An Ostensive Information Architecture to Enhance Semantic Interoperability for Healthcare Information Systems

Hua Guo

Submitted in partial fulfilment of the requirements of
University of Reading for the degree of
Doctor of Philosophy

Informatics Research Centre
Henley Business School

September 2022

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Name: Hua Guo

Certificate of Readiness to Be Included in Library

I grant powers of discretion to the University Librarian to allow this thesis to be copied in whole or in part without further reference to me. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgement.

Acknowledgements

"Life was like a box of chocolates. You never know what you're going to get." When I decided to apply for PhD at the University of Reading in 2018, I never knew such a fantastic academic journey was waiting for me. I feel very lucky to have met so many great mentors and friends and spent four fulfilling years in the Informatics Research Centre. I am grateful to all who have helped and encouraged me throughout my PhD journey.

Firstly, I would like to give my sincerest gratitude to my principal supervisor, Professor Kecheng Liu, for all guidance and support in my PhD journey. He led me on the road of scientific research and made me understood the true values of being a scientific researcher. The fantastic opportunities and academic experiences he gave me in the four years have gradually changed my career direction. When I decided to pursue academic development, he gave me great help and support. The four years I spent with him changed the trajectory of my life. Secondly, I would like to thank Dr Michael Scriney, my co-author, for his professional suggestions from a computer science perspective. It was a wonderful memory to collaborate with him working on a paper. I would also like to thank Prof Yinshan Tang, Prof Weizi Li and all the other academic and administrative staff in the Informatics Research Centre for supporting me throughout my PhD journey.

Moreover, I would like to thank my beloved family, my mother, my husband and my son. Their love and unconditional supports are the driving force behind my pursuit of academic development. I am so grateful for what they have done for me, making me feel surrounded by happiness all the time.

Lastly, special thanks to my PhD colleagues, Dr Muhammad Dashtban, Dr Sumayyah Jad, and Dr Ning Jing, for all the help and companionship. The time with them is the pearls of my memories.

Related Publications

Published

- Kecheng Liu, **Hua Guo** (2020) Artificial Intelligence and Data Analytics in Digital Business Transformation Before, During and Post COVID-19, in Monica Billio, Simone Varotto (eds), A New World Post COVID-19, Lessons for Business, the Finance Industry and Policy Makers, Edizioni Ca'Foscari, Venezia
- LIU, K. & **GUO, H.** 2020. Digital Innovation and Transformation to Business Ecosystems. In: J. FILIPE, M. SMIALEK, A. BRODSKY & S. HAMMOUDI (eds.) Enterprise Information Systems: 22th International Conference, ICEIS 2020. Springer.
- **Guo, H.** and Liu, K. (2019) Hierarchical ontology graph for solving semantic issues in decision support systems. In: International Conference on Enterprise Information Systems, 3-5 May 2019, Crete, Greece, pp. 483-487

Working papers

- **Guo,H**. Liu, K and Scriney,M. (2022) An Ostensive Information Architecture to enable semantic interoperability for healthcare information systems, Information Systems Frontier (accepting for publication)

- Checkfong Tan, Kecheng Liu and **Hua Guo** (2022) Information architecture for transforming healthcare organisations: A case study of a UK hospital.

Abstract

Health information systems (HISs) manage healthcare data and support decision-making in order to improve the quality of health service. The nature of healthcare demands, particularly resulting from patient-centred care policies and evidence-based medicine, necessitates the efficient management and usage of healthcare resources. Given the benefits that technological advancements such as sensor-based technologies and ubiquitous computing environments give to healthcare services, the HIS faces various hurdles. The Interoperability across diverse information systems, particularly the complexity introduced by continuously increasing data sources, is one of the greatest obstacles. This research focuses on the semantic ambiguity in information exchange processes and then proposes an ostensive information architecture based on the widely adopted international standard FHIR (Fast Health Interoperability Resources) to improve the semantic interoperability between healthcare information systems.

The semantic interoperability concerns the capacity of systems to share and interpret the meaning of information exchanged in ecosystems. Motived by the semantic ambiguity generated in FHIR implementation, this research investigates the semiotic sources semantic ambiguity. It then proposes an ostensive information architecture to reduce ambiguity caused by applying FHIR specifications in various contexts for various objectives.

The FHIR specification specifies a collection of present information models referred to as *resources, as well as* four paradigms for exchanging these *resources* amongst heterogeneous health information systems. As the FHIR specification seeks to cover all medical service scenarios, its conceptual descriptions lean towards generality and universality. As a consequence, the FHIR specification is rather flexible and implementors are free to utilise FHIR *resources*. In this context, the semantic ambiguity resulting from differing implementers' interpretations of the FHIR specification causes interoperability issues between systems. In order to decrease such ambiguity, this research introduces an ostensive method as an additional means of elucidating the semantics through the use of examples.

An ostensive approach defines concept through direct example. This method is often applied in language and philosophy, and it is considered particularly effective in clarifying semantics. In this research, the ostensive approach shows how implementers employ FHIR *resources* to represent local healthcare data. Thus, one healthcare service can be explained by 1) explaining its lexical-semantic by the FHIR knowledge graph; 2) explicating the semantics through the explicit correspondences between FHIR and local data attributes, and 3) showing examples of these attributes by pointing to local dataset values. This research builds an FHIR knowledge graph-based Semantic Engine using Neo4j to provide semantic interpretation and examples. The proposed information architecture has been tested using MIMIC III and diabetes datasets.

Overall, This study provides a semiotics-based information architecture approach to address semantic ambiguity and enhance the semantic interoperability. The suggested ostensive information architecture naturally separates semantic explanation and data storage, ensuring data privacy and enhance data security. The final chapter discusses benefits and future research directions.

Contents

List of Figures

List of Tables

List of Command box

List of Notations and Abbreviations

Symbol or Abbreviations	Meaning
AR	Augmented Reality
BBIS	Blood Bank Information System
CDISC	Clinical Data Interchange Standards Consortium
CDR	Clinical Data Repository
CDSS	Clinical Decision Support System
CIMI	Clinical Information Modelling Initiative
CIS	Clinical Information System
DIC	Data Integration Centre
EHR	Electronic Health Record
FA	Federated Architecture
GECCO	German Corona Consensus Dataset
HIE	Health Information Exchange
HIM	Health Information Management system
HIT	Health Information Technology
HIMSS	Healthcare Information and Management Systems Society
HIS	Health Information System
HL7	Health Level 7 international
IoT	Internet of Thing
IoMT	Internet of Medical Things
JSON	JavaScript Object Notation
LIS	Laboratory Information System
LOINC	Logical Observation Identifiers Names and Codes
IS	Information System
MDD	Model-Driven Development
MII	Medical Informatics Initiative
MIMIC	Medical Information Mart for Intensive Care
MR	Mixed Reality
MSA	Micro Service Architecture
PCC	Patient Centric Care
PMS	Patient Management System

Symbol or Abbreviations	Meaning
REST	Resource Representational State Transfer
SDO	Standards Development Organisation
SMART	Substitutable Medical Applications, Reusable Technologies
SNOMED-CT	Systematized Nomenclature of Medicine - Clinical Terms
SOA	Service Oriented Architecture
TOGAF	The Open Group Architecture Framework
VNA	Vendor Neutral Archive
VR	Virtual Reality

Chapter 1 Introduction

This chapter explains the context and rationale of designing an information architecture for healthcare ecosystems. This study investigates interoperability between healthcare information systems as a result of the demand for sharing healthcare data and the development of IT technology. Research issues and questions are discussed from both a theoretical and an empirical perspective. Consequently, the purpose and objectives of the research are outlined. The structure of this thesis is sketched at the end of the chapter.

1.1 Background and Motivation

A health information system (HIS) consists of interconnected components that gather, analyse, store and disseminate healthcare data to support decision-making, coordination, control, analysis and visualisation in order to improve the quality of health services (Stair and Reynolds, 2020). In light of the benefits provided to healthcare services by advances such as sensor-based technology and the ubiquitous computing environments for multiple HIS users, including physicians, patients, funders of healthcare, and regulatory bodies (He *et al.*, 2019), the HIS faces multiple challenges (Haux, 2006). How to efficiently manage and employ healthcare resources to provide quality care is one of the greatest difficulties. The nature of healthcare demands, particularly resulting from patient-centred care policies (Stewart, 2001, Håkansson Eklund *et al.*, 2019) and evidence-based medicine (Sackett, 1997, Martini, 2021), requires the efficient data exchange across domains.

As the HIS landscape has significantly expanded, with its complexity increasing exponentially, in response to increased levels of connectivity and stakeholder demand, HISs are evolving into healthcare ecosystems (Liu and Guo, 2021); facilities of this nature should have the capacity to content with multiple domain knowledge (Blobel, 2019), particularly heterogeneous data collected by novel medical devices or sensors (Kankanhalli *et al.*, 2016, Ristevski and Chen, 2018). In this context, system interoperability, which facilitates intercommunication and enables data sharing across disparate information systems (Geraci, 1990, Mouttham *et al.*, 2012, Lehne *et al.*, 2019b) is a major concern for enhancing the efficiency of health data usage.

The diversity of datasets generated by information obtained from wearable devices, telehealth, and digital therapeutics (Aungst and Patel, 2020, Li *et al.*, 2015) requires exchangeability, not only of the data but also the information they contain. Consequently, both academia and industry are paying increading attention to the issue of the interoperability of digital ecosystems (Grimson *et al.*, 2000, Lehne *et al.*, 2019b). According to the survey, the lack of interoperability across medical devices costs $ 35 billion annually (West Health Institute, 2013). Therefore, there is a need to enhance the interaction of complex systems in healthcare domain, particularly facilitate health data sharing from disparate HISs or data sources.

From information network connectivity to application interaction, interoperability can be categorised into three forms, specifically: technical, syntactic, and semantic (Joshi *et al.*, 2017, Tolk *et al.*, 2007). In the dimensions of technical and syntactic interoperability, consensus solutions have, to some extent, been developed; for example, information exchange protocols such as REST (Resource Representational State Transfer) API (Application Programming Interface) and the unified data formats are, in practice, becoming more widely adopted. However, there remain significant challenges for semantic interoperability (Ashrafi *et al.*, 2018, Geraci *et al.*, 1991, HIMSS, 2014); this concerns the capacity of systems to interpret the meaning of the exchanged information within ecosystems. Because the applications of artificial intelligence, advanced data analytics and wearable technologies are becoming increasingly common within healthcare ecosystems (Rehman *et al.*, 2022, Knight *et al.*, 2021), the subject of interaction between heterogeneous applications and systems has attracted the interest of many academics. Therefore, the motivation of this research is to explore the meaningful information exchange between two or more entities within a healthcare ecosystem at the semantic level (Ouksel and Sheth, 1999, Liu and Li, 2015).

Due to the complex nature of health information representations, international standards have been produced to achieve their semantic interoperability, such as HL7 v2 and v3 (HL7 International, 1987), open EHR (open EHR, 2003) and CEN/ISO 13606 (ISO). Although these standards claim to answer the problem of semantic exchange, they are implemented at different levels of interoperability, namely syntactic, semantic and pragmatic from the perspective of information exchange (Liu and Li, 2015). However, the semantic ambiguity continues to exist (Jiang *et al.*, 2015, Jiang *et al.*, 2016, Dolin *et al.*, 2018). In the background, FHIR (Fast Healthcare Interoperability Resources) is adopted as a research foundation to investigate the semantic ambiguity in the application of this specification because its high adoption in industry (Information Technology Industry Council, 2018).

As FHIR is still in the process of continuous development, numerous scholars and practitioners have highlighted semantic ambiguity difficulties for FHIR implementation (Kubick, 2016, Dolin *et al.*, 2018, Kraus, 2018, Beale, 2019). This research devoted to enhancing the semantic operability for FHIR. The current solutions to reduce semantic ambiguity can be roughly classified into three categories: 1) improving the harmonisation of FHIR resource usage through review processes (McClure *et al.*, 2020, Tute *et al.*, 2021) ; 2) developing detailed usage specifications on the FHIR according to the characteristics of local health data (Rosenau *et al.*, 2022, NHS Digital, 2021); 3) automatic FHIR specification conversion for specific databases (Pfaff *et al.*, 2019, Sayeed *et al.*, 2020). However, these solutions are limited by high development costs in terms of time and manpower, as well as a restricted application scope. Regarding the research gap, this research examines the semantic interoperability of FHIR from semiotics perspective and investigate a cost-effective and easily implementable solution.

This research is not concerned with establishing a new health information system to satisfy demand, but rather about with optimising the existing solution. In addition to the FHIR-based semantic interoperability, the research will investigate an appropriate research strategy that can expand the FHIR's adoption boundaries.

1.2 Research Problems and Questions

Given the necessity to explore the healthcare information architecture to respond to the requirements of healthcare data users and the importance of semantic interoperability between heterogeneous healthcare information systems, this research firstly investigates the interoperability challenges in health information systems; therefore, the first research question of this study is:

Q1. What is the current state of health information systems and the interoperability challenges they face?

Within the context of multi-system interaction, semantic interoperability is investigated in depth, with a primary focus on gaining a knowledge of the obstacles currently faced by health information systems and how those obstacles might be overcome. Thus, the second question of this research is:

Q2. How to improve the semantic interoperability regarding to the multidisciplinary and cross-organisational healthcare delivery?

After comparing the healthcare interoperability standards released by various international standards organisations from the perspective of supporting information exchange, Fast Healthcare Interoperability Resources (FHIR) is chosen as the basis of the study. From a literature review, the limitations of FHIR in terms of supporting semantic interoperability are identified. Regarding the semantic ambiguity generated in FHIR implementation, the semiotics theory is adopted to facilitate understanding the root cause of semantic ambiguity. Thus, the third question of this study is:

Q3. How to enhance semantic interoperability of FHIR?

Considering the urgent requirements of using healthcare data for high-quality service or academic research, the fourth question of this research is:

Q4. How can FHIR-underpinned healthcare information platform integrate data from heterogeneous local systems with a unified schema for multiple purposes?

These three questions concern the semantic interoperability in healthcare ecosystems from the information system architecture point of view and the perspective of implementation to respond to the requirements of healthcare data use.

1.3 Research Aims and Objectives

The aim of this research is to improve the semantic interoperability of digital healthcare ecosystems on the basis of FHIR, aiming to support patient-centred care (Stewart, 2001) and evidence-based medicine (Sackett, 1997).

In order to achieve this aim, this research targets to propose an optimised information architecture based on FHIR to enable the collaboration between heterogeneous local health information systems. The proposed information architecture could synergy dispersed health information systems, including the microdata sources such as wearable devices and monitors, in a collaborating way of working. To align with the research aims, the specific objectives of this research are:

1. To comprehend the complexity of healthcare ecosystem, which is driven by sophisticated technology and the demands for optimal utilisation of health data.
2. To examine the existing interoperability solutions for health information systems and understand their limitations.
3. To investigate the underlying source of semantic ambiguity generated by FHIR implementation.
4. To develop an information architecture for HIS based on the identified underlying causes of semantic ambiguity and an assessment of the literature on multi-agent information architecture.
5. To validate the efficacy of proposed information architecture in decreasing semantic ambiguity and facilitating health data exchange from heterogeneous data sources.

1.4 The Expected Outcome

Due to the fact that the FHIR design based on the Internet protocol has rapidly gained a large number of adoptions, this research will maximally retain the benefits of the current design to the greatest extent possible; therefore, this research will not alter the structure of FHIR but will instead explore a solution from a broader perspective in terms of a new information architecture.

Since this ecosystem will continue to increase in complexity and is still in the process of developing naturally, it will surely continue to receive new data sources from a variety of sources. Therefore, the anticipated information system architecture can reconcile disparate data sources; in other words, it can manage dynamically heterogeneous data sources. This system architecture will reduce the amount of effort necessary for data migration and system installation by taking into consideration the diversity and complexity of existing systems.

1.5 Thesis Outline

This thesis is structured and follows the falsification research process. The nature of this research has resulted in the subscription of falsification as its logical reasoning approach to guide the development of the ostensive information architecture for healthcare information systems. The falsification process enables researchers to propose a solution that solves the limitations of existing solutions while inheriting the advantages of existing solutions. In Figure 1, chapters are organised according to research activities to address research aims.

