*: Target Concept that belongs to the matching target ontology.
- *A*: Children and properties match of the matching concepts.
- *EC*: A specific number of children and properties match for source and target concepts match.
- *MV*: A value between 0 and 1 that is based on CBSB and PBSB results match.

These metrics enable easy integration of new bridge algorithms into the system that prolongs the system lifetime. State of the art matching techniques and new methodologies can be plug and play to the proposed system, without disturbing the design of the system.

6.3 EHR Standards Transformation

We have evaluated ARIEN system on datasets of Diabetes and Cancer patients encounter information gathered from local hospitals. The scenario of HMIS complaint to CDA and CDSS compliant to vMR is continued in this section. When the HMIS wants to query CDSS for guideline provision, the information is provided in CDA format. The CDSS can only process the information when it is in vMR Input format. So conversion from CDA to vMR Input is performed by ARIEN system. CDSS process the information and generates guidelines in the form of vMR Output. This vMR Output is not understandable format for HMIS therefore conversion from vMR Output to CDA is again performed by the proposed system. The experiments are conducted by converting CDA to vMR Input and vMR Output to CDA and the detailed description of the results are provided as follows:

6.3.1 Accuracy of Mappings in Transformation Process

6.3.1.1 CDA to vMR Input Conversion

The portion of the CDA document that is most related with vMR document is CDA body part. This part contains the patient observations details which requires processing by CDSS for recommendations or guidelines generation. Therefore, we observed smooth transformation process, as the number of vMR transformation constructs are subset of CDA transformation constructs. Also, this type of conversion mechanism is used for generation of *vMR Input* record from the CDA. The result of mapping constructs (concepts, attributes, and their values) precision, recall, and f-measure in the transformation process from CDA to vMR Input record is shown in Table 6.5 (Table for Precision, Recall, F-Measure (a)) and graph (Graph for Precision, Recall, F-Measure (b)) formats. Accuracy is measured by using the formula

Accuracy = [True Positives (TP) + True Negatives (TN)] / [True Positives (TP) + True Nega-

tives (TN) + False Positives (FP) + False Negatives (FN)]

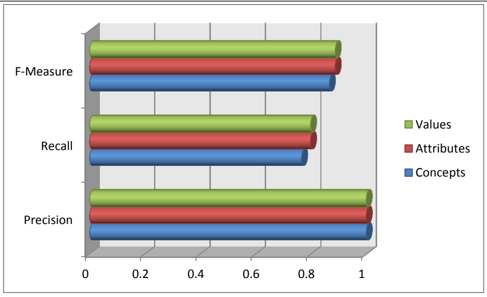
Table 6.5: CDA to vMR Input Transformation Process Results

Table for Precision, Recall, F-Measure (a)

	Precision	Recall	F-Measure
Concepts	1	0.767	0.867
Attributes	1	0.799	0.888
Values	1	0.799	0.888

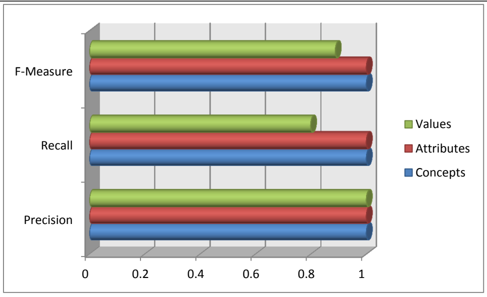
Overall Accuracy = 79%
(*MBO* Mappings with SPHeRe)

Graph for Precision, Recall, F-Measure (b)



	Precision	Recall	F-Measure
Concepts	1	1	1
Attributes	1	1	1
Values	1	0.799	0.888

Overall Accuracy = 93%
(*MBO* Mappings with SPHeRe and P-DCM Approach)



There are no false positives and true negatives in the transformation process therefore the precision of concepts, attributes and values is 1. Initially, we considered only SPHeRe's mappings generated and stored in *MBO*. The recall of these mapping constructs lies between 0.75 and 0.8 therefore approximately 0.877 is the overall F-Measure. The only concepts that are not properly transformed include **ClinicalStatementRelationships** and its sub-concepts, their attributes and values. This is because the mappings for these mappings constructs were not stored in the *MBO* as the SPHeRe matching systems of Accuracy Mapping Engine was unable to find suitable corresponding mapping concept in CDA. The overall accuracy achieved in this case is 79%. The concept **ClinicalStatementRelationships** and its sub-concepts,its attributes were identified with P-DCM approach by involving conformance information. The overall accuracy has improved to 93% with the only the values of the attributes of **ClinicalStatementRelationships** remained unresolved. A strategy to resolve this problem is discussed in Section 6.3.2 that will further increase the level of accuracy.

Organization specific conformance information is stored in P-DCM ontology. This information is used to generate customized mappings that have contributed in the overall increase in accuracy level from 79% to 93%. Therefore, organization conformance information handling is necessary for generating customized mappings to support generalized mappings for complete data interoperability.

6.3.1.2 VMR Output to CDA Conversion

The vMR Input is processed by CDSS to generate recommendations/guidelines in vMR Output format. The format of guidelines is not compatible with HMIS therefore vMR Output conversion to CDA format is required. The information of vMR Output is mapped with CDA body portion completely; however, there exists mapping constructs in CDA that are necessary for parsing document. These constructs are handled by structural mappings but transformation is faced with the values assignment problem. Values to the concepts and their attributes that are transformed using structural mappings cannot be assigned automatically. We provided a solution to this problem by involving human interventions. The details of the results of conversion process without human intervention are provided in Table 6.6.