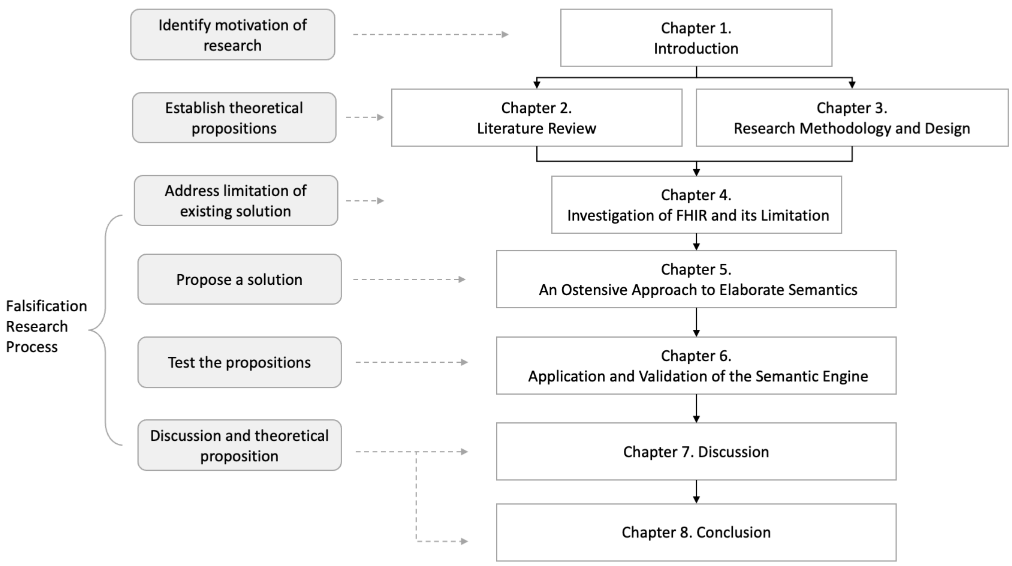


Figure 1 Thesis outline

Chapter 2 reviews the relevant literature about digital healthcare ecosystems. In the beginning, this chapter addresses the complexity of the healthcare ecosystem and the impacts of emerging technologies on healthcare services. Then the chapter reviews the state-of-the-art health information systems from the perspective of their goals and challenges. After that, this chapter introduces the concept of health informatics and particularly emphasises the interoperability challenges of health information systems, which provides an antecedent for understanding the importance of interoperability to high-quality health care. Following with the present of the interoperability of the healthcare ecosystem from the perspective of the levels of information exchange and summarises the interoperability levels of existing international standards in the healthcare domain. Finally, based on the interoperability level, this chapter selects two specifications for semantic interoperability for a detailed comparison and then chooses one as the basis for this study.

Chapter 3 illustrates the philosophical stance of this research. Firstly, this chapter explains the author's empirical observations and philosophical position derived from the observations. Then the chapter studies the mainstream research paradigms and compares falsificationism with others. From the comparison, falsificationism can be identified as it provides research philosophy and methods for innovating existing theories, which can guide the research of information system design. As sign/language is one of the critical aspects influencing the interaction between participants and information systems, which could generate semantic ambiguity in some circumstances, this research particularly explores the language used as a sign system in information system design. In the last, this chapter introduces the research design of the present thesis, which is carried out according to the falsificationist research process.

Chapter 4 identifies the fundamental course of semantic ambiguity generated in the FHIR (Fast Healthcare Interoperability Resources) implementation through a semiotic analysis of information interaction. The first section lays the groundwork for proposing a novel way to decrease semantic ambiguity. The following sections introduces FHIR including its evolutionary history and distinctions from previous HL7 versions. Following a detailed introduction of how FHIR applies in the healthcare information systems and the FHIR compliance challenges caused by semantic ambiguity. The subsequence sections then investigate the limitation of FHIR. The following section reviews the related works to improve FHIR compliance. As the ontology technique is typically used to promote the information exchange between multiple domains, the final section examines the ontology-related work employed in healthcare information systems.

Chapter 5 presents an ostensive information architecture, which enhances semantic interoperability and decreases semantic ambiguity during the process of FHIR implementation. Firstly, this chapter introduces the MSA (Micro-Services Architecture) in modern information system design and explains how the knowledge graph reflects the design idea of MSA in this research. Then, this chapter depicts the federated architecture of information systems, which is the infrastructure model adopted in this study. After that, this chapter demonstrates how federated architecture and MSA contribute to the deliberation of the ostensive information architecture through the descriptions of each technique and steps to construct a Semantic Engine and mapping the local clinical data to the Semantic Engine.

Chapter 6 validates the capacities of the ostensive information architecture through two case studies. The first shows how semantics are enhanced through the ostensive approach facilitated by the Semantic Engine; the second shows how the Semantic Engine works in the heterogeneous data exchanging situation to facilitate semantic interoperability between MIMIC III (Medical Information Mart for Intensive Care version III) datasets and diabetes datasets. The last section of this chapter describes the applications of the Semantic Engine in terms of supporting computational and analytical tasks.

Chapter 7 provides a summary of this research's contributions from theoretical, methodological, and practical viewpoints, as well as its answers to the four research questions.

Chapter 8 summarises this research from the perspective of knowledge accquisition. In the section that follow, the limitations of this study and their implications for future research are discussed. In the final section, a brief reflection of this research journey is presented.

Chapter 2 Literature Review

The literature review of this thesis will be undertaken from the perspectives of digital health information systems, information system architecture and semiotics at different stage of study. As seen in **Error! Reference source not found.**, the three domains are paired to conduct in-depth research on different perspective of semantic interoperability in HIS. The integration of information system architecture and digital HIS allows for an architectural understanding of interoperability difficulties within digital HIS. On the basis of identified interoperability concerns, a review of the semiotics domain helps to comprehend the use of signs in information systems, hence facilitating the investigation of the underlying causes of communication barriers in information exchange. The integration of semiotics with information system design facilitates the resolution of recognised semantic problems.

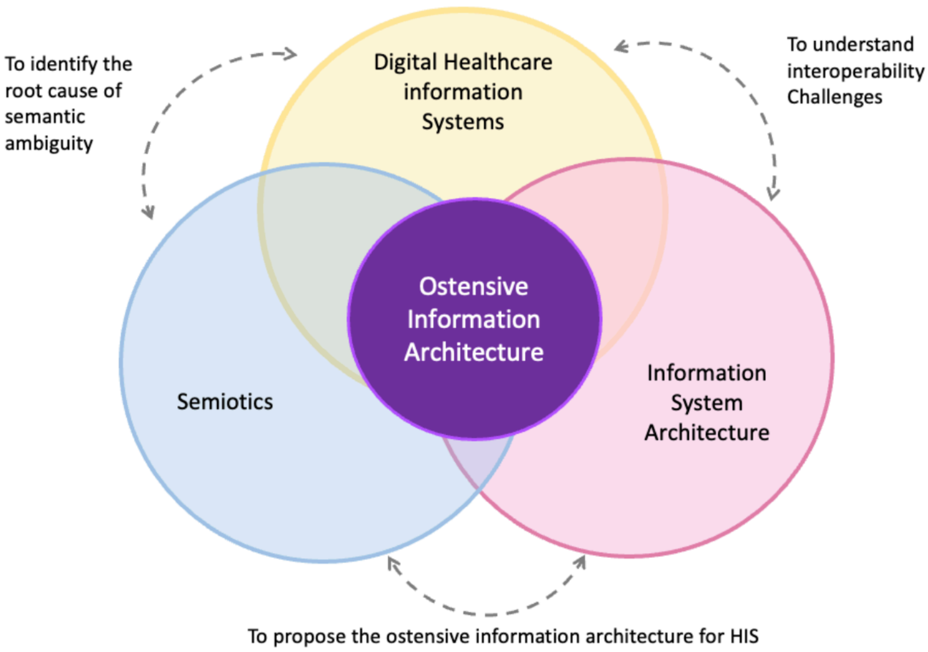


Figure 2 Literature review structure

In Chapter 2, health information systems are examined in order to comprehend the present state of digital HIS and the available solutions for semantic interoperability. In Chapter 4 and 5, the architecture and semiotics of information systems are examined, along with the process of identifying a solution to semantic interoperability.

In the beginning of Chapter 2, the complexity of the healthcare ecosystem is addressed, particularly in the context of health information system transformation driven by advanced technologies, such as Bigdata, AI, and Cloud computing, as well as by the demands of precision medicine, ubiquitous monitoring, and wearable devices. The direction of expanding the capacity of healthcare information systems through the use of advanced technologies can be identified, and the external needs for digital healthcare information systems also can be clarified. Section 2.2 examines state-of-the-art healthcare information systems, including the goals and challenges of healthcare information system. Section 2.3 examines the breadth of health informatics and the issues of interoperability. Section 2.4 addresses the classification of interoperability from the information systems perspective and examines the published standards for semantic interoperability. Based on sorting out the interoperability level of existing protocols, Section 2.5 selects two widely adopted international standards at the semantic interoperability level and compares them from the perspective of technology architecture, openness, scalability, flexibility, and portability. FHIR is selected as the foundation of this research to discuss semantic interoperability in healthcare information systems.

2.1 The Complexity of Healthcare Ecosystem

The healthcare ecosystem is profoundly changed by an explosion of innovation in digital technologies (illustrated in Figure 3), which enhance the understanding of patient health in both temporal and spatial dimensions (Mamom and Daovisan, 2022, Kaushik *et al.*, 2021), renovate the interaction between the healthcare providers and patients (Thornton, 2010, Sink *et al.*, 2022), and reshape the practices of diagnosis (Firouzi *et al.*, 2018, Li *et al.*, 2022), and even impact on the medical research paradigm (Practice, 2006, Kidholm *et al.*, 2021).

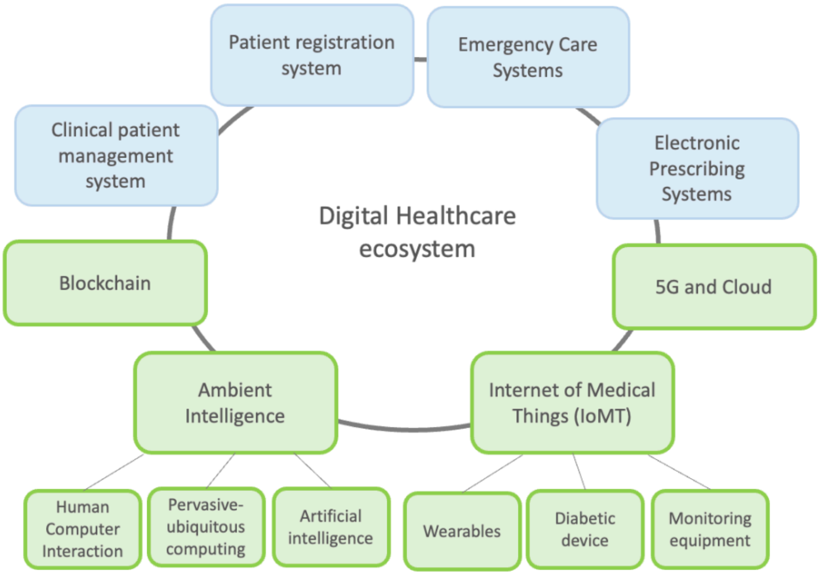


Figure 3 The complexity of digital healthcare ecosystem

From the onset of digital health, the “unquenchable needs for more and greater access to healthcare consumers” (Wen and Tan, 2003) are released and, in turn, promote the development of health information technology (HIT). Digital health shifts the paradigm of healthcare provision aiming at a more efficient, cost-effective, and time-saving mode. Empowered by digital health technologies, patients can make better-informed health decisions about their personalised health status, such as disease prevention and chronic conditions management outside traditional healthcare settings. In addition to providing such a personalised and integrated healthcare experience, HIT is also perceived as a strategic necessity for increasing provider productivity, engaging formal and informal caregivers, and improving equity and affordability. HIT is the booster to drive the evolution of the digital health ecosystem.

In this research, the digital health ecosystem is defined as a set of services and capabilities that benefit all participants in the healthcare value chain from the multi-dimensional data sharing enabled by a virtual data backbone (McKinsey&Company, 2019) . The following content discusses the components of the virtual data backbone and their implicit impact on health care.

A virtual data backbone is a general concept that refers to the infrastructure of the telecom network and the IT capabilities it carries. It is named a virtual data backbone because it is not a physical network, but a logically integrated network composed of multiple networks or communication technologies. From the data management point of view, the virtual data backbone comprises the data collection layer, data transmission and exchange layer, data security and privacy layer, data storage and analysis layer, and the top layer of human-machine interaction. As data are collected from various layers of a health ecosystem, its hierarchical structure must be considered, as each layer emphasises distinct data sources and application scenarios. Table 1 lists the emerging technologies in each layer.

Table 1 The promising technologies in the hierarchical model of data existing in network

Purpose and functions	Technology and means
Human-Machine interaction	Digital health (mobile APP, VR, AR and MR)
Data security and privacy	Blockchain
Data storage and analysis	Cloud, edge computing, AI, and Machine Learning
Data transmission and exchange	5G
Data collection	IoT, IoMT

- Data collection layer

At the level of data collection, underpinned by the development of sensor technologies, IoMT (Internet of medical things), sometimes referred to as IoT (Internet of things) in healthcare, brings diverse data sources into the healthcare ecosystem. IoMT enables wireless and remote devices to collect health and relevant data, such as environmental data, and communicate over the Internet, which is the primary dimension for personalised and integrated care.

IoMT reshapes the healthcare service fundamentally, and the changes in service provision can be seen most notably when deploying IoMT on-body, in-home, in the community, and in-hospital (Al-Turjman *et al.*, 2020). IoMT devices provide real-time and personalised data to support flexible medical data analysis and diagnosis decisions, particularly in health condition management, independent ageing, epidemic management, and chronic health condition management.

IoMT is a new category of data generation source for dynamic monitoring of personal healthcare (Mohapatra and Sahoo, 2022, Alemdar and Ersoy, 2010), and the more advanced product modes, such as smart pills, are capable of providing very precise personal health data. Technological and business innovations in this area are constantly emerging. According to the survey of a market research firm, the overall investment in IoMT market is expected to reach $185 billion by the end of 2028 (Data Bridge Market Research, 2022)

This collection of medical devices becomes an essential part of the heterogeneous profile of a patient, and it enables plenty of possibilities for improving the quality of care (Gubbi *et al.*, 2013, Pustokhina *et al.*, 2020). Chamola *et al.* (2020) summarised the role of IoMT in contact tracing, measuring respiratory rate, and telemedicine during the COVID-19 pandemic. However, the various data sources located in a large number of IoMT isolated systems bring the complexity of data standardisation, integration, and interaction. Thus, the interoperability issue is more prominent (Jabbar *et al.*, 2017, Ullah *et al.*, 2017, Ganzha *et al.*, 2017) in the IoMT embedded health information systems.

- Data transmission and exchange layer

The wireless network plays a decisive role in the Internet of Things (IoT) ecosystem. Commissioned by the Global System for Mobile Communications Association (GSMA), a report by PwC in 2015 estimates that digital health could save a potential €99 billion in healthcare costs in the European Union resulting from the increasement of connectivity through wireless networks (PwC, 2015).

The fifth generation (5G) mobile network, in particular, aims to address the limitations of previous cellular standards in terms of downlink and uplink data rate, bandwidth, the connectivity of machine-type devices, and end-to-end network latency. 5G becomes a key enabler for future IoMT by the capacities of eMBB (enhanced mobile broadband), mMTC (massive machine type communications) and URLLTC (ultra-reliable low latency communications). Telemedicine, remote monitoring, thermal imaging, and remote surgery (Chen *et al.*, 2017, Magsi *et al.*, 2018, Lacy *et al.*, 2019, Chamola *et al.*, 2020) such products and service forms, gain tremendous potential under the empowerment of 5G networks. Chettri and Bera (2019) explain the massive opportunities for IoMT through 5G new technologies, such as new radio (NR), multiple-input-multiple-output antenna, and mm-wave. 5G sits in the middle of IoMT and Cloud, supporting data transmission and exchange; therefore, in addition to fixed-line Internet, the diversity of future healthcare services brought by the 5G wireless network should be considered.

- Data storage and analysis

Cloud is another vital enabler of IoMT to make ubiquitous and on-demand data access possible and promote collaborative consultation through sharing data across organisations (Hoang and Chen, 2010, Thota *et al.*, 2018). Wan *et al.* (2022) review the related work in IoMT data analysis, emphasise the importance of fast, comprehensive, and accurate health data to quality healthcare services, especially in COVID-19 prevention and control, and discuss the Cloud, fog, and edge computing architecture for IoMT. Cloud is regarded as a distributed repository, providing the fundamental capacity in terms of storage and computing. Cloud and IoMT amalgamation provide an efficient solution for data and information exchange between heterogeneous devices and handling ever-increasing data demands in healthcare applications.