Table 6.6: VMR Output to CDA Transformation Process Results

Table for Precision, Recall, F-Measure (a)

	Precision	Recall	F-Measure
Concepts	1	1	1
Attributes	1	1	1
Values	1	0.856	0.922

Overall Accuracy = 95%

Graph for Precision, Recall, F-Measure (b)

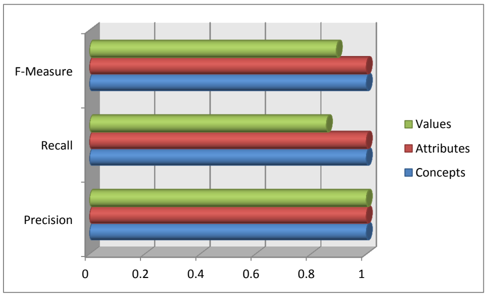


Table 6.6 (Table for Precision, Recall, F-Measure (a)) and graph (Graph for Precision, Recall, F-Measure (b)) shows the precision, recall and f-measure for concepts, attributes and their values conversion as shown in Table 6.6. Unlike CDA to vMR Input conversion, this type of conversion has precision, recall, and f-measure equal to 1 for concepts and attributes except values. The values recall and f-measure are 0.856 and 0.922 respectively. The overall accuracy for transformation process is 95%. The only problem of values transformation at run-time is faced because of

the structural mappings transformation process in which values cannot be automatically converted due to lack of information and human interventions becomes inevitable. If the values to these concepts are provided beforehand the conversion process shows 100% conversion result. As vMR output only contain recommendations and it uses less concepts than vMR input therefore, generalized mappings transform vMR ouput to CDA with 95% accuracy having no need of customized mappings.

6.3.2 Discussion

The proposed system transformation process performs structural mappings, sequential mappings, and data transformation mappings as explained in the previous sections. These mappings involve transformation mapping constructs (concept, attributes/properties and their values) conversion between different standards format. The more the mappings are stored in the *MBO*, the accurate and more complete will be the transformed format. We faced some challenges in the mapping process, the detail of their cause and our solution are the focus of this section. The discussion is based on the level of transformation from vMR to CDA format.

The vMR to CDA conversion process involves the transforming mapping constructs of *vMR Output*. The *vMR Output* consists of the guidelines or recommendations generated for patients after processing by the CDSS. In this case the number of transformation mapping constructs in vMR is less while due to some structural transformations requirements CDS transformation mapping constructs are more. The problem occurs only in the value assignments stage for particular attributes of concepts that are part of the structural mappings transformation process. Table 6.7 shows *Author* and *Custodian* classes of CDA standard that are necessary for parsing the CDA document. In CDA RMIM *"an author is a person in the role of an assigned author (AssignedAuthor class). The entity playing the role is a person (Person class) or a device (AuthoringDevice class)"* [73]. On the other hand, *"Custodian represents the organization that is in charge of maintaining the document. Every CDA document has exactly one custodian"* [73]. These classes values while conversion from vMR Output to CDA are performed by structural mappings but the Author class values for attributes time, address, city, assignedPerson name, and other (as shown in Table 6.7 (a)) and also Custodian class values for attributes id, name, address and others (as shown in Table 6.7 (b)). To address this problem, human intervention is made and these particular attribute values are asked by the system to manually enter for conversion to CDA format. In this way with minimal human interventions, maximum accuracy of the transformation process is achieved for true data interoperability among communicating HISs.

Table 6.7: Author and Custodian Concepts Information in CDA

Author Class (a)	**Custodian Class (b)**
<author>	<custodian>
<time value=""/>	<assignedCustodian>
<assignedAuthor>	<representedCustodianOrganization>
<id extension="" root=""/>	<id root=""/>
<addr>	<name/>
<streetAddressLine/>	<telecom value="tel:" use=" "/>
<city/>	<addr>
<postalCode/>	<streetAddressLine/>
<country/>	<city/>
</addr>	<postalCode/>
<assignedPerson>	<country/>
<name>	</addr>
<given/>	</representedCustodianOrganization>
<family/>	</assignedCustodian>
<suffix/>	</custodian>
</name>	
</assignedPerson>	
</assignedAuthor>	
</author>	

The accuracy of the transformation process is directly related to the mappings stored and also their representation. If the mappings stored consist of more annotated information then the challenges can be easily handled in conversion process. Some concepts can map with multiple concepts, their identification and transformation is performed in our case by utilizing the annotated information stored with the matched concepts in the *MBO*. **ObservationFocus** attribute in vMR can be mapped with **code** and **title** attribute of CDA. Therefore, in **title** and **observationFocus** mapping, hierarchy information is annotated for conversion. This type of concept to multiple concepts mappings are handled by annotating parent information with concept. So, in CDA to vMR transformation, **title** attribute can only be converted to **observationFocus** when hierarchy when the parent class of title attribute exists in the stored mapping, otherwise **code** attribute of CDA will map with **observationFocus** attribute of vMR.

Chapter 7

Conclusion and Future Directions

7.1 Conclusion

Semantic heterogeneity problem exists in various domains due to difference in the data formats that are required for information exchange. A sophisticated strategy of mediation dealing with the diverse aspects for interoperability is the solution to resolve the issue of heterogeneity. Resolving semantic heterogeneity among heterogeneous health-care systems is one of the key goal achieved by this thesis. Semantic heterogeneity aspects are observed from generic standards structure and semantics and also organization's involvement using the conformance guide. The SPHeRe and P-DCM approaches resolves this kind of semantic heterogeneity by dealing with generalized and customized mappings respectively. While generalized mappings are more towards the generation and representation of mappings between heterogeneous health-care standards; customized mappings more focuses on the conformance guide of the organization that provides information about conformance issues of organization with concepts related to different standards and localized concepts. This semantic heterogeneity is mainly resolved by the semantic reconciliations between generalized and customized mappings.

Semantic reconciliation applicability is another factor of the proposed approach that helps in resolving the semantic heterogeneity between heterogeneous health-care systems. The problem is resolved by providing effective mappings storage and representation scheme in the form of flexible and extendible design patterns oriented MBO. The design patterns results in higher level of maintainability of the system. This can be observed by the comparison of design factors of MBO such as low coupling factor, low rate of change and high polymorphism values as compared to existing systems. The validity of these semantic reconciliations information stored in the MBO can only be performed when higher degree of accuracy is achieved in the transformation process between different standard formats. This leads to the provision of an interoperable solution that can manage the exchange of information among different heterogeneous health-care systems.

SPHeRe an ontology matching system is one of the contribution of thesis for providing the generalized mappings. Its use of object oriented design patterns with ontology alignment design patterns defines its uniqueness to the existing approaches. Also, the sequence of bridge algorithms execution for maintaining accuracy with design patterns dealing with the maintainability aspect is another uniqueness of the system. SPHeRe also participated in the OAEI initiative and precision of 86% is achieved in large biomedical ontologies dataset. It was successful in running all the tasks provided in the large biomedical ontologies track.