Artificial intelligence (AI) brings a paradigm shift to the healthcare domain. The algorithm-enabled analysis archives remarkable results in medical data analysis (Kooi *et al.*, 2017, Karimi *et al.*, 2020, Heller *et al.*, 2021, Liu and Guo, 2020). For example, Jiang *et al.* (2017a) discussed AI applications in early stroke detection, diagnosis and treatment. Yu *et al.* (2018) systematically reviewed AI applied in medical practice regarding clinical decision support, medical imaging, and integrated healthcare from the technical breakthrough perspective.

- Data security and privacy

Blockchain is one of the most disruptive technologies in the IT domain. Although it empowers a few controversial applications in different fields, the adoption of blockchain in healthcare is accelerated by its characteristics of decentralised mechanism to ensure the security of data and the appropriate flexibility in using privacy data (Wood *et al.*, 2016, Agbo *et al.*, 2019). The use of blockchain technology in healthcare information systems has many potential application scenarios and is of high practical value (Mettler, 2016). For example, blockchain can be used to provide access to medical data (Azaria *et al.*, 2016), control privacy (Yue *et al.*, 2016), and track the changes in patients' medical history records (Zhang *et al.*, 2018). Aloini *et al.* (2022) review the related work and summary the opportunities of blockchain in healthcare process innovation.

- Human-Machine interaction

The immersive technologies, such as augmented reality (AR), virtual reality (VR), and mixed reality (MR), provide real-life simulations to enable the user to practice different actions in a safe and controlled environment. Such immersive technologies are gradually being adopted more and more in two mainstream applications (Yeung *et al.*, 2021). One is to alleviate symptoms related to ageing and neurodegenerative problems, such as Parkinson's disease (PD), stroke, depression and PTSD (post-traumatic stress disorder); another one is in the domain of surgical visualisation and education (Pottle, 2019), such as surgical training (Kneebone, 2003), image-guided surgery, virtual patient, and robotic surgery (Zendejas *et al.*, 2013, Yeung *et al.*, 2021).

From the above discussions, it can be seen that due to the development of connection technology, communication technology, data storage and computing technology, the forms of medical services have been diversified (Rodrigues *et al.*, 2018, Manogaran *et al.*, 2018, Iyawa *et al.*, 2016), and they are still rapidly proliferating, and there is no sign of convergence at present. This kind of innovation using ecosystem synergy will continue for a long time, so the complexity of the medical ecosystem will continue to increase. In this context, it is necessary to discuss the challenges of health information systems from a holistic and harmonised viewpoint.

In the following sections, firstly, the state of the art of health information systems in the environment of new technology outbreaks will be reviewed, and then the current work related to semantic interoperability in the digital healthcare ecosystem will be discussed.

2.2 State of the Art of Health Information Systems

The health information systems (HIS), as one of the typical systems for information management, represents a wide range of information systems challenges that directly influence business decisions and quality of work, particularly the quality of healthcare (Haux, 2006). HIS is expected to have social and economic benefits by leveraging the information technologies for patients, healthcare providers, researchers and policymakers (Jennett *et al.*, 2003). Furthermore, there is considerable interest in exploring the optional of HIS (Currie and Finnegan, 2011).

HIS is recognised as the cornerstone of achieving quality of care, multi-party access and healthcare equity. The importance of HIS has been addressed by academia and industry (Haux, 2006, WorldHealthOrganisation, Sligo et al., 2017). The goals of HIS can be briefly summarised as follows:

Table 2 The goals of HIS (based on Eysenbach's seminal work (Eysenbach, 2001))

No	The goals of HIS	The 10 e's in "e-health"
1	improve the quality of medical services	**enhancing quality** of care
		evidence based - proven by rigorous scientific evaluation
		ethics - e-health involves new forms of patient-physician interaction and poses new challenges and threats to ethical issues
		equity
2	reduce the cost of operation	**efficiency -** thereby decreasing costs
3	empower patient to involve in healthcare or treatment process and to decide about the use of their data	**empowerment of consumers and patients**
		encouragement of a new relationship between the patient and health professional
4	safely and effectively use medical cases and data for multiple purposes such as scientific research, education, and service quality improvement	**education of physicians through online sources and consumers**
5	standardise information exchange and commutation in an extendable scope of healthcare ecosystem	**enabling** information exchange and communication in a standardized way between health care establishments.
		extending the scope of health care beyond its conventional boundaries

Haux (2006) identifies seven directions of HIS development over time. They are: (1) to shift from paper-based processing and storage to digital-based; (2) to evolve from local to global information system architectures; (3) to include patients as target users of HIS; (4) to extend HIS data serving scope from primary patient care to healthcare planning, and data-enabled clinical research; (5) to move the focus from technical aspects to strategic information management of HIS; (6) to broadly involve new types of data in diagnosis and research; (7) to keep steady track on new technologies and their application in the medical field, even as yet unseen.

Well-functioning healthcare, such as patient-centred care (Little *et al.*, 2001, Stewart, 2001), requires the healthcare information are efficiently managed and utilised. The advent of rapid technological development and the demands for ubiquitous environments from multiple users of HIS, including physicians, patients, healthcare payers and regulatory bodies (He *et al.*, 2019), make the HISs more complex in adopting various emerging technologies. In response to the increased connectivity and demand from multiple stakeholders, HIS are evolving into healthcare ecosystems. Such healthcare ecosystems are capable of coping with multiple domain knowledge (Blobel, 2019), especially the heterogenous data collected by emerging medical devices or sensors. From the use and management of healthcare data point of view, in order to support advanced analytics, the availability, accessibility, flexibility, and quality of healthcare data are demanded by innovators in the healthcare domain.

HIS is a typical social technology to cope with clinical, technical, and cultural issues under different laws and regulations (Chaudhry *et al.*, 2006). In the face of rapidly expanding healthcare ecosystems, increasingly stringent supervision and regulation, and continuous improvement in the quality-of-care services, many technologies have been deployed. For example, in response to the widely recognised security and privacy issues of patient data, distributed storage, and federated learning (Yang *et al.*, 2019) working together can maximize the value of clinical data while ensuring data security and protecting user privacy. The support of these technologies means that heterogeneous medical information systems do not necessarily need to be physically integrated in order to exchange information.

However, after decades of development, digital health technologies cannot fulfil all the requirements of the stakeholders within the healthcare ecosystem because the digital healthcare ecosystem is a social technology network, which is dynamic and synergetic complex. As addressed by the healthcare ‘iron triangle’ (Kissick, 1994), including three competing elements: access, quality and cost containment in the healthcare ecosystem, cannot be improved at any angle without compromising the other two. The adoption of IT in healthcare is particularly slow and has been lagging behind that of the leading industry for around 10-15 years (Goldschmidt, 2005). This result is mainly attributable to the failure of HIS implementation (Heeks, 2006) and resistance to using new technologies by healthcare professionals (Blumenthal and Tavenner, 2010). Tummers *et al.* (2021) summarise the obstacles to the low adoption rate of HISs in five points. (1) The technical issues caused by the mediocre design of HISs are attributed to the inadequate infrastructure of information and communication technologies. (2) The poor system usability results from low user satisfaction and lack of training on new systems. (3) The interoperability problems block HISs adoption and use on a scale. (4) The operational functionality provided by HISs cannot meet user demands due to the rapid evolvement of Internet technologies and the increasing complexity of health care. (5) The system maintenance and support issues are caused by the lack of professionals and poor documentation of HISs. Based on the statistics (Tummers *et al.*, 2021), nearly two out of three low adoptions of HISs are related to technical issues and poor system usability. Therefore, there is a significant gap between the state-of-the-practice and the state-of-the-art.

This paper aims a better design of the architecture of HIS, echoing the first obstacle identified by (Tummers *et al.*, 2021), to establish a unified information exchange and communication platform.

2.3 Health Informatics and Interoperability Challenges

Health informatics is a field of science and engineering that aims at applying information technologies for the acquisition, processing, analysis and study of health data to support diagnosis and clinical decision-making (Imhoff, 2002). Health informatics is a wild spectrum of multidisciplinary fields that includes studying the design, development and applications of computational innovations to improve health care (Nadri *et al.*, 2017). Accordingly, health informatics can be considered a discipline that studies applied technologies rather than gains new knowledge.

In the early 1960s, Electronic medical record systems (EMRs) came out for the first time in the Health Information System (HIS) created by IBM in Akron Children's Hospital, and then the ambulatory electronic medical record systems outside hospitals were deployed in the late 1960s. The two cases are regarded as the earliest adoption of health informatics in practice (Braunstein, 2018). Another crucial early adoption is the clinical decision support system (Shortliffe, 1976). With the rapid development of computer technology, health informatics is a fertile domain in academia and industry.

As multiple entities are involved in the clinical data exchange, health information exchange (HIE) is designed to enable clinical information can be moved among disparate healthcare information systems and consolidated and accessed on demand. Kuperman (2011) describes the primary rationale for HIE is to address the critical healthcare problems that 'siloed' healthcare information systems do not solve. Thus, healthcare data sharing and information exchanging can be regarded as the foundation for improving the quality and efficiency of healthcare services from the perspective of applying health informatics.

Underpinned by HIE, ideally, all patient-related data can be organised around this patient. The relevant workflows connect multiple organisations, like a social network, but the patient is in the centre of the web. Baker (2001) points out the differences between the traditional and patient-centric healthcare from the following eleven perspectives.

Table 3 The comparison between traditional and patient-centred healthcare models (based on Institute of Medicine 2001 (Baker, 2001, Benson and Grieve, 2016))

Aspect	Traditional healthcare	Patient-centred healthcare
Focus of care	Discrete visits/episodes Clinician makes diagnoses independently on what investigations and treatment to order	Continuous healing relationships Health care system should be always responsive (24 hours/7days)
Variation mainly due to	Professional autonomy	Patient needs and values Have the capability to respond to individual patient choices and preferences
The controller in	Professionals	Patients Health information systems accommodate

Aspect	Traditional healthcare	Patient-centred healthcare
treatment		differences in patient preferences and encourage shared decision making
Information sharing	Partially shared	Fully shared Unfettered access
Decisions based on	Professionals' training and experiences Care may vary illogically	Evidence-based decision making Care is consistency
Safety	Individual responsibility	A system property
Transparency	Partially	Fully Transparent subject to patient privacy
Reactivity	React to patient needs	Anticipate patient needs
Operational focus	Decrease cost	Eliminate waste
Collaboration	Demarcation	Cooperation
System architecture	Silos	Interoperability

The patient-centred healthcare describes the evolving direction of HISs, and the fundamental prerequisite is system interoperability (INTEROP NoE, 2007). Janett and Yeracaris (2020) present the challenges of a single EMR in U.S. primary care and stress the importance of interoperability to enable disruptive technologies and applications in healthcare information systems. Noon *et al.* (2021) point out that the lack of interoperability among disparate EMRs is one of the identified challenges of HISs applied in scale and interoperability is a crucial application scenario for blockchain solutions. Kruse *et al.* (2018) conducted a systematic literature review with 55 primary studies on how electronic health records support public health and identified that interoperability is listed as the second top barrier following clinical data missing. Because the interoperability issue prevents stakeholders from accessing patients' records, thereby increasing operational costs (Iroju *et al.*, 2013). Moreover, medical errors could be caused by inadequate availability of patients' information (Walker *et al.*, 2005).

West Health Institute (2013), a non-profit medical research organisation, specifically scanned the interoperability between medical devices and patient data repositories and device-to-device interoperability in U.S. hospitals. They estimated in 2013 that the problems caused by the lack of medical device interoperability cost $35 billion annually. For example, the medication errors in drug ordering, order transcription, drug dispensing and administration accounted for around $400 million. Such adverse events, including diagnostic errors and failure to prevent injury, costs a potential $2 billion in total. In addition, the redundant testing resulting from inaccessible information costs $3 billion; increasing clinician productivity and capacity of the treatment to shorten the length of stay can save $ 12 billion and $18 billion, respectively.

2.4 Interoperability of Healthcare Ecosystem

With the introduction of ambient intelligence in healthcare, digital environments are more adaptive and responsive to making self-care and precision medicine a reality. Significantly as ambient intelligence (AmI), mainly represented by IoMT, is growing tremendously and promising the pervasive diffusion of intelligence in our surrounding environment, the complexity of the healthcare ecosystem grows exponentially. The need for robust data pipelines to communicate across disparate health systems and devices is becoming increasingly urgent. This chapter will review healthcare ecosystems' interoperability standards, especially semantic interoperability.

2.4.1 Interoperability Standards

According to the definition of interoperability standards made by the healthcare information and management systems society (HIMSS), there are four levels of interoperability in the healthcare domain. The **foundational level** establishes the inter-connectivity requirements for securely exchanging data between applications and systems or among systems. The **structural level** defines the format, syntax, and organisation of data with the interpretations at the data field level. The **semantic level** provides standard underlying models and codifications of data, upon which consensus has been reached so that the users could share understandings and meanings of data through them. The **organisational level** sets up operational rules to facilitate secure, seamless, and timely data exchange within or between organisations, entities, and individuals. The operating rules include data governance, data usage policy, and social, legal and organisational considerations (HIMSS, 2022). These four levels of interoperability reflect the understanding of how clinical data is organised in healthcare information systems from the application perspective. As a result, various interoperability standards have been formulated by different organisations driven by the need for diversified clinical data application scenarios. According to HIMSS, there are over 40 standards development organisations (SDOs) in the health IT arena. Some SDOs develop standards, such as health level seven (HL7), systematized nomenclature of medicine (SNOMED) international, and the clinical data interchange standards consortium (CDISC). Other SODs develop specific functions or use cases as complementary standards over these fundamental standards, which promote the adoption of the fundamental standards. In the following sections, HL7 and its promulgated standards will be introduced in detail.

The fundamental standards developed by different SDOs have varying compositions and processes to meet the specific industry or market requirements yet generally follow the design principle of public standards. From the application perspective, these standards can be divided into five categories, including vocabulary/terminology standards, content standards, transport standards, privacy and security standards, and identifier standards (HIMSS, 2022). These standards unify medical terminologies, specifications for medical records, interaction methods and related management specifications for using these medical records. However, from the perspective of information systems, the fundamental standards mix the syntax, semantics, and even pragmatics of clinical data. For example, the clinical document architecture (CDA) intends to specify clinical documents' encoding, structure, and semantics for exchange, providing a framework for defining the full semantics. The framework contains terms related to medical activities, e.g., allergies, care team, encounters, family history, functional status, health concerns, immunizations, interventions, and medical equipment, used to describe the medical practice in detail. According to the HIMSS classification, the CDA belongs to the content standard. While from the perspective of informatics, the CDA unifies both the syntax (the format of data, e.g., string, integer, float) and the semantics (the meaning of concept). The syntax used for data interaction in information systems is relatively straightforward and has little ambiguity, while semantics is context-dependent, and the ambiguity often appears in the use process. As addressed by Liu *et al.* (2009), the code/value pairs of the observation element may have many different meanings depending on the context of the usage. The following sections will review the clinical standards from the information systems perspective to distinguish their applications at the syntax, semantic and pragmatic levels, respectively.

2.4.2 Clinical Standards of Semantic Interoperability

Regarding the request for collaboration between multiple dispersed health information systems, the semantics exchange turns out to be one of the fundamental requirements for health information exchange. Semantic interoperability (Geraci *et al.*, 1991, HIMSS, 2014) ensures the seamless information exchange in terms of interpreting the data in the same manner across any system or device regardless of its proprietary architecture. Semantic interoperability plays a vital role in the healthcare ecosystem, which enables the ubiquitous forms of information shared among disparate systems.

Semantic interoperability is the pre-condition of adopting advanced medical information technology (Lehne *et al.*, 2019a) to improve healthcare quality. For instance, health sensing, big data analytics and cloud computing, these three categories are promising technologies in future health information systems summarised by Yang *et al.* (2015). All of these require data and information exchange without ambiguity. However, the standard which is supposed to address the issue of semantic interoperability amongst heterogeneous systems of the digital ecosystem still remains uncertainty (El-Sappagh *et al.*, 2018). Currently, several fundamental standards are proposed by different SDOs with concerns from various perspectives. For instance, OpenEHR (openEHR, 2002) specifies the information system architecture required for interoperable communications between systems and services. SNOMED-CT (Systematized Nomenclature of Medicine Clinical Terms) unifies a glossary of clinical vocabulary used in EHR across medical institutes. The logical observation identifiers naming and coding (LOINC) system is an international coding system for identifying health measurements and clinical observations, which provides a precise clinical vocabulary to make data portable across health systems. The HL7 V3 clinical document architecture (CDA) defines the structure and semantics of clinical documents between hospital and patient. In addition, the clinical information modelling Initiative (CIMI) is a shared repository of detailed clinical information models defined by the HL7 workgroup to improve the interoperability between health systems.

In detail, the standards like SNOMED-CT and LOINC are ontologies of reality, which describe the clinical terminologies' independence on any information system. The others describe the ontologies of clinical information, which are organised by different approaches reflecting the various application scenarios. Most of them are mutually overlapping or exclusive to some extent. In summary, the ontologies of reality can be employed by any ontology of information. However only one ontology of information is normally adopted by a single information system.

Even though the standards mentioned above all claim that they are developed for semantic interoperability, they describe more than one level of interoperability defined by semiotics (Liu and Li, 2015).

2.4.3 The Levels of Semantic Interoperability

When the healthcare ecosystem is treated as an integrated information system, the level of interoperability can be explained by the framework of the semiotic ladder (Liu and Li, 2015). According to the theory of semiotics, Charles Peirce interpreted objects by the triadic world method (Pierce, 1998), which uses a tuple (Signified, Interpretant, Object) to define an object with its contextual information. Applying this philosophical approach to investigate the semantic interoperability information system, we can assume that Ii and Ij are two independent health systems that will communicate semantic information about the two objects Oi and Oj presented by the signs Si and Sj respectively. The ambiguity that might happen during the process of communication between Ii and Ij are:

1. Empirical ambiguity: The format of information sent by Ii cannot be translated by Ij. Therefore, the two health systems cannot exchange information due to the absence of information.

2. Syntactic ambiguity: There are two possibilities here leading to ambiguity. a) Two syntactically equivalent signs signify different objects. AcronymFinder (AFwebpage) provides 34 expansions of the acronym "BT" in the context of Science and Medicine, including Bacillus Thuringiensis, Behaviour Therapy, Blood Transfusion, Blood Test, Blood

Type, Bleeding Time (test of blood), Borderline Tuberculoid etc. b) The signs, Si and Sj, for the same objects are different. Examples can be found in the use of different languages (English and Chinese) or in different narrative medicine systems (Western medicine and Chinese medicine).

3. Semantic ambiguity: Two syntactically equivalent signs signify the same object but have different meanings within different contexts. For example, a child's heart beats faster than an adult's. Thus, judging whether a heart rate reading is in the normal range depends on the test subject's age.

4. Pragmatic ambiguity: In an information system, pragmatic interoperability refers to a system that can process and understand the information sent by another system and respond to the query with correct information. Pragmatic ambiguity arises when a medical diagnosis model has been established, but the diagnosis results fail to be obtained by entering relevant parameters.

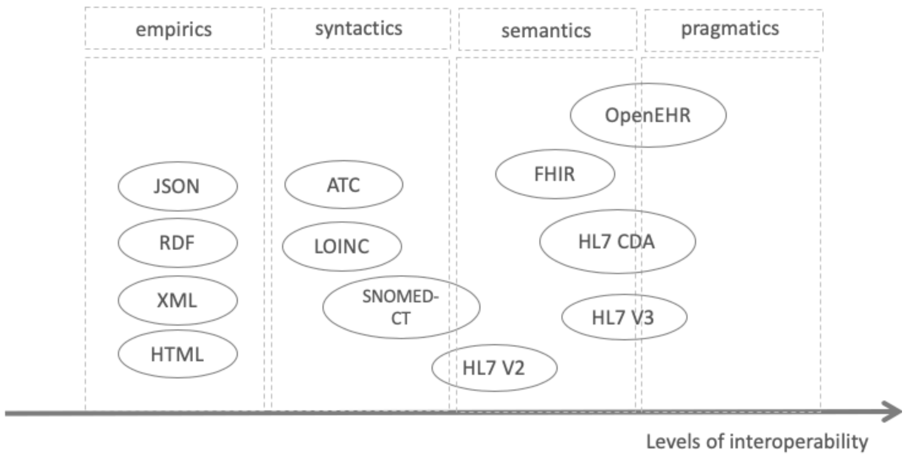


Figure 4 The interoperability levels of clinical standards

As shown in Figure 4, this study classifies the published standards in healthcare domain according to the level of interoperability. From left to right, the horizontal axis indicates the operability level gradually increasing, from empirics to pragmatics.

The definition of data format is concerned with interoperability at the level of empirics. Standards such as HTML, XML, RDF and JSON provide formats which can be used in the IT industry for data interchange, which are located on the far left of the horizontal axis. The level of syntactic defines the clinical terminologies, including the anatomical therapeutic chemical (ATC), LOINC, and SMOMED-CT, which are the coding systems to indicate organs, drugs, or symptoms. At the semantic level, the standards are used to describe the clinical operations or the healthcare services, which are context-dependent and with more ambiguities than the standards at the syntax level. HL7 series standards and OpenEHR are located at this level. At the pragmatic level, the OpenEHR, HL7 CDA and HL7 V3 provide some functions of the medical diagnosis model. Therefore, these three standards cross the semantic level and pragmatic level.

This study focuses on semantic interoperability, and as FHIR (Fast Health Interoperability Resources) is the latest version of HL7, the following section compares the FHIR and OpenEHR with the purpose of identifying a research base.

2.5 OpenEHR vs FHIR

OpenEHR and FHIR are two mainstream standards adopted in the industry. Both are designed for full interoperability of the digital health and social care ecosystem. FHIR and OpenEHR have massive open-source implementations and a strong supportive community for implementors, which may make them possible to become the universal standard of the future medical industry.

2.5.1 Introductions of OpenEHR and FHIR

The OpenEHR specification defines a completed architecture of EHR, including the format of data structure, the canonical information model, the query language, and the APIs for external applications regarding the increasingly complex healthcare ecosystem. In contrast, FHIR (HL7, 2022b) has a narrow scope, specially designed for semantic interoperability.

The core idea of the OpenEHR specification is to separate the medical domain knowledge from the specific clinical information through a two-level modelling approach. This approach divides models into the archetype model (AM) and the reference model (RM).

The RM defines the stable and unchanging concepts in the information system, and the basic data types and structures required for information expression. The AM includes archetypes and templates. The archetype defines clinical content and expresses domain knowledge by adding constraints to the RM; the template meets practical application requirements by constraining and customizing the archetype. By separating the RM and the AM, the dependence of the information system on the domain knowledge is effectively reduced so that the system based on OpenEHR can adapt to the changes of the domain knowledge and has the characteristics of reusability, scalability, and easy maintenance.

The separation of RM and AM seemed to be aspirational because developers and clinicians found themselves could work together regarding both clinical and technical concerns through modelling the RM and designing AM. The architecture of OpenEHR is a best-of-breed framework that aims for all forms of interoperability in incumbent systems. However, its implementation, particularly in large-scale projects, is challenging and complicated because modelling the domain knowledge was too complex to adapt to various end-users (Christensen and Ellingsen, 2016).

FHIR alternatively selects another approach to handle this issue, which aims to simplify the implementation process without losing information integrity. FHIR starts from an engineer's point of view to consolidate all possible categories of data by leveraging the existing models. FHIR also provides a widely adopted IT mechanism (RESTful interface) for exchanging health data between healthcare applications.

FHIR can be regarded as a shared schema with rich data representation in the digital healthcare ecosystem. It works as a communicational protocol for the systems and devices to exchange healthcare data. Therefore, the entities can understand each other only if they follow the protocol defined by FHIR. Hence, FHIR fills the semantic interoperability gap between heterogeneous healthcare entities, through which the systems or the devices in the ecosystem can understand each other.

2.5.2 Comparison between OpenEHR and FHIR

The significant difference between OpenEHR and FHIR reflects in the design philosophy. OpenEHR takes a top-down approach, which solves the problem of semantic interoperability from the perspective of system architecture. In comparison, FHIR offers an alternative approach starting from the bottom, starting with data already in the ecosystem and specifying the data schema to support exchange between health systems (McNicoll *et al.*, 2019). The two distinguished design philosophies determine different ways of adoption in the industry.

From the statistics of the number of published papers, it can be identified that FHIR has received significantly more attention than OpenEHR (Lehne *et al.*, 2019a). The reason is that FHIR is more practical and has less impact on existing systems. The following (Table 2) will compare the two standards from the following five aspects: technology architecture, openness, scalability, flexibility and portability (Blobel *et al.*, 2006).

Table 4 The comparison between OpenEHR and FHIR

Standards / Aspects	OpenEHR	FHIR
Technology architecture	Complete and comprehensive system architecture; a top-down approach	Data schema for information exchange between systems; a bottom-up approach
Openness	RM (Reference Model) is compatible with any clinical data model, and CDR (Clinical Data Repository) is flexible to any language and database. System architecture, software stack and its applications are closed source.	Internet-based standard; Providing standard RESTful API as an interface for Web services.
Scalability	SOA architecture, dealing with semantic scalability by defining RM	Leave 20% extension of resources to deal with semantic scalability
Flexibility	CDR can be easily exchanged among different vendors regardless of the database or language in use	RESTful API + OAuth2 for all web services and mobile applications
Portability	High portability, embedded with specific query language (ADL)	High portability, without query language

1. Technology architecture

OpenEHR is a typical standard that adopts model-driven development (MDD) approach (Selic, 2003, Raghupathi and Umar, 2008). All architecture specifications of OpenEHR are presented as a set of abstract models with UML notations. The medical concepts in OpenEHR are defined by RM and AM hierarchy, which is called the two-level modelling approach (Beale, 2002). RM defines data types and structures, and AM describes medical concepts through RMs. For instance, RM defines “Quantity” with numbers and “Text” with strings. The medical concept “Blood pressure” can be expressed with four “Quantity” terms - systolic, diastolic, mean arterial pressure and pulse pressure, and one “Text” term to indicate the name of this terminology - comment. The RM/AM design method meets the basic idea of Service-Oriented Architecture (SOA) (Arsanjani, 2004), which is broadly adopted by IT companies and has been proven to be a successful architecture. Accordingly, it shows the consistency of OpenEHR as a digital transformation standard for medical systems and mainstream IT design ideas.

Similarly, the previous version of FHIR, the HL7 V3, did follow the top-down MDD design. However, it failed to scale because of its high complexity and high cost of implementation (Bender and Sartipi, 2013). In order to be more practical and implementable, HL7 FHIR (HL7 International, 2011a) intentionally adopted an incremental and iterative approach to transforming industry best practices into standards. Easy implementation is the critical reason for FHIR becoming the leading standard (Lehne *et al.*, 2019a) for interoperability. In Section 4, FHIR is discussed in detail.

In summary, from the architecture design point of view, the OpenEHR standard is advanced in system design and converges with Internet best practices. In comparison FHIR is easy to implement and is widely adopted by the industry.

2. Openness

The openness of the OpenEHR standard is polarized. On the one hand, it is incredibly open. The Reference Model (RM) supports and is compatible with any OpenEHR clinical data model or archetype. Also, the clinical data repository (CDR) can be described in any computer language, stored in any type of database, and defined by any vendor, as long as it complies with RM specification defined by the OpenEHR (McNicoll *et al.*, 2019)

Furthermore, the CDRs can be easily swapped between any CDR vendor systems without data migration implementation (Atalag *et al.*, 2016). On the other hand, it has obvious limitations; the OpenEHR clinical system and its applications, as well as the software stack, are not open-source to the implementor.

FHIR is a repository of medical resources, which appears as a communication format for organising data. It can be adopted by any system or device, and therefore, FHIR is usually used as an intersystem communication protocol in the digital healthcare ecosystem. For instance, FHIR can be used to bridge the OpenEHR and non-OpenEHR systems (McNicoll et al., 2019). In order to increase openness, FHIR adopts RESTful (Resource Representational State Transfer) API (HL7 International, 2011b) to communicate with other systems and applications. RESTful is widely used in Web services, through which FHIR can flexibly interconnect with any application on the Internet. SMART (Substitutable Medical Applications, Reusable Technologies) on FHIR (SMART, 2009) is an excellent example to show the openness of FHIR. Based on FHIR, SMART enables healthcare providers to utilise more apps underpinned by the existing digital healthcare systems and empowers software developers to create more apps to meet the needs of a wider audience (Mandel *et al.*, 2016). Furthermore, a few third parties deploy FHIR on the cloud for test and sandbox implementations (HL7 International, 2019b) to engage broader developers and cultivate the innovation environment. To summarise, in the aspect of openness, the FHIR standard does better than the OpenEHR.

3. Scalability

Scalability refers to the capacity of a system that can gracefully grow to manage the increased demand. The OpenEHR standard offers a genius architecture, two-level modelling (openEHR, 2007), which separates data instances from the clinical domain knowledge and guarantees the flexibility to face the demands of medical development and software development. The architecture of OpenEHR is hierarchical modelling within a service-oriented software framework ensuring enough scalability to handle the services at different granularities (Khan, 2010).

FHIR is designed with the 80/20 rule to describe the clinical resources. Extensions and customizations are allowed to exist in the FHIR profile to adapt to the needs of specific use cases. Also, the RESTful API of FHIR ensure the system's horizontal scalability from the angle of system architecture design (HL7, 2011a).

The comparison shows that OpenEHR has a structured advantage and fundamentally solves the scalability problem. FHIR takes a relatively simple approach; extensions and RESTful API guarantee system scalability.

4. Flexibility

OpenEHR's flexibility is guaranteed from the level of design philosophy, which allows CDR to be easily communicated among different vendors regardless of the database or language used.

Form release 3, FHIR started to support RESTful API and concurrently brought OAuth2 into use, which means the FHIR healthcare system can authorize apps through the SMART protocol on a large scale (Alterovitz *et al.*, 2015, Wagholikar *et al.*, 2017). OAuth2 + a standard API provides sufficient flexibility for developers of medical applications (Mandel *et al.*, 2016, Warner *et al.*, 2016, Bloomfield Jr *et al.*, 2017, Wagholikar *et al.*, 2017).

In short, the OpenEHR and FHIR are comparable in the flexibility perspective.

5. Portability

The path mechanism of OpenEHR defines an archetype-based X-path compatible syntax, which enables the health information to be queried by Archetype Definition Language (ADL). This mechanism is one of the distinguishing features of OpenEHR. However, FHIR focuses more on information exchange and lacks a systematic view to solve the semantic issues through a complete system architecture.

Chapter 3 Research Methodology and Design

This chapter elaborates on the research paradigm and the methodology adopted, with a particular focus on the philosophical stances of this research. The gap between the observed world and the laws revealed via scientific activities motivates the author to rethink the approach to gaining knowledge in the domain of information systems. The characteristics of information systems are studied and analysed from the perspective of the theory of signs, i.e., semiotics. Furthermore, as the dependence of information systems on context and culture, the research approaches adopted by different IS research are compared and discussed. Lastly, this chapter details the adopted paradigm, approaches, methods, and techniques in this research.

3.1 Empirical Observations

Scientific research aims to explore the observed world and discover the laws and theories that can explain and predict natural or social phenomena (Horrobin, 1969). This process also can be defined as knowledge acquisition that gradually improves as people deepen their understanding of the observed world. However, due to the distance between our understanding and the observed world, the theories or laws we use to understand or predict phenomena should only be reckoned as postulate theories or working hypotheses. The suitable premises of theories and laws may change because people's understanding of the observed world may differ with the changes in scale and dimensions of observation. It occurs not only in the field of social sciences but also in natural sciences. For instance, classical mechanics based on the foundational works of Sir Isaac Newton can describe the motion status of physical bodies at the macroscopic level (Stump, 1990). The basic assumptions of classical mechanics are that space is continuous and time is absolute; while the effects of gravity on the fabric of space-time must be considered when studying extremely massive objects according to Einstein's theory of general relativity (Sciama, 1964).

The general theory of relativity made a quantitatively minor modification to Newton's theory but qualitatively changed people's view of time and space.

Conversely, only quantum mechanics applies if studying the physical properties of nature at the scale of atoms and subatomic particles. Currently, the Grand Unified Theory (Ellis, 1986) is trying to use a unified theory to explain the observed world at a large and micro-scale. Still, the Grand Unified Theory is a hypothesis, and there is no experimental evidence to support this theory yet. Thus, the scale of the physical space where the research object is located is the vital premise in physics research.