Mediation Bridge Ontology is another contribution of the thesis that deals with the semantic reconciliation problem. The design aspects of MBO is superior as compare to the existing systems with low coupling factor value of 0.068, high polymorphism value of 7, and less rate of change value of 0.064 comparing to Logmap and FALCON systems. The factors that shows uniqueness of the approach with existing approaches are; effective mapping storage and representation, structure for management of ontology, flexibility for accommodating new matching dimensions, understandability for expert verifications, and mappings personalization for customized mappings.

The proposed approach provides the higher level of accuracy between heterogeneous health-care standards using the approaches described above. The experiments run on CDA and vMR standards showed approximately 80% of accuracy when only SPHeRe system was used and the level of accuracy increased to above 90% when SPHeRe is used with P-DCM approach. Finally, all these generalized mappings, customized mappings, mapping representation and storage, mapping evolution and transformation among standard formats are packaged into a single middleware framework called Health-care Semantic Reconciliation Framework. This framework provides interoperable solution among different heterogeneous health-care systems compliant to different standards and achieves the overall objective of semantic data interoperability.

7.2 Future Directions

Semantic interoperability has two main aspects that leads to true semantic interoperability i.e. data and process. Semantic process interoperability compliments semantic data interoperability by handling semantic reconciliations in the process artifacts of different health-care standards. Workflows of different standards requires alignment for semantic reconciliations on communication patterns for data exchange. This thesis work will be extended to semantic process interoperability as future work. The role of ontology alignment design patterns will be investigated for automation in the process workflows among different organizations. Also, the mediation ontology will be extended to deal with the process as well as data interoperability aspects for true interoperability.

The proposed work will also be extended for investigating evaluation aspects for interoperable systems. This is the grey area where benchmarks are required to absorb the continuous evolution of health-care standards and models. It will also help in improving the limited matching capabilities of ontology matching systems. Experts intervention should be limited in regulating the information exchange for automation in the integration platform. Evaluation framework for interoperable systems will require knowledge or model about related health-care standards, to perform routing of information among heterogeneous standards.

Bibliography

[1] M. Hussain, A. Khattak, W. Khan, I. Fatima, M. Amin, Z. Pervez, R. Batool, M. Saleem, M. Afzal, M. Faheem *et al.*, "Cloud-based smart cdss for chronic diseases," *Health and Technology*, vol. 3, no. 2, pp. 153–175, 2013.

[2] G. Freriks, "Semantic interoperability artefact modelling standard: Clinical information models," Electronic Record Services BV, Tech. Rep., June 2012, (Last visited in November 2012). [Online]. Available: http://www.en13606.org/wiki/images/1/1c/ERS_SIAMS_%28most_recent_version%29.pdf

[3] W. Hsu, R. K. Taira, S. El-Saden, H. Kangarloo, and A. A. Bui, "Context-based electronic health record: toward patient specific healthcare," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 16, no. 2, pp. 228–234, 2012.

[4] G. Tyson, A. Taweel, S. Zschaler, T. Staa, and B. Delaney, "A model-driven approach to interoperability and integration in systems of systems," in *Proc. of Workshop on Model-Based Software and Data Integration (MBSDI)*, 2011.

[5] M. Masud, M. S. Hossain, and A. Alamri, "Data interoperability and multimedia content management in e-health systems," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 16, no. 6, pp. 1015–1023, 2012.

[6] M. Yuksel, A. Dogac, C. Taskin, and A. Yalcinkaya, "A case for enterprise interoperability in healthcare it: Personal health record systems," *Revolutionizing Enterprise Interoperability Through Scientific Foundations*, p. 192, 2014.

[7] B. E. Dixon, D. J. Vreeman, and S. J. Grannis, “The long road to semantic interoperability in support of public health: Experiences from two states,” *Journal of biomedical informatics*, 2014.

[8] F. Adebesin, R. Foster, P. Kotzé, and D. Van Greunen, “A review of interoperability standards in e-health and imperatives for their adoption in africa,” 2013.

[9] E. Commission, “Ict standards in the health sector: Current situation and prospects, 2008,” http://goo.gl/KOKZn, (Last visited in September 2014).

[10] “E-health standards and interoperability, 2012,” http://alturl.com/xonvr, (Last visited in September 2014).

[11] L. Otero-Cerdeira, F. J. Rodríguez-Martínez, and A. Gómez-Rodríguez, “Ontology matching: A literature review,” *Expert Systems with Applications*, 2014.

[12] S. Pavel and J. Euzenat, “Ontology matching: state of the art and future challenges,” *Knowledge and Data Engineering, IEEE Transactions on*, no. 25, pp. 158–176, 2013.

[13] J. Euzenat and P. Shvaiko, “The matching problem,” in *Ontology Matching*. Springer, 2013, pp. 25–54.

[14] ——, “Classifications of ontology matching techniques,” in *Ontology Matching*. Springer, 2013, pp. 73–84.

[15] E. Jiménez-Ruiz, C. Meilicke, B. C. Grau, and I. Horrocks, “Evaluating mapping repair systems with large biomedical ontologies.” in *Description Logics*, 2013, pp. 246–257.

[16] C. Martínez Costa, M. Menárguez-Tortosa, and J. T. Fernández-Breis, “Clinical data interoperability based on archetype transformation,” *Journal of Biomedical Informatics*, vol. 44, no. 5, pp. 869–880, 2011.

[17] “Cimi history and overview,” (Last visited in Sep 2014). [Online]. Available: http://www.semantichealthnet.eu/SemanticHealthNet/assets/File/CIMI%20Background%202014%2003%2010.pdf

[18] W. T. Goossen, "Detailed clinical models: Representing knowledge, data and semantics in healthcare information technology," *Healthcare informatics research*, vol. 20, no. 3, pp. 163–172, 2014.

[19] T. O. Joey Coyle, Yan Heras and S. Huff, "Clinical element model," http://www.clinicalelement.com/docs/CEReference20081114.pdf, Tech. Rep., November 2008, (Last visited September 2014).