In the domain of social science, the study objects are highly connected with cultural and contextual settings and more objective understanding involved. Thus, the research conclusions in social science are heavily dependent on their research premises. Both positivism and interpretivism will inevitably introduce subjective understanding into research. For instance, in positivist research, distinguishing dependent and independent variables, qualitative data collection and experimental controls all entail certain subjective understandings. Comparatively, interpretivist research investigates multiple realities through interactions between researchers and participants. Due to the fact that information systems are highly intertwined with social contexts (Chaudhry *et al.*, 2006), multi-reality unquestionably exists in the research of information system domain. Consequently, there is no universally applicable solution in the design of information systems. Continuous development based on the best solution available could be a more feasible strategy for IS continuous improvement.

In addition, the understanding of culture and the background of research questions are subjectively influenced by the language researcher used. Especially in the field of social science research, the impact of language on scientific research should be re-evaluated as language is a sign system for communication where semantic ambiguity inhabits. Hence, there is no entirely objective scientific research activity; therefore, scientific research results may only be viewed as workable hypotheses within current cognitive settings, and there is always room for improvement. In other words, each existing information system solution has the potential to be enhanced. These empirical observations on philosophical issues motivate me to examine the philosophical stance of my research.

3.2 Research Paradigms

Research paradigms are different philosophical ways of thinking (Kuhn, 1962) reflected by a set of commonly held beliefs, values, assumptions and practices within a research community (Johnson and Christensen, 2019). This whole set of philosophical stances significantly affect the manner of conducting research. For a specific research problem, it can be understood or interpreted from numerous perspectives, and its relevant research points can also be addressed from different angles. The adopted research paradigm can precisely reflect the way of researcher's understanding of knowledge acquisition, the angle of addressing the research question and the relationships between the research problem and existing theories.

A research paradigm is defined as an intellectual structure and its underlying assumptions upon which research and development in a field of inquiry are based (Kuhn, 1962). A paradigm constitutes a world view that structures the research approach and influences how a research community perceives their field of study. The underpinned beliefs and assumptions can be characterised by three philosophical concerns: ontology, epistemology, and methodology (Guba, 1990).

Ontology is a philosophical study of the nature of reality and aims to identify the specifications of concepts and the relations among them (Smith, 2012). Epistemology addresses the nature of valid knowledge and how people communicate knowledge to each other (Burrell and Morgan, 2017). Methodology is the systematic study of methods to guide the implementation of research within a discipline. In Summary, ontology and epistemology create a holistic view of how knowledge is extracted from this world and how ourselves in relation to this knowledge; the methodology is the strategy to discover knowledge. These three philosophical stances help researchers: 1) to define how the world works, how knowledge is extracted from this world, and how one is to think, write, and talk about this knowledge; 2) to define the types of questions to be asked and the methodologies to be used in answering; 3) to decide what is published and what is not published; 4) to structure the world of the academic worker; 5) to provide its meaning and its significance (Dills and Romiszowski, 1997).

3.2.1 Positivism and Interpretivism Paradigms in IS

The discipline of Information Systems (IS) focuses on solving problems in businesses through the use of information and communication services. IS is established in social organisations and is significantly impacted by social science (Mingers, 2004). The IS research underpinned by different research paradigms addresses various concerns (Goldkuhl, 2012). Some research aim for technology development to promote the reuse of solutions for various requirements; or for the understanding of 'existing meaning systems shared by the actors' (Orlikowski and Baroudi, 1991) and the social and historical context of this system; or for the intervention and change (Braa and Vidgen, 1999). In IS research, positivism and interpretivism are the two mainly adopted paradigmatic views (Chen and Hirschheim, 2004). In positivism, social phenomena can be quantifiably measured by variables and possibilities. The purpose of positivist research is to reveal the inherent causal relationships and explain the cause of behaviour (Orlikowski and Baroudi, 1991). However, interpretivism reckons social phenomena is intentional, which depends upon meaningful actions of individuals. Interpretivist seeks the meaning of actions (Davoudi, 2020). As IS seats in the middle of social and natural sciences, the debates of research methods are located in positivism and interpretivism. The comparison between positivism and interpretivism from the ontological, epistemological, and methodological perspectives has been shown in the following table.

Table 5 Comparison between Positivism and Interpretivism Paradigm, Adapted from (Chen and Hirschheim, 2004, Vaishnavi and Kuechler, 2004)

Philosophical assumption	Positivism	Interpretivism
Ontological	A single reality Existing objectively and independent from human experiences Knowable Logical reasoning Probabilistic	Multiple realities Socially constructed Containing subjective meanings and can be constructed or reconstructed through a human and social interaction process (Burrell and Morgan, 2017)
Epistemological	Objective Dispassionate Sensory experience Detached observer of truth	Subjective Values and knowledge emerge from researcher-participant interaction by which the subjective meaning of the reality is constructed (Walsham, 1995)

Philosophical assumption	Positivism	Interpretivism
	Concerns with the hypothetical-deductive testability of theories	
Methodological	Observation Quantitative Statistical Replicable	Qualitative Interpretive knowledge Need to engage in the social setting investigated and learn how the interaction takes place from the participants' perspective by adopting qualitative research (Howe, 1988) Focusing on the transferability instead of generalisation

Positivism believes knowledge can be obtained from observational data interpreted by scientific methods. Knowledge value-neutrally exists, and any subjective factor cannot intervene in investigations (Henderson, 2011). They believe truth is an independent part of the whole and understand factual accounts of phenomena from possibilities calculation. Accordingly, the quantitative, cross-sectional, and survey-oriented methods are adopted by the positivist approach to measuring the social world. Positivist IS researchers assume the physical and social world exists objectively and is independent of human experiences. As such, its nature can be apprehended, identified, and measured. When a set of descriptive studies has been identified in the positivist tradition, the problematic nature of the studied phenomena is hard to be recognised. That is why the positivism approach is criticised for ignoring the complexity of a social system Buttery and Buttery (1991). Positivist IS researchers are attempting to restore a social system by a limited number of variables and describe it through the relationships between variables. The most important thing is that positivist IS researchers neglect the adaptability of the social system, which is driven by the learning ability of the participants in social organisations. The results delivered by the positivist approach often lack the ability to accommodate the changing facts. Furthermore, the implicit assumption of positivist IS research is the ideal environment in which positivist researchers can abstract laws that can be widely applied to many phenomena. In short, the positivists are committed to identifying time- and context-free generalisations or nomothetic statements that can be multiply applied (Keat and Urry, 2011, McKenna *et al.*, 2011). However, in social science, the static and ideal environments never exist, and the prescriptions cannot be multiply applied.

Conversely, interpretive research is based on the assumption that the social reality is not singular or objective but rather shaped by culture, social context, the use of language, and human activities. Interpretive researchers examine the phenomenon of interest within cultural and contextual situations, and thus various subjective interpretations can be reconciled in their socio-historic context. Rather than seeking to determine the time- and context-free generalisations, the interpretivist seeks to determine meanings and reasons in time- and context-bound (Geertz, 2008) to restore the complexity of human "sense-making" in social activities.

Information system (IS) is a typical multi-paradigmatic research discipline. With the evolution of information technology, the term IS has been dramatically changed. IS started primarily as a dependent system that automates repetitive work and gradually evolved into a part of a social system that penetrated all aspects of daily life. In this context, IS can be considered a socio-technical system that heavily involves the interactions between human behaviour and technology in social groups. The emergence of adopting interpretivism in IS can be traced back to the 70s (Boland Jr, 1979), and Zuboff (1988) emphasises the social implications of IS. After that, interpretivist researchers gradually occupy a significant proportion of the field of IS research (Orlikowski and Baroudi, 1991, Walsham, 2006, Sarker *et al.*, 2013). IS reflects the designer's understanding of the world through computer-based systems, in which new conversations, connections and commitments are produced (Walsham, 1995). One IS reflects one perspective of the multiple realities which is mentally perceived by a designer. Therefore, the use, design and study of IS can be understood as a hermeneutic process (Boland Jr, 1979, 1985) of applying information technologies in human society, in which the creation and interpretation of symbols are involved. The functionalities of language and symbols in IS will be discussed in the next section.

To summarise, the difference between positivism and interpretivism in IS lies in the philosophical assumptions adopted by IS researchers. Positivist researchers believe an IS can be abstracted from realities and applied to multiple scenarios. They aim for generalisability and replicability. While interpretivist researchers believe an IS is a time-and context-bound reality. They focus on the in-depth understanding of the phenomenon examined from a multidimensional view and acknowledge the reality of multiple realities.

3.2.2 Inductivism and Falsificationism

The interpretivist researchers often qualitatively adopt inductive reasoning in research (Denzin and Lincoln, 1995) in IS to explore and understand the system requirements in organisational, social and cultural contexts. They try to develop wider generalisations from specific observations, and the process of research is to detect the patterns and regularities led by tentative hypotheses (Williams, 2000). The inductive method emphasises the accurate knowledge acquisition from a great number of observations, and as the number of observed cases approaches infinity, the tentative hypothesis is more likely to be accepted as a general (Russell, 2001). However, David Hume (1711-1776) challenges the basis of inductive inferences as confirmatory evidence cannot provide full certainty, which is commonly known as a logical fallacy and exemplified by an example of black swans. Thereafter, Karl Popper, as a major critic of inductivism, refutes the classical positivist account of the scientific method by replacing induction with the falsification principle. The following two figures feature the research processes of the inductive method (Figure 5) and the falsificationist method (Figure 6).

As shown in Figure 5, inductive researchers start from a large number of observations or experiments regarding particular phenomena, identify the patterns by data analysis, and make generalisations that account for every instance. The extraction of generalisation from evidence is the target of inductivist research. From the view of inductivism, knowledge may be obtained by this method.

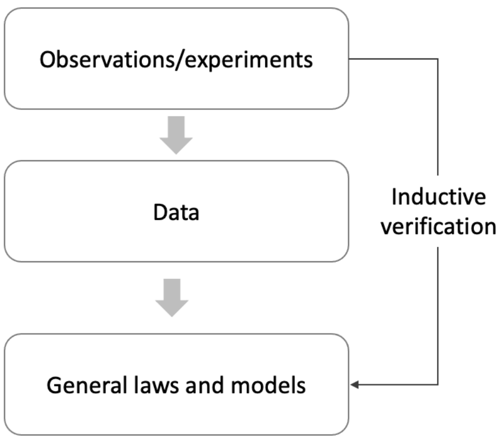


Figure 5 Inductive research process (Pietsch, 2021)

Karl Popper was one of the critics of the inductive research process. He accepted that Hume's problem could not be circumvented and proposed criteria of demarcating scientific statements from other forms of intellectual inquiry, which is named falsification. Falsificationism argues the philosophy of science about the validity of induction as no amount of positive evidence can ensure that negative instances may be found. From the view of the falsificationist, there are no theories that can be considered as correct eternally, merely the least wrong. All scientific knowledge is provisional and conditional, and a scientist can only attempt to approach the truth with bounded context and never can assert what the truth is. In other words, if a theory is to be considered scientific, it must be able to be tested and conceivably proven false. Thus, fallibility is regarded as a criterion for demarcating science from non-science.

Based on Popper's theory of falsification, Magee (1973) developed a process of falsificationist research, shown in Figure 6. It suggests that scientific research should start from deductive reasoning to identify the flaw or limitation of the existing theory and aim to address the flaw or limitation, and then propose a new postulate which should be testable. The third step is to articulate the testable proposition and then test. The last step is to select the least-wrong theory based on reasonable inferences from competing theories.

Table 6 compares inductivism and falsificationism from the five perspectives, which reflect the key differences between the two research paradigms for conducting scientific research. More and more researchers argue the appropriateness of applying inductivism in IS science and are inclined to adopt falsificationism (Farhoomand, 1987, Chen and Hirschheim, 2004, Salovaara and Merikivi, 2015, Lee and Hubona, 2009). Falsificationism emphasises that scientific knowledge is provisional and context-bound. Lee (1989) addresses information systems are context sensitive and expresses a similar understanding of the falsificationist approach to Magee (1973). Lee (1989) summarises four steps of applying falsificationism in the case study research in information systems. Gregor (2006) also discusses the applicability of falsificationism in information systems design. Falsificationism seeks the boundary of existing theories and increases knowledge by re-examining published studies.

This objective of falsificationism can meet the aim of reusing software architecture in IS domain (Garlan *et al.*, 1995). Thereby falsification as a scientific research method is increasingly adopted in the study of information systems or software system design (Salovaara and Merikivi, 2015, Lee, 2004, Lee and Hubona, 2009).

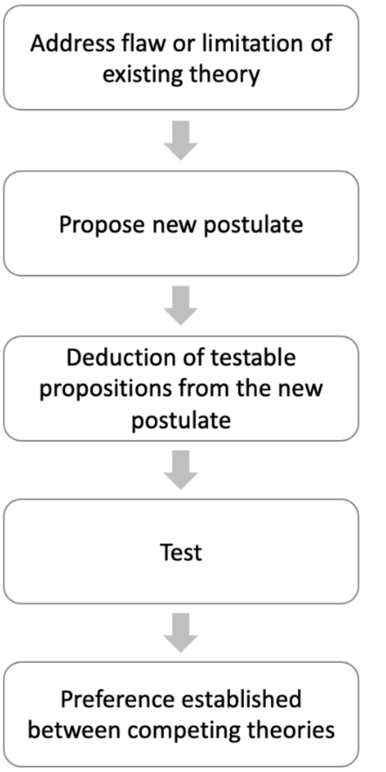


Figure 6 Falsificationist research process (Magee, 1973)

Table 6 Comparing inductivism and falsificationism from the five perspectives

Features	Inductivism	Falsificationism
Key assumption	Inductivist methodology assumes that the group of true statements yield a general universal statement	Scientific knowledge is provisional; Truth is only ever to be approached and never be claimed absolutely we cannot obtain general theories that are universally true
Purpose of scientific inquiry	Extract generalisations from evidence	attempt to disprove a theory thereby improve scientific theory
Validation	Context free	Context bounded
Key distinction	No general justification	A scientific theory must be able to be tested and conceivably proven false

Contribution to scientific research	Revolutions	Science vs. Non-science Constant improvement

As an IS reflects the needs of participants for business operations, in addition to technical considerations, the interactions between participants and information systems account for a significant part of IS research. For all IS studies, there is a key assumption that an IS should be effectively viewed as a carrier medium for social interaction, not just a model that reflects reality (Stamper, 1991, Goguen and Burstall, 1992, Ngwenyama and Lee, 1997, Ulrich, 2001, Liu and Li, 2015). Hence, the use of sign, language and their representations are critical to a successful information system.

3.3 Sign, Language, Representation and Research Paradigm

Signs have been widely applied to every facet of human society, enabling communications and information exchange. The use of signs marks the birth of human civilisation and is also the primary carrier for the continuation of human civilisation (White, 1940). A sign is the smallest unit of meaning; anything can be a sign as long as it denotes, signifies and represents something (Liu and Li, 2015). A sign can be simple or complex, persistent or transient, and static or dynamic. There are three forms of signs: symbolic, iconic and indexical (Peirce, 1931). Words are one of the typical symbolic signs, and language is a typical signifying system deploying signs. Words, phrases, and characters are used to convey meanings and emotions and produce actions in symbolic sign systems. Language can be observed to change across space and social groups and also varies across time. Therefore language is a socially constituted system of signs varying over time (Holdcroft, 1991).

In order to understand how language is employed in information systems, broadly accepted as ‘technical systems with social implications’ (Lyytinen, 1985), semiotics divides the use of signs in IS into three divisions: syntactic, semantics and pragmatics (Liu and Li, 2015). ***Syntactics*** concerns the structure of signs and the rules to compose complex signs. From the use of language perspective, syntactics can be understood as grammar. In IS, syntactics or grammar is the data format in terms of type and length. ***Semantics*** decodes meanings from data through mapping signs onto objects they denoted. When interpretants have reached a shared consensus, the meanings of signs can be perceived. Otherwise, semantic ambiguity is generated. ***Pragmatics*** concerns the behaviour of the agents after receiving signs, which is underpinned by the shared common knowledge and assumptions. From the three levels of using signs as the carrier for information exchange, we can tell that more ambiguity develops as the level rises. The informative nature of signs decides the complexity of communications between agents, based on the sign production, reception, and circulation in all forms.