[20] C. Tao, C. G. Parker, T. A. Oniki, J. Pathak, S. M. Huff, and C. G. Chute, "An owl meta-ontology for representing the clinical element model," in *AMIA annual symposium proceedings*, vol. 2011. American Medical Informatics Association, 2011, p. 1372.

[21] M. Horridge, M. E. Aranguren, J. Mortensen, M. A. Musen, and N. F. Noy, "Ontology design pattern language expressivity requirements." in *WOP*, 2012.

[22] R. Lenz, M. Peleg, and M. Reichert, "Healthcare process support: achievements, challenges, current research," *International Journal of Knowledge-Based Organizations (IJKBO)*, vol. 2, no. 4, 2012.

[23] G. A. Lewis, E. Morris, S. Simanta, and L. Wrage, "Why standards are not enough to guarantee end-to-end interoperability," in *Composition-Based Software Systems, 2008. ICCBSS 2008. Seventh International Conference on*. IEEE, 2008, pp. 164–173.

[24] C. Ohmann and W. Kuchinke, "Future developments of medical informatics from the viewpoint of networked clinical research," *Methods Inf Med*, vol. 48, no. 1, pp. 45–54, 2009.

[25] "Clinical information modeling initiative," http://www.amia.org/the-standards-standard/2012-volume3-edition1/clinical-information-modeling-initiative, September 2013.

[26] M. Ba and G. Diallo, "Large-scale biomedical ontology matching with servomap," *IRBM*, vol. 34, no. 1, pp. 56–59, 2013.

[27] S. Schulz and C. Martínez-Costa, "How ontologies can improve semantic interoperability in health care," in *Process Support and Knowledge Representation in Health Care*. Springer, 2013, pp. 1–10.

[28] T. Kirsten, A. Gross, M. Hartung, E. Rahm *et al.*, "Gomma: a component-based infrastructure for managing and analyzing life science ontologies and their evolution." *J. Biomedical Semantics*, vol. 2, p. 6, 2011.

[29] E. Jiménez-Ruiz and B. C. Grau, "Logmap: Logic-based and scalable ontology matching," in *The Semantic Web–ISWC 2011*. Springer, 2011, pp. 273–288.

[30] D. Faria, C. Pesquita, E. Santos, M. Palmonari, I. F. Cruz, and F. M. Couto, "The agreementmakerlight ontology matching system," in *On the Move to Meaningful Internet Systems: OTM 2013 Conferences*. Springer, 2013, pp. 527–541.

[31] W. Hu and Y. Qu, "Falcon-ao: A practical ontology matching system," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 6, no. 3, pp. 237–239, 2008.

[32] P. Lambrix and H. Tan, "Sambo-a system for aligning and merging biomedical ontologies," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 4, no. 3, pp. 196–206, 2006.

[33] Y. R. Jean-Mary, E. P. Shironoshita, and M. R. Kabuka, "Asmov: Results for oaei 2010," *Ontology Matching*, vol. 126, 2010.

[34] J. A. Maldonado, D. Moner, D. Boscá, J. T. Fernández-Breis, C. Angulo, and M. Robles, "Linkehr-ed: A multi-reference model archetype editor based on formal semantics," *International Journal of Medical Informatics*, vol. 78, no. 8, pp. 559–570, 2009.

[35] R. Sahay, R. Fox, A. Zimmermann, A. Polleres, and M. Hauswrith, "A methodological approach for ontologising and aligning health level seven (hl7) applications," *Availability, Reliability and Security for Business, Enterprise and Health Information Systems*, pp. 102–117, 2011.

[36] B. Xu, P. Wang, J. Lu, Y. Li, and D. Kang, "Bridge ontology and its role in semantic annotation," in *Cyberworlds, 2004 International Conference on*. IEEE, 2004, pp. 329–334.

[37] P. Wang, B. Xu, J. Lu, Y. Li, and D. Kang, "Theory and semi-automatic generation of bridge ontology in multi-ontologies environment," in *On the Move to Meaningful Internet Systems 2004: OTM 2004 Workshops*. Springer, 2004, pp. 763–767.

[38] I. F. Cruz, W. Sunna, N. Makar, and S. Bathala, "A visual tool for ontology alignment to enable geospatial interoperability," *Journal of Visual Languages & Computing*, vol. 18, no. 3, pp. 230–254, 2007.

[39] I. F. Cruz, F. P. Antonelli, and C. Stroe, "Agreementmaker: efficient matching for large real-world schemas and ontologies," *Proceedings of the VLDB Endowment*, vol. 2, no. 2, pp. 1586–1589, 2009.

[40] D. Ngo, Z. Bellahsene, and R. Coletta, "A flexible system for ontology matching," in *IS Olympics: Information Systems in a Diverse World*. Springer, 2012, pp. 79–94.

[41] D. Ngo and Z. Bellahsene, "Yam++: a multi-strategy based approach for ontology matching task," in *Knowledge Engineering and Knowledge Management*. Springer, 2012, pp. 421–425.

[42] F. Song, G. Zacharewicz, and D. Chen, "An ontology-driven framework towards building enterprise semantic information layer," *Advanced Engineering Informatics*, vol. 27, no. 1, pp. 38–50, 2013.

[43] Y. Jean-Mary, E. Shironoshita, and M. Kabuka, "Ontology matching with semantic verification," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 7, no. 3, pp. 235–251, 2009.

[44] D. H. Teodoro, R. Choquet, D. Schober, G. Mels, E. Pasche, P. Ruch, and C. Lovis, "Interoperability driven integration of biomedical data sources," *Studies in health technology and informatics*, vol. 169, pp. 185–9, 2011.

[45] D. Teodoro, E. Pasche, J. Gobeill, S. Emonet, P. Ruch, and C. Lovis, "Building a transnational biosurveillance network using semantic web technologies: requirements, design, and preliminary evaluation." *Journal of medical Internet research*, vol. 14, no. 3, pp. e73–e73, 2011.

[46] P. Xu, Y. Wang, and B. Liu, "A differentor-based adaptive ontology-matching approach," *Journal of Information Science*, vol. 38, no. 5, pp. 459–475, 2012.

[47] J. Xie, F. Liu, and S.-U. Guan, "Tree-structure based ontology integration," *Journal of Information Science*, vol. 37, no. 6, pp. 594–613, 2011.