Language is a sign system that needs to be built on a shared consensus, and shared contexts and assumptions help decrease ambiguity in communication. The idea that language can reshape perception and thought can be traced back to the 1930s, the hypothesis of the "Language Relativity" (Leavitt, 2010). The hypothesis holds that language is not only the basic expression of thought, but also has a profound impact on it, and even determines the formation of our worldview. Empirical evidence (Ahearn, 2021) shows a positive relationship between language usage and people's thoughts and decisions. The categories of describing things or reality in a language system are unique, and different from those in other language systems. Hence, a language represents a nation, reflects the thinking, beliefs and attitudes of speakers of the languages (Berger, 1967, Halliday, 1973), and can be treated as social semiotics (Halliday, 2014). In social life, language is used in different ways for different purposes. As language keeps changing across social groups, space, and time, ambiguity is pervasive at syntactic, semantic, and pragmatic levels. Piantadosi *et al.* (2012) conclude this is due to the informative context and the re-use of words. Aiming for effective communication, Bassham (2011) proposes the elements of language expression shown in Table 7.

Table 7 Elements of language expression to reduce ambiguity, adapted from (Bassham, 2011)

Elements of language expression	Explanation
Clarity	Avoid faulty assumptions Reduce uncertainty (Chartier, 1981)
Precision	Use exactly the correct words to convey the intended thoughts Depends on the native ability of a certain language in terms of the number of different words and grammatical structures that are available to use to convey intentions (Mason, 2019)

Elements of language expression	Explanation
Accuracy	Related to the ability to use language in terms of choosing words, phrases, and sentences that convey ideas, which is impacted by the cultural lexicon (Mason, 2019)
Relevance	Expression fits audience and matches purpose The words, phrases are correct or suitable for a particular purpose The sentences are connected with what is happening or being discussed
Consistency	Logical consistency: saying or believing are consistent in language expression Practical consistency: words and deeds are consistent with each other (Bassham, 2011)
Logical correctness	Same conclusion can be deduced by audiences from the premises and assumptions which the representor provided (Bassham, 2011)
Completeness	All facts and all relevant key elements have been included to prevent the need for further communication, amending and elaborating
Fairness	Free of distorting biases and preconceptions (Bassham, 2011)

Bassham (2011) discusses the standards of effective expression to reduce ambiguity in the use of language. However, ambiguity still exists because of the different cognitive backgrounds of the recipient (Chandler, 1994). Therefore, effective information exchange is a dual form of communication that is decided by the informative nature of signs.

The semiotic principles (particularly Peircean semiotics), as discussed above, lend it naturally to the philosophical and methodological selection in this research. This selection can be characterised by subjectivity in ontology, abductive reasoning in the research process, and falsificationism in research validation. As can be seen in the research design, these characteristics will influence the whole process of this PhD research.

3.4 Research Design

Semiotics in this research is adopted as a scientific approach to deal with semantic interoperability. More importantly, it is also being adopted as a philosophical paradigm (Liu and Li, 2015, Houser, 2020) in research design.

This research will address the semantic ambiguity in healthcare information systems, so understanding the signs used in the domain of healthcare is the primary task. Since there is no strict rule for using signs that the healthcare industry has agreed upon, the interpretivism paradigm is adopted to direct this research to understand the existing sign systems in the healthcare field and the problems researchers address.

The sign system mentioned in this research refers to the communication protocols which aim to provide information exchange standards for the entire healthcare domain. However, there is no perfect communication protocol in practice, and improvement room for each communication protocol always exists because of the informative nature of signs. Therefore, the falsificationist research process (Magee, 1973) is deemed more appropriate for addressing problems in order to get closer to the perfect communication protocol at the philosophical level. This research is devoted to improving the existing protocols to make them closer to perfect ones. The following sections will detail the method of identifying the research target and determining the research questions and process.

3.4.1 Understanding the Use of Signs from the Perspective of Interpretivism

The research questions posited in this thesis require the researcher to 1) understand the use of signs in health information systems; 2) identify the reason for semantic ambiguity from the perspective of the use of signs; 3) apply the research process to develop information architecture; 4) justify the information architecture by ambiguity reduction. This research has indicated that problems in existing communication protocols caused by the subjective understanding of the use of signs in the development of information systems are critical for this research to identify the limitation of existing theory/protocols. Then a solution or new protocol is expected to be proposed underpinned by a comprehensive understanding of the use of signs from an individual’s perspective.

This research adopts the interpretivism paradigm to study the use of signs in the healthcare ecosystem. As interpretivism assumes reality is socially constructed by individuals, the use of signs reflects the individual’s understanding of the context of healthcare activities. The aim of this research is to decrease semantic ambiguity in healthcare information systems that use FHIR. The FHIR specification is regarded as a consensus agreement in healthcare information systems; the semantic ambiguity is generated from the diverse understandings of users on the sign system of FHIR. According to the theory of semiotics, different users interpret the same sign differently, influenced by their perceptions of context or experiences (Blumer, 1986). As human language is the most well-developed sign system (De Saussure, 2011), such ambiguity is common in specifications expressed in language. Thence, the stance on interpretivism, which recognises the existence of multiple understandings, enables the researcher to understand the causes of semantic ambiguity from different participants' perspectives, which is the foundation for the researcher to propose the solution to improve the use of FHIR in healthcare information systems. Table 8 illustrates the philosophical assumptions of adopting interpretivism as a research paradigm from the ontological, epistemological, and methodological points of view of this research.

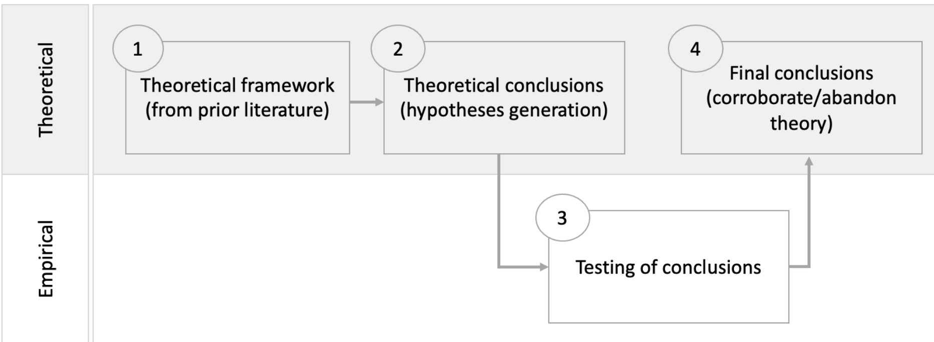
Table 8 Philosophical Assumptions of this Research

Philosophical Assumptions	Descriptions
Ontological	The information used in healthcare ecosystems is socially constructed and highly associated with the actions or the demands of participants/stakeholders. The nature of information in healthcare ecosystems should be studied under the premise of fully considering its context and usage scenarios and be conducted from the perspective of users.
Epistemological	Information is represented by signs in healthcare information systems to be exchanged between dispersed information systems, and between users and systems. Knowledge is gained from the understanding of the use of signs in the context of these interactions. The obtained knowledge is applied to improve semantic interoperability and decrease semantic ambiguity during the healthcare information exchanging process in this study.
Methodological	Qualitative research methods will be dominantly adopted in this research, and the falsification research process will be applied to lead the research conduction and the structure of this thesis. This process will be further discussed in the next section.

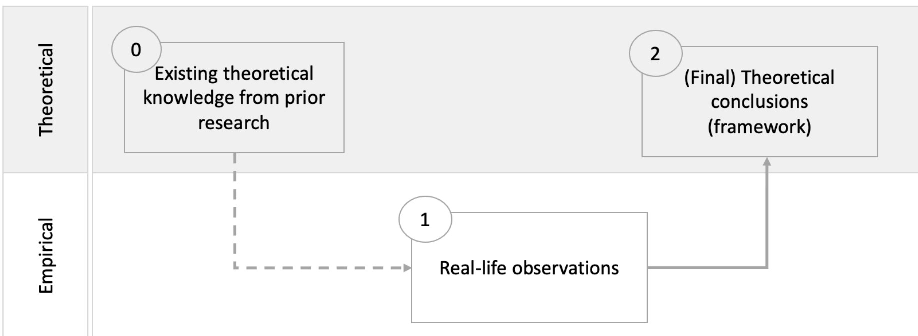
3.4.2 Falsification Research Process

This research aims to increase the semantic interoperability between healthcare information systems that employ FHIR as a communication protocol by decreasing semantic ambiguity. Regarding the research objectives, the falsification research process is adopted as a cornerstone of research methodology in this study to guide the research design. Before discussing the details of the falsification research process, the author would like to discuss the relationships between the falsification research process and other traditional research processes.

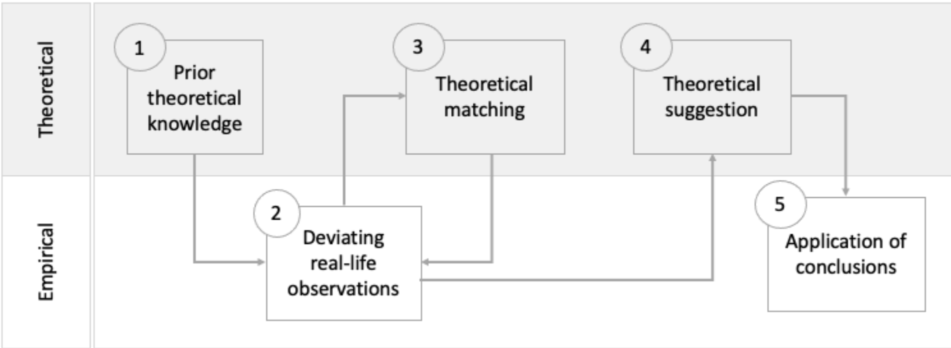
Inductive and deductive reasoning are two general approaches researchers adopt in scientific research. Inductive reasoning seeks to establish generalisations based on the observations of specific instances; deductive reasoning seeks to examine generalisations with particular cases. Most often, interpretivism follows an inductive process and acquires knowledge from the qualitative investigation. Some scholars also discuss the deductive reasoning in qualitative research to test a theory by the "pattern matching" approach (Hyde, 2000). In addition, some recent research shows that interpretivism is intertwined with abduction (Muggleton *et al.*, 2014, Petty *et al.*, 2012, Lukka and Modell, 2010). Abductive reasoning is adopted in the scientific inquiry (Peirce, 1931) persistently seek evidence that confirms or refutes the existing theoretical propositions. Figure 7 shows the deductive, inductive, and abductive research processes.



(a) Deductive research process



(b) Inductive research process



(c) Abductive research process

Figure 7 Deductive, inductive and abductive research processes (Kovács and Spens, 2005)

Deductive research 1) derives theoretical conclusions from prior literature and generates hypotheses and propositions; 2) tests the hypotheses and propositions by empirical research; 3) corroborates or abandons the theoretical conclusions. The logical order of deductive research is from theory to practice to results (Danermark *et al.*, 2019). Inductive research follows the opposite logical order: starting from the observations and then generalising theoretical frameworks, which follows the pattern of practice-result-theory (Danermark *et al.*, 2019). In contrast, abductive research is substantially different from deductive and inductive research processes. Abductive reasoning has five steps: 1) derives conclusions from prior theoretical knowledge; 2) observe the instances in real life; 3) match the observation facts with existing theory; 4) suggests new theory when not exactly match; 5) apply the new hypotheses or propositions into empirical settings. Abductive reasoning emphasizes the “theory matching” (Dubois and Gadde, 2002), which is a process to search for a suitable theory to an empirical observation (Kovács and Spens, 2005). An abductive approach is concerned with the particularities of a specific instance that deviate from the general structure of such situations (Danermark *et al.*, 2019). Most researchers employ abduction to examine the applicability of the generalised theory with situational environmental factors, especially in social science research. The "theory matching" process helps researchers distinguish which aspects of the application scenarios can be generalised when applying these theories and which aspects are only relevant to the specific situation itself (Liu *et al.*, 2014, Tan *et al.*, 2018). The purpose of abductive research is to propose a new matching framework or extend the prior theory to a new phenomenon (Andreewsky and Bourcier, 2000). Compared with inductive reasoning and deductive research approaches, the abductive research process is more appropriate for information architecture study because developing a software system architecture is not only about theories but is more critical to understanding the needs of stakeholders and the business. To a certain extent, the development of the system architecture can be understood as the matching process of the existing technical architecture and the needs of social activities. Knowledge is acquired from extending the application boundaries of existing technical architecture.

As information systems are context-bounded, there is always room for improvement and optimisation. Falsificationism provides a method to constantly improve the existing technical architecture by challenging flaws or limitations of existing solutions. In general, the falsification research process can be regarded as subsequent actions of mismatching between current theories and observations. A new postulate theory aiming to reconcile the deviation will be proposed and tested to ensure that the new postulate theory expands the explanation of mismatching beyond what was explained by the original theory. In short, the essence of the falsification research process is to describe how to improve existing theories through scientific practice and bring better explanations of the observed world. Figure 8 transforms falsification research processes proposed by Magee (1973) into a swim-lane diagram in the theoretical and practical fields.

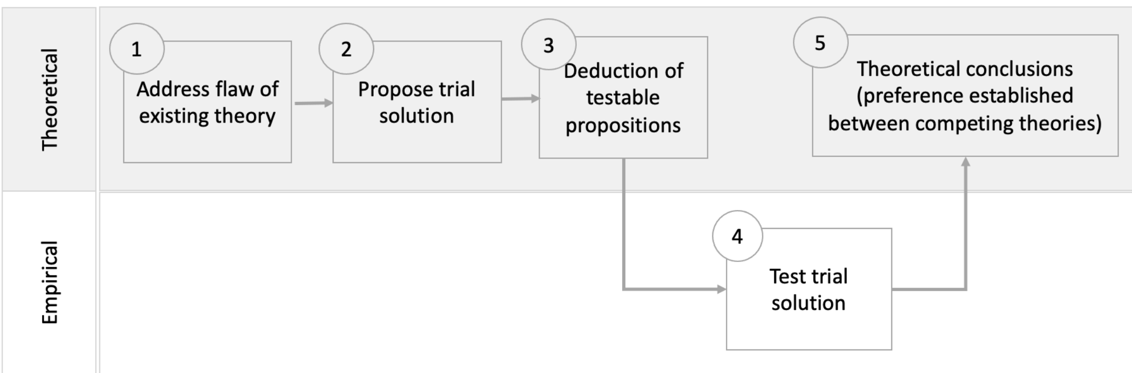


Figure 8 Falsification research processes (Magee, 1973)

Table 9 compares abduction and falsification from five perspectives. The similarity of the two research processes is significant. In most of research practices, abduction is used as a basic form of logical reasoning (Minnameier, 2010), and falsification is usually used to validate a theory. The method of attempting to disprove a theory, thereby improving scientific theory, is the essence of falsificationism. As the falsification research method is gradually accepted and adopted by the research in the IS field (Lee, 2004, Lee and Hubona, 2009, Salovaara and Merikivi, 2015), IS can be gradually improved through continually examining the boundary conditions of existing software architecture and extending the boundary by developing new software architecture.

Table 9 Abduction and falsification

Features	Abduction	Falsification	Sources
Logic	Propose hypotheses that explain more practical observations	Falsify a theory or propose more generalised hypotheses that explain more practical observations	(Lopez, 2013, Salovaara and Merikivi, 2015, Magee, 1973)
Data analysis	Start from theory to data to propose new hypotheses	Start from theory to data to abandon existing theory or propose new hypotheses	(Yu, 1994, Salovaara and Merikivi, 2015)
Mode of discovery	Theory-informed	Theory-informed	(Kovács and Spens, 2005, Magee, 1973, Gregor *et al.*, 2013)
Knowledge generation	Yes	Yes	(Kovács and Spens, 2005, Magee, 1973, Popper, 2013)
Possibility of human creativity	High	High	(Kovács and Spens, 2005, Popper, 2013)

In this research, the aim is to improve the semantic interoperability between healthcare information systems by solving the limitation of FHIR in implementation. To adapt the falsification research process proposed by Magee (1973) and combine with the abductive reasoning, this research will be formulated into the process as shown in Figure 9.

In Figure 9, the vertical flow chart on the left side of the figure is falsification research process which combines the proposing test propositions and testing of Figure 8 into one step. Consequently, this research is divided into four phases, with several sub-steps in each. This thesis's research steps are depicted in the swim lane diagram on the right side of the figure.