[48] O. Šváb-Zamazal, V. Svátek, and L. Iannone, "Pattern-based ontology transformation service exploiting oppl and owl-api," in *Knowledge Engineering and Management by the Masses*. Springer, 2010, pp. 105–119.

[49] F. Scharffe, O. Zamazal, and D. Fensel, "Ontology alignment design patterns," *Knowledge and Information Systems*, pp. 1–28, 2013.

[50] A. Dogac, G. B. Laleci, S. Kirbas, Y. Kabak, S. S. Sinir, A. Yildiz, and Y. Gurcan, "Artemis: Deploying semantically enriched web services in the healthcare domain," *Information Systems*, vol. 31, no. 4, pp. 321–339, 2006.

[51] E. D. Valle, D. Cerizza, V. Bicer, Y. Kabak, G. B. Laleci, and H. Lausen, "The need for semantic web service in the ehealth," in *In W3C workshop on Frameworks for Semantics in Web Services*, 2005.

[52] R. Sahay, W. AKhtar, and R. Fox, "Ppepr: Plug and play electronic patient records," in *Proceedings of the 2008 ACM Symposium on Applied Computing (SAC)*. Fortaleza, Ceara, Brazil: ACM, March 16-20 2008.

[53] C. Carr and S. Moore, "Ihe: a model for driving adoption of standards," *Computerized medical imaging and graphics*, vol. 27, no. 2, pp. 137–146, 2003.

[54] F. Wozak, E. Ammenwerth, A. Hörbst, P. Sögner, R. Mair, T. Schabetsberger, and E.-W. Center, "Ihe based interoperability for service oriented architectures." in *GI Jahrestagung (1)*, 2008, pp. 59–64.

[55] A. Ducrou, "Complete interoperability in healthcare: technical, semantic and process interoperability through ontology mapping and distributed enterprise integration techniques," *Doctor of Philosophy Thesis*, 2009.

[56] A. Magni, R. de Oliveira Albuquerque, R. de Sousa Jr, M. Hans, and F. Magni, "Solving incompatibilities between electronic records for orthodontic patients," *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 132, no. 1, pp. 116–121, 2007.

[57] W. A. Khan, M. Hussain, K. Latif, M. Afzal, F. Ahmad, and S. Lee, "Process interoperability in healthcare systems with dynamic semantic web services," *Computing*, vol. 95, no. 9, pp. 837–862, 2013.

[58] O. Kilic and A. Dogac, "Achieving clinical statement interoperability using r-mim and archetype-based semantic transformations," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 13, no. 4, pp. 467–477, 2009.

[59] O. Bodenreider *et al.*, "Biomedical ontologies in action: role in knowledge management, data integration and decision support," *Yearb Med Inform*, vol. 67, p. 79, 2008.

[60] A. Hein, O. Nee, D. Willemsen, T. Scheffold, A. Dogac, G. B. Laleci *et al.*, "Saphire-intelligent healthcare monitoring based on semantic interoperability platform-the homecare scenario," in *1st European Conference on eHealth (ECEH06), Fribourg, Switzerland*, 2006, pp. 191–202.

[61] M. B. Amin, R. Batool, W. A. Khan, S. Lee, and E.-N. Huh, "Sphere: A performance initiative towards ontology matching." *The Journal of Supercomputing*, vol. 68, no. 1, pp. 274–301, 2014.

[62] W. A. Khan, M. B. Amin, A. M. Khattak, M. Hussain, and S. Lee, "System for parallel heterogeneity resolution (sphere) results for oaei 2013," *Ontology Matching*, p. 184, 2013.

[63] W. A. Khan, M. Hussain, M. Afzal, M. B. Amin, M. A. Saleem, and S. Lee, "Personalized-detailed clinical model for data interoperability among clinical standards," *Telemedicine and e-health*, vol. 19, no. 8, pp. 632–642, 2013.

[64] W. Hasselbring and S. Pedersen, "Metamodelling of domain-specific standards for semantic interoperability," *Professional Knowledge Management*, pp. 557–559, 2005.

[65] W. A. Khan, A. M. Khattak, M. Hussain, M. B. Amin, M. Afzal, C. Nugent, and S. Lee, "An adaptive semantic based mediation system for data interoperability among health information systems," *Journal of medical systems*, vol. 38, no. 8, pp. 1–18, 2014.

[66] M. Hussain, W. Khan, M. Afzal, and S. Lee, "Smart cdss for smart homes," *Impact Analysis of Solutions for Chronic Disease Prevention and Management*, pp. 266–269, 2012.

[67] M. Hussain, M. Afzal, W. A. Khan, and S. Lee, "Clinical decision support service for elderly people in smart home environment," *12th International Conference On Control, Automation, Robotics and Vision (ICARCV 2012)*, 2012.

[68] R. Qamar and A. Rector, "Semantic issues in integrating data from different models to achieve data interoperability," *Studies in Health Technology and Informatics*, vol. 129, no. 1, p. 674, 2007.

[69] P. A. de Clercq, J. A. Blom, H. H. Korsten, A. Hasman *et al.*, "Approaches for creating computer-interpretable guidelines that facilitate decision support," *Artificial intelligence in medicine*, vol. 31, no. 1, pp. 1–28, 2004.

[70] A. A. Boxwala, S. Tu, M. Peleg, Q. Zeng, O. Ogunyemi, R. A. Greenes, E. H. Shortliffe, and V. L. Patel, "Toward a representation format for sharable clinical guidelines," *Journal of biomedical informatics*, vol. 34, no. 3, pp. 157–169, 2001.

[71] M. Samwald, K. Fehre, J. De Bruin, and K.-P. Adlassnig, "The arden syntax standard for clinical decision support: Experiences and directions," *Journal of biomedical informatics*, vol. 45, no. 4, pp. 711–718, 2012.

[72] G. W. Beeler, S. Huff, W. Rishel, A.-M. Shakir, M. Walker, C. Mead, and G. Schadow, "Message development framework," Version 3.3, Health Level Seven, Inc., Tech. Rep., December 1999.

[73] R. H. Dolin, L. Alschuler, S. Boyer, C. Beebe, F. M. Behlen, P. V. Biron, and A. S. Shvo, "Hl7 clinical document architecture, release 2," HL7 V3 Normative Edition 2011, Health Level Seven, Inc., Tech. Rep., 2011.

[74] ——, "Hl7 clinical document architecture, release 2," *Journal of the American Medical Informatics Association*, vol. 13, no. 1, pp. 30–39, 2006.

[75] "Domain analysis model: Virtual medical record for clinical decision support (vmr-cds), release 1," Health Level Seven, Inc., Tech. Rep., January 2012, http://www.hl7.org/v3ballot2012jan/html/dams/uvvmr/uvvmr.html.

[76] "Design principles," http://www.openehr.org/releases/1.0.1/html/ architecture/overview/Output/design_principles.html, (Last visited in June 2011).

[77] "The openehr archetype model, the template object model (tom)," http://www.openehr.org/releases/1.0.1/architecture/am/tom.pdf, Tech. Rep., March 2007, (Last visited March 2011).

[78] T. Beale and S. Heard, "The openehr archetype model, archetype definition language, adl 2," http://www.openehr.org/releases/1.0.2/architecture/am/adl2.pdf, Tech. Rep., March 2007, (Last visited March 2011).

[79] HL7Wiki, "Openehr datatypes mapping," (Last visited in March 2011). [Online]. Available: http://wiki.hl7.org/index.php?title=OpenEHR_datatypes_mapping

[80] "The openehr reference model, ehr information model," Tech. Rep., April 2007, (Last visited March 2011). [Online]. Available: http://www.openehr.org/releases/1.0.1/ architecture/rm/ehr_im.pdf

[81] "The openehr reference model, demographic information model," http://www.openehr.org/releases/1.0.1/architecture/rm/demographic_im.pdf, Tech. Rep., Febuary 2006, (Last visited March 2011).

[82] "The openehr reference model, data structures information model," http://www.openehr.org/releases/1.0.1/architecture/rm/data_structures_im.pdf, Tech. Rep., September 2006, (Last visited March 2011).

[83] "The openehr reference model, common information model," http://www.openehr.org/releases/1.0.1/architecture/rm/common_im.pdf, Tech. Rep., April 2007, (Last visited March 2011).

[84] "The openehr reference model, data types information model," http://www.openehr.org/releases/1.0.1/architecture/rm/data_types_im.pdf, Tech. Rep., April 2007, (Last visited March 2011).

[85] "The openehr reference model, support information model," http://www.openehr.org/releases/1.0.1/architecture/rm/support_im.pdf, Tech. Rep., April 2007, (Last visited March 2011).

[86] "Sphere: System for parallel heterogeneity resolution," http://uclab.khu.ac.kr/sphere, (Last visited in June 2013).

[87] "Wordnet: A lexical database for english," http://wordnet.princeton.edu/, (Last visited in June 2013).

[88] L. Li and Y. Yang, "Agent-based ontology mapping and integration towards interoperability," *Expert Systems*, vol. 25, no. 3, pp. 197–220, 2008.

[89] J. Euzenat, C. Meilicke, H. Stuckenschmidt, P. Shvaiko, and C. Trojahn, "Ontology alignment evaluation initiative: six years of experience," in *Journal on data semantics XV*. Springer, 2011, pp. 158–192.

[90] J. Vlissides, R. Helm, R. Johnson, and E. Gamma, "Design patterns: Elements of reusable object-oriented software," *Reading: Addison-Wesley*, vol. 49, p. 120, 1995.

[91] A. K. M. A. S. L. Wajahat Ali Khan, M Bilal and E. S. Kim, “Object oriented and ontology alignment patterns based expressive mediation bridge ontology (mbo),” *Journal of Information Science*, 2014.

[92] J. Euzenat and P. Shvaiko, *Ontology matching*. Springer Berlin, 2007, vol. 18.

[93] G. Navarro, “A guided tour to approximate string matching,” *ACM computing surveys (CSUR)*, vol. 33, no. 1, pp. 31–88, 2001.

[94] A. Gangemi and V. Presutti, “Ontology design patterns,” in *Handbook on Ontologies*. Springer, 2009, pp. 221–243.

[95] V. Presutti and A. Gangemi, “Content ontology design patterns as practical building blocks for web ontologies,” in *Conceptual Modeling-ER 2008*. Springer, 2008, pp. 128–141.

[96] F. Scharffe, “Correspondence patterns representation,” Ph.D. dissertation, University of Innsbruck, 2009.

[97] H. Thomas, D. O’Sullivan, and R. Brennan, “Evaluation of ontology mapping representation,” in *Proceedings of the Workshop on Matching and Meaning*, 2009, pp. 64–68.

[98] K. Offit, “Personalized medicine: new genomics, old lessons,” *Human genetics*, pp. 1–12, 2011.

[99] “Health information exchange,” (Last visited in March 2012). [Online]. Available: http://en.wikipedia.org/wiki/Health_information_exchange

[100] “Health information exchange definition_defined by experts,” (Last visited in September 2012). [Online]. Available: http://www.medical-record-coding.com/2011/12/health-information-exchange-definition/

[101] “Detailed clinical model,” (Last visited in September 2012). [Online]. Available: http://www.detailedclinicalmodels.nl/dcm-en

[102] W. Goossen, “Using detailed clinical models to bridge the gap between clinicians and hit,” *Stud Health Technol Inform*, vol. 141, pp. 3–10, 2008.

[103] C. Parker, R. Rocha, J. Campbell, S. Tu, and S. Huff, "Detailed clinical models for sharable, executable guidelines," *Medinfo*, vol. 11, no. Pt 1, pp. 145–8, 2004.

[104] "Clinical information modelling initiative goes with archetypes & uml profile," (Last visited in March 2012). [Online]. Available: http://www.openehr.org/326-OE

[105] "Cimi details," (Last visited in March 2012). [Online]. Available: http://wiki.hl7.org/images/9/99/CIMI-Seeking-Clarity-20120108.pdf

[106] "Architecture overview," (Last visited in October 2012). [Online]. Available: http://www.openehr.org/releases/1.0.1/

[107] I. Health Level Seven, "Hl7 implementation guide for cda release 2: Nhsn healthcare associated infection (hai) reports, release 2," http://www.cdc.gov/nhsn/PDFs/CDA/CDAR2L3_IG_HAIRPT_R2_D2_2009FEB.pdf, Tech. Rep., 2009, (Last visited October 2012).