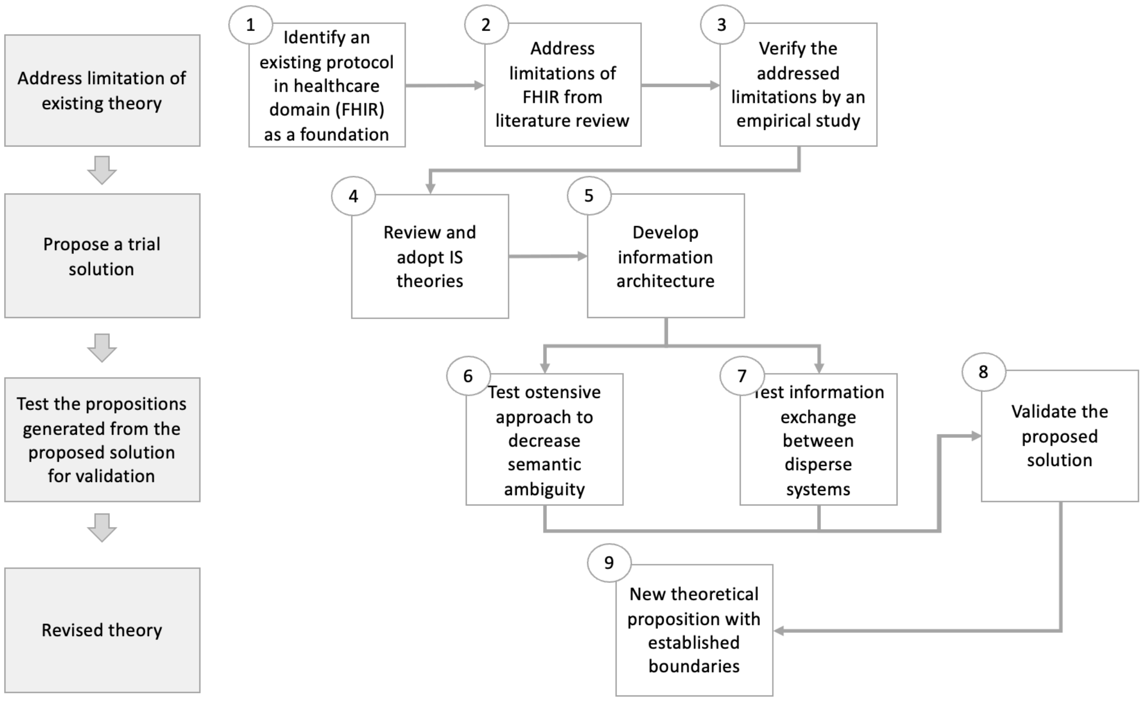


Figure 9 Research design based on abductive reasoning and falsificationism

Address limitation of existing theory

The first stage of this research is to identify the limitation of the existing widely adopted communication protocol in healthcare information systems. This stage consists of three sub-steps. The first sub-step is to identify one communication protocol that stresses the sematic interoperability as the research object of this study (Section 2.4 and 2.5). An extensive literature review is conducted because there are many protocols in the domain of healthcare information systems that claims to support semantic interoperability. The second sub-step is to identify limitations of selected communication protocol of semantic interoperability from the literature review (Section 4.3). And the last sub-step of this phase is to verify the limitations that are addressed in literature through applying selected communication protocol to imperial data (Section 4.3).

Propose a trial solution

After identifying the limitations of the existing solution in the first stage, an extensive literature review of the related work helps to develop the proposed solution for the specified constraints. The first step is to explore the reason for the restrictions (Section

Chapter 5). Based on the identified root causes for the limitations, the second step is to conduct literature review of IS theories (Section 5.3 and 5.2) aiming to solve the problems. The third step is to develop a new information architecture to solve the problems in the existing solution (Section 5.4 and 5.5)

Test the proposed information architecture and validation

This stage is to design test cases for the proposed architecture against the limitations of the existing solution. In this research, there are test cases are conducted. The first one is to verify the capacity of the proposed architecture in terms of decreasing semantic ambiguity through show the examples (Section 6.1). The second case is used to demonstrate the capacity of the proposed architecture to handle the information exchange between disperse systems (Section 6.2). The last step in this stage is to discuss the proposed architecture and compare it with the existing solutions (Section 6.3 and 6.4).

New theoretical proposition with established boundaries

When the proposed solution has been tested and validated, the theoretical proposition for this new solution can be addressed (Chapter 7). Furthermore, the limitations and implications for future research are discussed at the end to present a complete view of this research outcome.

3.4.3 Summary

This chapter has discussed the critical aspects of the research framework that provides the research paradigm, approaches, and methods for this research. In addition to conventional research paradigms, this research particularly discussed falsificationism and its applicability of this research. Regarding the research topic, this chapter explored the capacity of a sign system in information system design. Through the sign system, users interact with information systems, and it is precisely in the middle of this interaction that semantic ambiguity is generated, which affects semantic interoperability. The context-dependent property of a sign system makes its applicability have clear boundaries. This research aims to extend the boundaries of an international specification of semantic interoperability by addressing its limitation, which presents the validity of falsificationism in this research. The research activities are also structured with the guide of the falsification research process.

Chapter 4 Investigation of FHIR and its limitations

This chapter investigates the FHIR specification from the following aspects: the evolution of HL7 series specifications, the design ideas of FHIR, the interoperability paradigms, the FHIR architecture patterns and its limitations. This chapter provides a comprehensive understanding of FHIR in order to take full advantage of FHIR's design benefits and propose improvements to its limitations.

4.1 FHIR chosen as the research platform

On the basis of the comparative analysis of OpenEHR and FHIR in Section 2.5, the OpenEHR design paradigm (openEHR, 2002) can be summarised by three "separations": ontological separation, separation of responsibilities and separation of viewpoints (The OpenEHR foundation, 2003), which have been discussed from the five aspects. However, it is challenging to deploy OpenEHR on outdated systems because the entire system must be changed. OpenEHR's industry appeal is impeded by its expensive implementation, particularly for large projects. FHIR, on the other hand, is rapidly adopted in the health care industry due to its easy-to-start features and usage of Internet interface protocols as its external interfaces. Although FHIR's architecture is still needs to be improved, its easy-to-operate allows it to be rapidly adopted by industry applications and swiftly occupy most of the market share. Consequently, the benefits of FHIR in industrial applications is substantial. The Google trend depicted in Figure 10 and 11, as well as the industry discussions from ALCIDION (2022), echo the results of comparative analysis in Section 2.5. In the consideration of industry acceptance, this study employs FHIR as its research foundation.

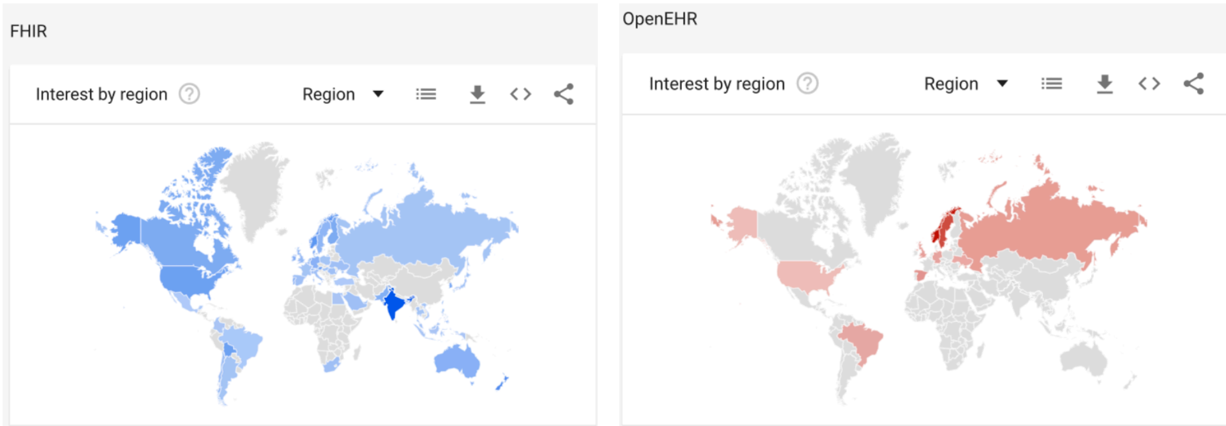


Figure 10 Interest by region-FHIR & OpenEHR

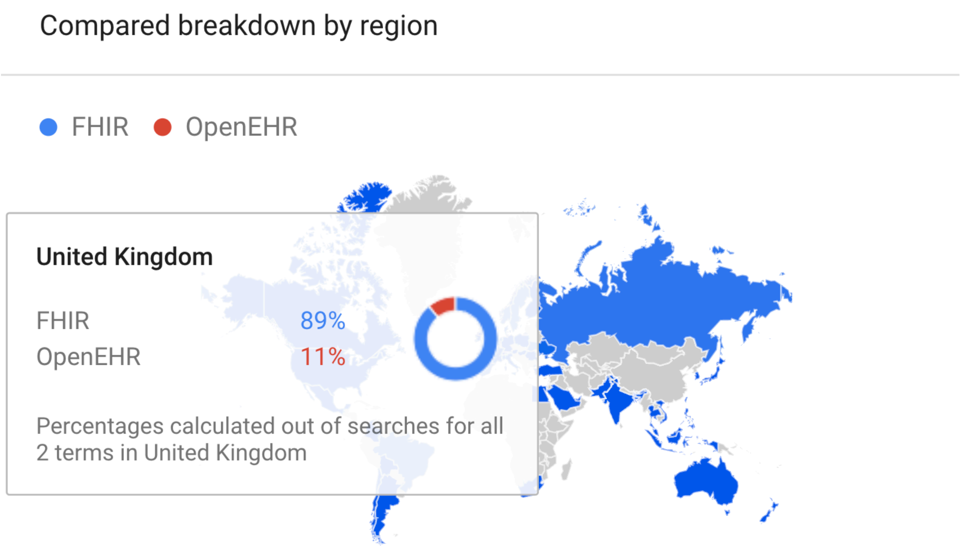


Figure 11 Interest in UK- FHIR & OpenEHR

4.2 Fast Healthcare Interoperability Resources (FHIR)

Fast Healthcare Interoperability Resources (FHIR) is proposed by HL7 International to improve the interoperability of systems in the healthcare domain and to facilitate information exchange between the stakeholders of healthcare ecosystems. FHIR is an open suite of software specification and implementation, comprising two elements: information models entitled '*resources*', and a specification for the exchange of these resources. The goal of FHIR is to render all health data accessible to large-scale analytics in order to improve the quality of healthcare services. FHIR distinguishes itself from the previous versions of HL7 standards (HL7 v2, HL7 v3 and CDA) because of the better leveraging of modern web technologies. In this chapter, HL7 series standards are reviewed in the first subsection in order to explore the evolution of interoperation standards and the driven factors. Then the FHIR standard is discussed in detail in the following subsection because it is expected to overcome the limitations of previous versions. As the basis of this research, FHIR architectural patterns and the limitations of FHIR are discussed in the last two sections.

4.2.1 Health Level 7 (HL7) and its Evolution

Health Level Seven International is a non-profit ANSI-accredited (American National Standards Institute) standards development organisation, which was founded in 1987 and has published a set of international standards for the transfer of clinical and administrative data between healthcare organisations. Health level seven international is abbreviated as HL7. This name is derived from the 7th level of the OSI (Open System Interconnection) model, shown in Figure 12. HL7 produces the international standards for healthcare interoperability widely adopted globally. The following will introduce HL7 Version 2, HL7 Version 3, HL7 Clinical Document Architecture (CDA) and FHIR, which are the four primary standards developed by HL7. The series of the four standards reflect the evolving history of HL7 protocols.

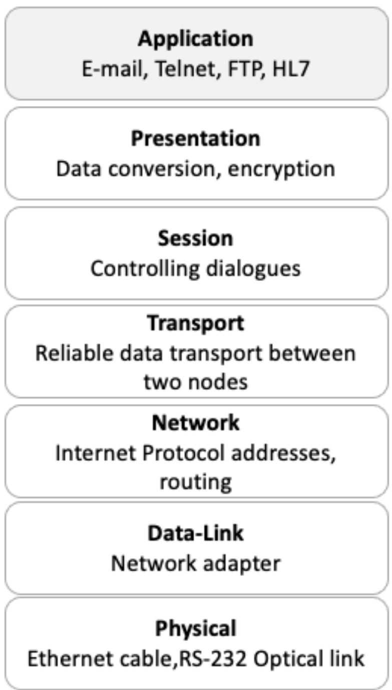


Figure 12 The interoperability levels of clinical standards based on OSI model (International Organization for Standardization, 1994)

- HL7 v2

The initial goal of HL7 is to develop grammars for messaging and standardise vocabulary. So, Health Level version 2 (HL7 v2) was developed in 1989, which adopts a pragmatic approach to define a set of specifications for clinical messages, mainly covering administrative and clinical systems. HL7 v2 is a pure message communication protocol without any underlying model. It defines a standard data structure to describe clinical events by pre-defined segments, fields, and data types, which consists of a message header and the content of clinical information partitioned by delimiter symbols like | or ^. Figure 13 shows an example of laboratory results.

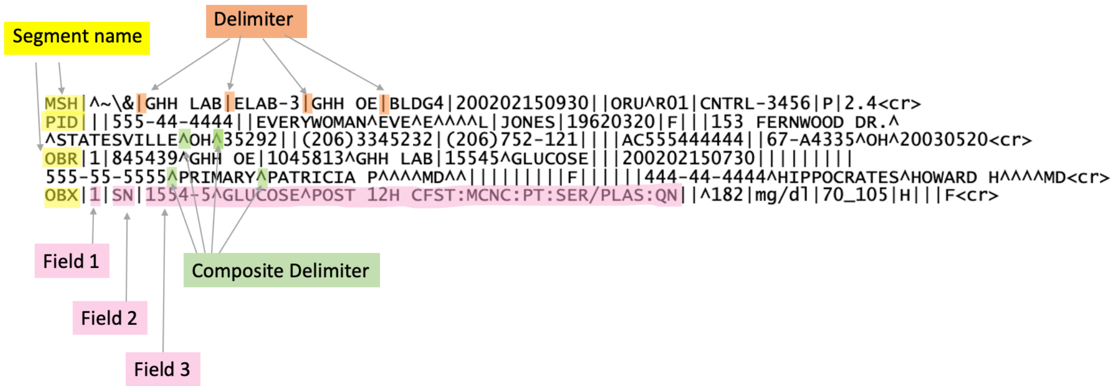


Figure 13 HL7 v2 example

In Figure 13, MSH (Message Header segment) is the message header; PID (Patient ID segment)identifies the patient, including the information of name and date of birth; OBR (Observation Request segment) is used to transmit information specific to an order for a diagnostic study or assessment; OBX (Observation/Result segment) carries information about observations in report messages. The standardised message can be easily identified by the information system through the pre-defined codes. As no other standards were available at the time, HL7 v2 grew rapidly when healthcare organisations and industry vendors realised the benefits it offered. By now, over 90% of American hospitals and many other countries have adopted HL7 v2 in message exchange (HL7, 2022c).

- HL7 v3

Based on the interoperability of message exchange provided by HL7 v2, HL7 v3 aims at a firmer and context-rich standard, which is expected to support healthcare workflows through data exchange. Therefore, HL7 v3 produces shared reference information models (RIMs) to enable more reuse, standardisation and format consistency. Unlike the HL7 v2 message, the HL7 v3 adopted XML to present clinical information with a semi-structured format, shown in Figure 14, which is a human-readable document.

```
<observationEvent classCode="OBS" moodCode="EVN" />
  <code code="POT" codeSystem="2.16.840.1.566.2.166" displayName="Potassium" />
  <statusCode code="completed" />
  <effectiveTime value="20100721082300" />
  <value xsi:type="PQ" value="120" unit="mg/dL" />
- <recordTarget typeCode="RCT">
  - <patient classCode="PAT">
      <id extension="2828811" root="2.16.840.1.566.2.701.67" />
      <statusCode code="active" />
    - <patientPerson classCode="PSN" determinerCode="INSTANCE">
      - <name>
          <given>John</given>
          <family>Smith</family>
        </name>
      </patientPerson>
    </patient>
  </recordTarget>
- <author typeCode="AUT">
    <time value="20100721115207" />
  - <assignedEntity classCode="ASSIGNED">
      <id root="2.16.840.1.566.2.701.2" extension="HEN03" />
    - <assignedPerson classCode="PSN" determinerCode="INSTANCE">
      - <name>
          <given>Peter</given>
          <family>Hendriksen</family>
        </name>
      </assignedPerson>
    </assignedEntity>
  </author>
</observationEvent>
```

Figure 14 HL7 v3 example

HLR v3 is not an incremental version of HLR v2 but a decided departure. HL7 Development Framework (HDF) is introduced by HL7 v3 as a development process to employ RIMs to develop clinical information models for various scenarios. The Clinical Document Architecture (CDA) is one of the proposed RIM standards, a widely adopted application of HL7 v3 for exchanging electronic documents that define the format and the shared meanings of clinical documents (Dolin *et al.*, 2001). This top-down Model Driven Architecture is expected to generate the messaging artefacts automatically;

however, the implementation of HL7 v3 is complex, and a customer compiler is usually needed to match RIMs with local specific models in most scenarios (Bender and Sartipi, 2013).