[108] "Hl7 reference information model," Tech. Rep., 2004, (Last visited October 2012). [Online]. Available: http://www.interopsante.org/offres/file_inline_src/412/412_P_15660_92.pdf

[109] W. A. Khan, M. Hussain, A. M. Khattak, M. B. Amin, S. Lee, and K. Latif, "Achieving interoperability among healthcare standards: Building semantic mappings at models level," in *ACM International Conference on Ubiquitous Information Management and Communication ACM ICUIMC (IMCOM)*, Febuary 2012.

[110] A. Khattak, Z. Pervez, W. Khan, S. Lee, and Y. Lee, "A self evolutionary rule-base," in *U-and E-Service, Science and Technology*. Springer, 2011, pp. 1–9.

[111] W. A. Khan, M. Hussain, M. Afzal, M. B. Amin, and S. Lee, "Healthcare standards based sensory data exchange for home healthcare monitoring system," in *34th Annual International Conference of IEEE Engineering in Medical and Biology Society*, August 2012.

[112] D. Bosca, J. A. Maldonado, D. Moner, and M. Robles, “Detailed clinical models to facilitate interstandard interoperability of data types,” in *23rd International Conference of the European Federation for Medical Informatics*, Oslo, Norway, August 2011.

[113] A. Khattak, P. Truc, L. Hung, V. Dang, D. Guan, Z. Pervez, M. Han, S. Lee, Y. Lee *et al.*, “Towards smart homes using low level sensory data,” *Journal of Sensors*, vol. 11, no. 12, pp. 11 581–11 604, 2011.

[114] O. 2013, “Results oaei 2013::large biomed track,” http://www.cs.ox.ac.uk/isg/projects/SEALS/oaei/2013/results2013.html, (Last visited in September 2014).

Appendix A

List of Publications

International Journal Papers:

[1] **Wajahat Ali Khan**, M Bilal, AM Khattak, Maqbool, Afzal, SY Lee and Eun Soo Kim, ”Object Oriented and Ontology Alignment Patterns based Expressive Mediation Bridge Ontology (MBO)”, Journal of Information Science, (SCIE, IF:1.08)(Accepted), 2014.

[2] **Wajahat Ali Khan**, AM Khattak, Maqbool Hussain, Bilal Amin, M Afzal, Christopher Nugent and Sungyoung Lee, ”An Adaptive Semantic based Mediation System for Data Interoperability among Health Information Systems”, Journal of Medical Systems (SCIE, IF 1.372), 38(8):28, 2014.

[3] **Wajahat Ali Khan**, Maqbool Hussain, Muhammad Afzal, Muhammad Bilal Amin, Muhammad Aamir Saleem and Sungyoung Lee, ”Personalized-Detailed Clinical Model for Data Interoperability among Clinical Standards”, Telemedicine and EHealth (SCI, IF:1.544), Vol. 19 Issue 8, pp.632-642, 2013.

[4] **Wajahat Ali Khan**, Maqbool Hussain, Khalid Latif, Muhammad Afzal, Farooq Ahmad, Sungyoung Lee, ”Process Interoperability in Healthcare Systems with Dynamic Semantic Web Services”, Computing Journal (SCI, IF:1.055), Vol. 95, Issue 9, pp 837-862, 2013.

[5] Asad Masood Khattak, Zeeshan Pervez, **Wajahat Ali Khan**, Adil Mehmood Khan, Khalid Latif, Sungyoung Lee, ”Mapping Evolution of Dynamic Web Ontologies”, Information Sciences, (SCI, IF 3.893) (Accepted), 2014.

[6] Maqbool Hussain, Taqdir Ali, **Wajahat Ali Khan**, Muhammad Afzal, Sungyoung Lee, Khalid Latif,;"Recommendations service for chronic disease patient in multi-model sensors home environment" Telemedicine and e-Health (SCI, IF:1.544), 2014

[7] Muhammad Bilal Amin, Rabia Batool, **Wajahat Ali Khan**, Sungyoung Lee, and Eui-Nam Huh. "SPHeRe:A Performance Initiative towards Ontology Matching by implementing Parallelism over Cloud Platform." The Journal of Supercomputing (SCI, IF:0.841), 68, no. 1 (2014): 274-301.

[8] Maqbool Hussain, AsadMasood Khattak, **Wajahat Ali Khan**, Iram Fatima, Muhammad Bilal Amin, Zeeshan Pervez, Rabia Batool, Muhammad Amir Saleem, Muhammad Afzal, Muhammad Fahim, Muhammad Hameed Saddiqi, Sungyoung Lee, and Khalid Latif, "Cloud-based Smart CDSS for Chronic Diseases", In Journal of Health and Technology - 3, no. 2 (2013): 153-175.

International Conference Papers:

[9] **Wajahat Ali Khan**, Muhammad Bilal Amin, Asad Masood Khattak, Maqbool Hussain, and Sungyoung Lee. "System for Parallel Heterogeneity Resolution (SPHeRe) results for OAEI 2013." In OM, pp. 184-189. 2013.

[10] **Wajahat Ali Khan**, Maqbool Hussain, Bilal Amin, Asad Masood Khattak, Muhammad Afzal, and Sungyoung Lee. "AdapteR Interoperability ENgine (ARIEN): An approach of Interoperable CDSS for Ubiquitous Healthcare." In Ubiquitous Computing and Ambient Intelligence. Context-Awareness and Context-Driven Interaction, pp. 247-253. Springer International Publishing, 2013.

[11] **Wajahat Ali Khan**, Maqbool Hussain, Muhammad Afzal, Muhammad Bilal Amin, and Sungyoung Lee. "Healthcare standards based sensory data exchange for Home Healthcare Monitoring System." In Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE, pp. 1274-1277. IEEE, 2012.

[12] **Wajahat Ali Khan**, Maqbool Hussain, Asad Masood Khattak, Muhammad Afzal, Bilal Amin, and Sungyoung Lee. "Integration of HL7 compliant smart home healthcare system and HMIS." In Impact Analysis of Solutions for Chronic Disease Prevention and Management, pp. 230-233. Springer Berlin Heidelberg, 2012.

[13] **Wajahat Ali Khan**, Maqbool Hussain, Asad Masood Khattak, Muhammad Bilal Amin, and Sungyoung Lee. "Saas based interoperability service for semantic mappings among healthcare standards." In 8th International Conference on Innovations in Information Technology. 2012.

[14] **Wajahat Ali Khan**, Asad Masood Khattak, Sungyoung Lee, Maqbool Hussain, Bilal Amin, and Khalid Latif. "Achieving interoperability among healthcare standards: building semantic mappings at models level." In Proceedings of the 6th International Conference on Ubiquitous Information Management and Communication, p. 101. ACM, 2012.

[15] **Wajahat Ali Khan**, Maqbool Hussain, Muhammad Afzal, Sungyoung Lee, Interoperability Engine for Smart CDSS, AAL Summit 2012, Spain

[16] Jamil Hussain, **Wajahat Ali Khan**, Muhammad Afzal, Maqbool Hussain, Byeong Kang and Sungyoung Lee, "Adaptive User Interface and User Experience based Authoring Tool for Recommendation Systems", Ubiquitous Computing and Ambient Intelligence (UCAmI), Belfast, Northern Ireland, December 2-4, 2014.

[17] Muhammad Afzal, Maqbool Hussain, Taqdir Ali, **Wajahat Ali Khan**, Sungyoung Lee, Byeong Ho Kang, "MLM-based Automated Query Generation for CDSS Evidence Support", Ubiquitous Computing and Ambient Intelligence (UCAmI), Belfast, Northern Ireland, December 2-4, 2014.

[18] Muhammad Bilal Amin, Mahmood Ahmad, **Wajahat Ali Khan**, and Sungyoung Lee, "Biomedical Ontology Matching as a Service", ICOST 2014, Advances in Cognitive Technologies, June 25-27, 2014

[19] Taqdir Ali, Maqbool Hussain, **Wajahat Ali Khan**, Muhammad Afzal, and Sungyoung Lee. ”Customized clinical domain ontology extraction for knowledge authoring tool.” In Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication, p. 23. ACM, 2014.

[20] Taqdir Ali, Maqbool Hussain, **Wajahat Ali Khan**, Muhammad Afzal, Byeong Ho Kang, and Sungyoung Lee. ”Arden syntax studio: Creating medical logic module as shareable knowledge.” In Innovations in Intelligent Systems and Applications (INISTA) Proceedings, 2014 IEEE International Symposium on, pp. 266-272. IEEE, 2014.

[21] Muhammad Afzal, Maqbool Hussain, **Wajahat Ali Khan**, Taqdir Ali, Sungyoung Lee, and Byeong Ho Kang. ”KnowledgeButton: An evidence adaptive tool for CDSS and clinical research.” In Innovations in Intelligent Systems and Applications (INISTA) Proceedings, 2014 IEEE International Symposium on, pp. 273-280. IEEE, 2014.

[22] Maqbool Hussain, Taqdir Ali, Muhammad Afzal, **Wajahat Ali Khan**, Sungyoung Lee, Lets Fire the CDSS integration issue with FHIR, 14th International HL7 Interoperability Conference (IHIC 2013) and FHIR Connectathon, Sydney, Australia, 27-29 October, 2013

[23] Muhammad Afzal, Maqbool Hussain, **Wajahat Ali Khan**, Taqdir Ali, Sungyoung Lee, and Hafiz Farooq Ahmad. ”Meaningful Integration of Online Knowledge Resources with Clinical Decision Support System.” In Inclusive Society: Health and Wellbeing in the Community, and Care at Home, pp. 280-285. Springer Berlin Heidelberg, 2013.

[24] Taqdir Ali, Maqbool Hussain, **Wajahat Ali Khan**, Muhammad Afzal, and Sungyoung Lee. ”Authoring tool: Acquiring sharable knowledge for Smart CDSS.” In Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE, pp. 1278-1281. IEEE, 2013.

[25] Maqbool Hussain, Muhammad Afzal, **Wajahat Ali Khan**, and Sungyoung Lee. ”Clinical decision support service for elderly people in smart home environment.” In Control Automation Robotics & Vision (ICARCV), 2012 12th International Conference on, pp. 678-683. IEEE, 2012.

[26] Rabia Batool, **Wajahat Ali Khan**, Maqbool Hussain, Jahanzeb Maqbool, Muhammad Afzal, and Sungyoung Lee. ”Towards personalized health profiling in social network.” In Information Science and Service Science and Data Mining (ISSDM), 2012 6th International Conference on New Trends in, pp. 760-765. IEEE, 2012.

[27] Muhammad Afzal, Maqbool Hussain, **Wajahat Ali Khan**, Sungyoung Lee, and Hafiz Farooq Ahmad. ”Social media canonicalization in healthcare: Smart cdss as an exemplary application.” In e-Health Networking, Applications and Services (Healthcom), 2012 IEEE 14th International Conference on, pp. 419-422. IEEE, 2012.

[28] Muhammad Bilal Amin, Aamir Shafi, Shujaat Hussain, **Wajahat Ali Khan**, and Sungyoung Lee. ”High performance Java sockets (HPJS) for scientific health clouds.” In e-Health Networking, Applications and Services (Healthcom), 2012 IEEE 14th International Conference on, pp. 477-480. IEEE, 2012.

[29] Maqbool Hussain, **Wajahat Ali Khan**, Muhammad Afzal, and Sungyoung Lee. ”Smart CDSS for smart homes.” In Impact Analysis of Solutions for Chronic Disease Prevention and Management, pp. 266-269. Springer Berlin Heidelberg, 2012.

[30] Muhammad Bilal Amin, **Wajahat Ali Khan**, Ammar Ahmad Awan, and Sungyoung Lee. ”Intercloud message exchange middleware.” In Proceedings of the 6th International Conference on Ubiquitous Information Management and Communication, p. 79. ACM, 2012.

[31] Asad Masood Khattak, Zeeshan Pervez, **Wajahat Ali Khan**, S. Y. Lee, and Y. K. Lee. ”A Self Evolutionary Rule-Base.” In U-and E-Service, Science and Technology, pp. 1-9. Springer Berlin Heidelberg, 2011.