In short, HL7 v3 provides interoperability through both message and document. Messages are generally used to support an ongoing process in real-time, while documents are used to process ‘static’ content. Comparing the broad adoption of HL7 v2, the adoption of HL7 v3 is limited. The first reason is that the adoption cost is high with a long adoption cycle as RIM model is over complex to implement (Worden and Scott, 2011, Benson and Grieve, 2016). The second reason is that HL7 v3 is not compatible with v2, which leads to two versions coexisting in the same system and also indirectly causes the problem of high maintenance costs (Mead, 2006).

- FHIR

FHIR was introduced in 2014 and is compatible with all previous versions, including HL7 v2, v3 and CDA. FHIR achieves interoperability through four levels: API (RESTful interface), message (similar to HL7 v2), document (similar to CDA) and Service. In the following subsections, FHIR will be introduced in detail. In summary, HL7 v2, v3 and FHIR are in an iterative relationship (Figure 15). The original intention of HL7 is to replace the old standard with a new one. However, due to practical issues such as implementation costs, multiple standards co-exist in practices.

From Figure 15, FHIR can be regarded as the latest standard developed by HL7, inheriting previous versions’ merits. The following subsection will introduce FHIR in detail.

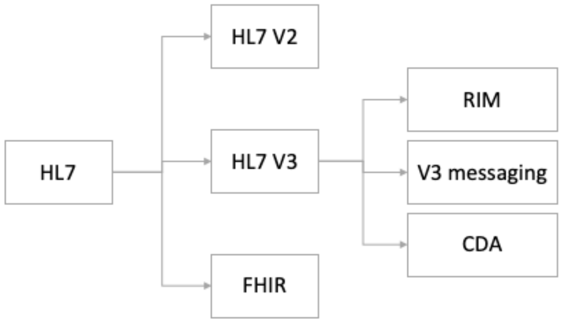


Figure 15 HL7 standards

4.2.2 HL7 FHIR Standard

FHIR is the next-generation interoperability standard of HL7, aiming for seamless and on-demand information exchange between heterogeneous healthcare systems. In order to improve the adoption, facilitate the mobility of healthcare data, and promote application innovation based on healthcare data, FHIR simplifies implementation complexity without sacrificing information integrity and thereby reduces integration costs. The birth of FHIR mainly stems from the lack of implementation of the HL7 v3 benefiting from the project named “fresh look task force”. This project is to examine how to improve HL7 messaging standards. Therefore, FHIR adopts a new approach to healthcare information exchange and then draws on the mature architecture of internet technology widely adopted in the industry. This approach is based on RESTful (Fielding, 2000), an open Internet standard widely used to engage applications running on mobile devices or web browsers. HL7 claims that HL7 v3 has not been abandoned but that FHIR was born because of learning HL7 v3.

The philosophy behind FHIR is to focus on implementers, leverage cross-industry technologies and make the healthcare information readable and freely available. Therefore, FHIR adopts a modern, internet-based approach to connect the discrete healthcare elements represented by a set of predefined ontological models. The following subsections will introduce the characteristics of FHIR in detail, starting with the introduction of these ontological models.

4.2.2.1 FHIR Resources

The fundamental units of FHIR representing clinical information are *resources*, which are information models featuring a set of predefined properties for a specific aspect of the domain. For example, the *resource* representing an individual patient has attributes including name, gender, address, and date of birth. Effectively, a *resource* can be identified as a schema which describes all the relevant attributes of a conceptual entity. Then HTTP-based RESTful API is developed to access these *resources*.

FHIR divides various resources into the following five categories (Figure 16), mapping to different layers of applications. With the FHIR release 4, there are 146 resources defined, a number that grows with every release. FHIR resources include *Foundation (30), Base (26), Clinical (39), Financial (16), and Specialized (35),* consolidating all categories of data with these predefined *resources,* which have already been in use or will be used in the healthcare information systems. FHIR, as a sign system, offers a defined lexical space in which clinical concepts, healthcare services, and FHIR resources are utilised.

These predefined *resources* currently cover 80 per cent of data elements used in existing healthcare systems, which are the common requirements of data usage scenarios. The remaining 20 per cent less used data elements are left to developers to define, known as *extensions*. This 80/20 rule avoids the proliferation of numerous, overlapping, and redundant resources.

A *resource* is the basic unit of interoperability, which can be likened to an alphabet as *resources* can be combined into a *profile*, just like alphabets forming words. When a single *resource* is not enough to represent the data needed for a certain application scenario, FHIR allows multiple *resources* bundled together to describe a complex concept. The profiling process refers to the combination of *resources* or *resources* and *extensions*. Thus, the profiling process involves subjective understanding, especially when resources describe clinical behaviours, such as workflows, updates, and services.

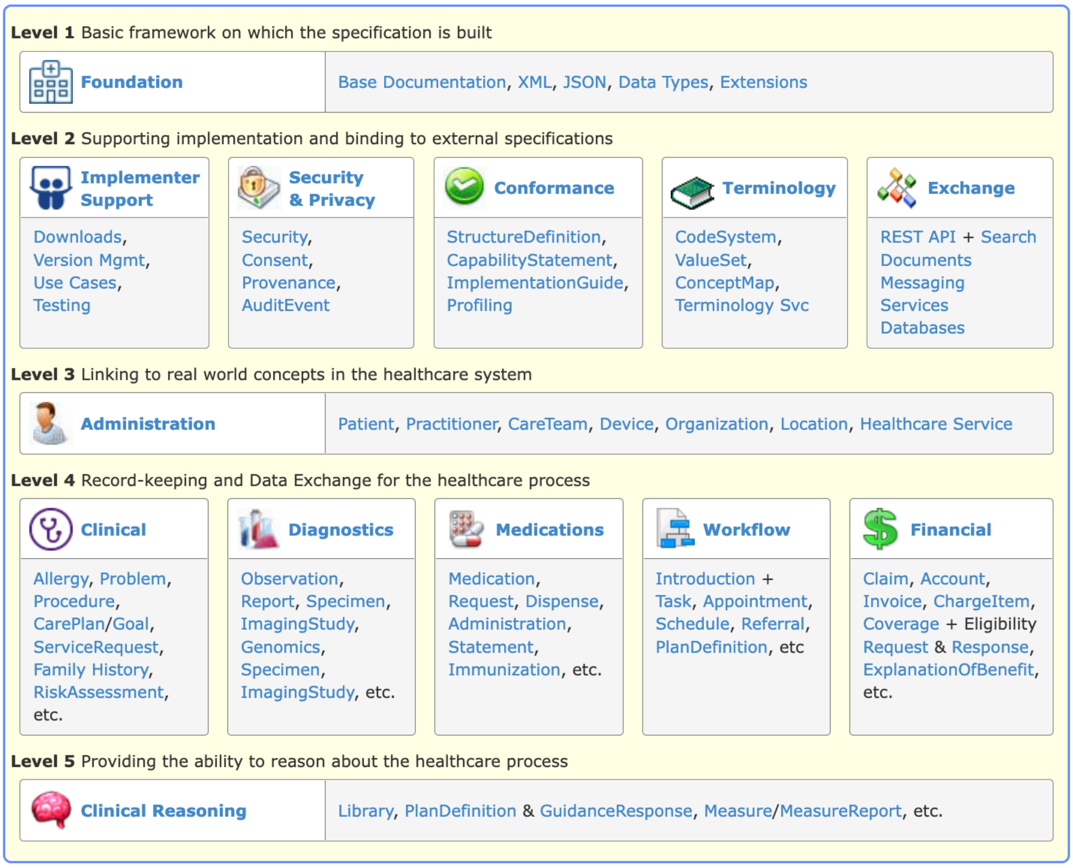


Figure 16 FHIR Resources

FHIR resources can be flexibly connected based on the structure of medical services. For example, the conceptual map of medication prescription services can be illustrated as Figure 17. The FHIR protocol is compatible with numerous medical terminology codes, including SNOMED CT, LOINC, and UCUM, among others (HL7, 2022a). In the example given below, the pharmaceutical code often uses SNOMED to represent the substance's name. As shown in Figure 15, the graph in which concepts are connected through different relationships exemplifies the basic connection of Knowledge Graph, making it feasible to convert the FHIR definition into a Knowledge Graph format.

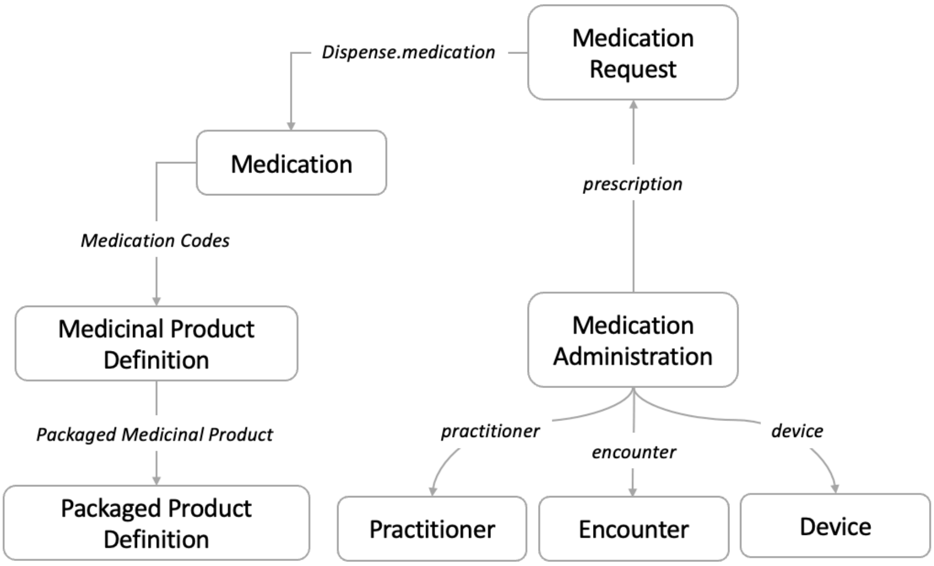


Figure 17 Health services related to medication

An example of the FHIR resource below (Figure 18) is about a patient. The key parts of this *resource* include four parts: Metadata, narrative, extension and elements.

1. Metadata

Metadata is an optional part of a resource, a set of information describing the technical and workflow context of the resource. In FHIR, Metadata contains the literal identity in the form of a URL and the date where the resource was last updated.

2. Narrative

The narrative provides the context and the content of essential clinical or business information for the resource in a human-readable expression. For instance, the narrative for a Medication Request (a *resource* of FHIR) cloud includes a summary regarding the referenced patient, prescriber and medication.

3. Elements and Extension

Elements and Extensions are the two types of data modules in FHIR, representing all necessary information regarding each resource. The Elements are a normative part of in FHIR standard, while the Extensions can be built individually by developers, vendors or interested parties developing their dedicated domain information and applications.

As shown in Figure 18, a *resource* contains a set of elements defined in a strict hierarchy. Elements either have child elements or a primitive value. Every element can have extensions containing either a value or other extensions. Figure 18 demonstrates the *resource* represented in XML format. Other formats, such as UML, JSON and Turtle, are naturally supported by FHIR.

As shown in Figure 19, the UML format of resource (patient) clearly shows the graphical properties of the healthcare data model expressed by nodes and relationships. It is for this reason that this study chose to use a graph database to describe FHIR resources and their relationships.

Besides the predefined content models, FHIR defines exchange specifications to facilitate information exchange, which herein are named interoperability paradigms.

Figure 18 Example FHIR Resources – Patient in XML format (HL7, 2022b)

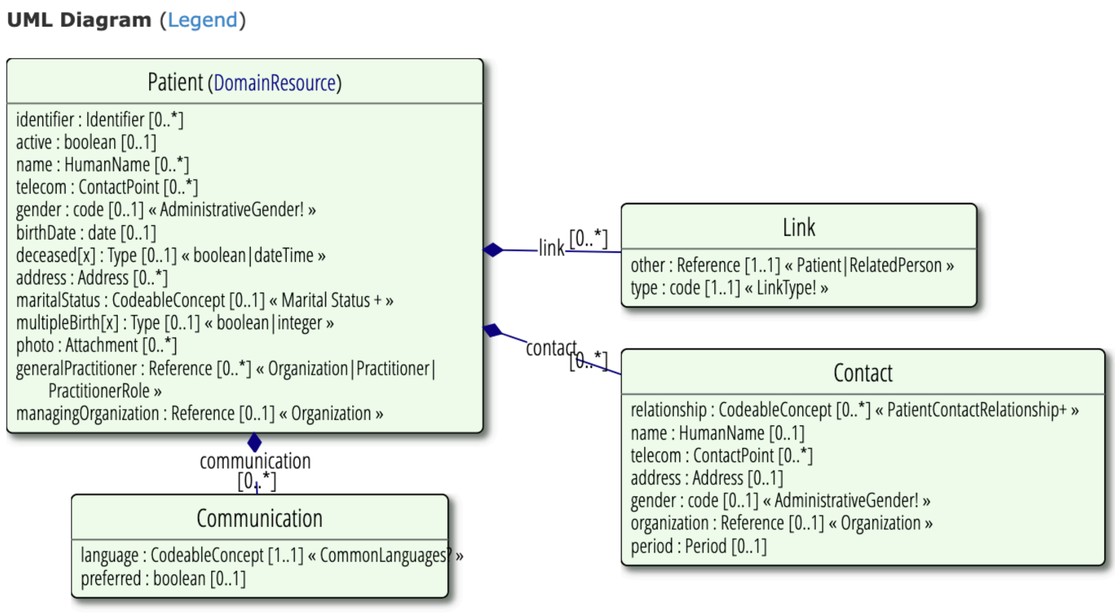


Figure 19 Resource-Patient in UML format

4.2.2.2 Interoperability Paradigms

As mentioned in the earlier subsection, FHIR establishes interoperability with four paradigms, including API, messaging, documents, and service. Strictly speaking, these four paradigms belong to two categories.

- API paradigm

To be precise, the API paradigm in FHIR refers to the RESTful interface, which conforms to the constraints of REST architectural style and enables interactions for web services. As a RESTful specification, FHIR is organised around the concept of a repository, which is a list of resources of a particular kind. For example:

http://myfhirserver.com/patient/122735

This URL consists of three parts structurally: [server-address]/[type]/[id], which indicates the record of *patient 122735* is stored in the server of *myfhirserver*.

In contrast to the RESTful interfaces used in other typical scenarios in the IT industry, the RESTful in FHIR is a general specification and is used to carry multiple types of information exchange. In general, there are three types of FHIR services underpinned by the RESTful interface: 1) instance service: allows clients to retrieve, update, delete and see the modification history of a resource; 2) type service: allows clients to create instances of a resource and search all existing resources, 3) system service: allows clients to determine which functions are available regarding specific use cases (Benson and Grieve, 2016). As described above, RESTful API employs pre-defined operations to deal with the small and light-weight exchanges between low coupling systems, especially suited for Mobile services.

- API enabled paradigm

Underpinned by RESTful API, messages, documents, and services are allowed to be exchanged with RESTful web services.

1) Message exchange

   A message consists of multiple resources in a single exchange and is exchanged between applications when a specific event occurs. The message interoperability paradigm is requested to complete complex operations than CRUD (Create, Read, Update and Delete) on multiple resources. A message is used to handle asynchronous communication scenarios and is employed to request or respond to a workflow. The structure of an FHIR message includes the message header and content. The message header describes the information of the message sender and receiver, and the content is comprised of one or multiple resources. Figure 20 shows a message consisting of a message header and three *resources* of FHIR.

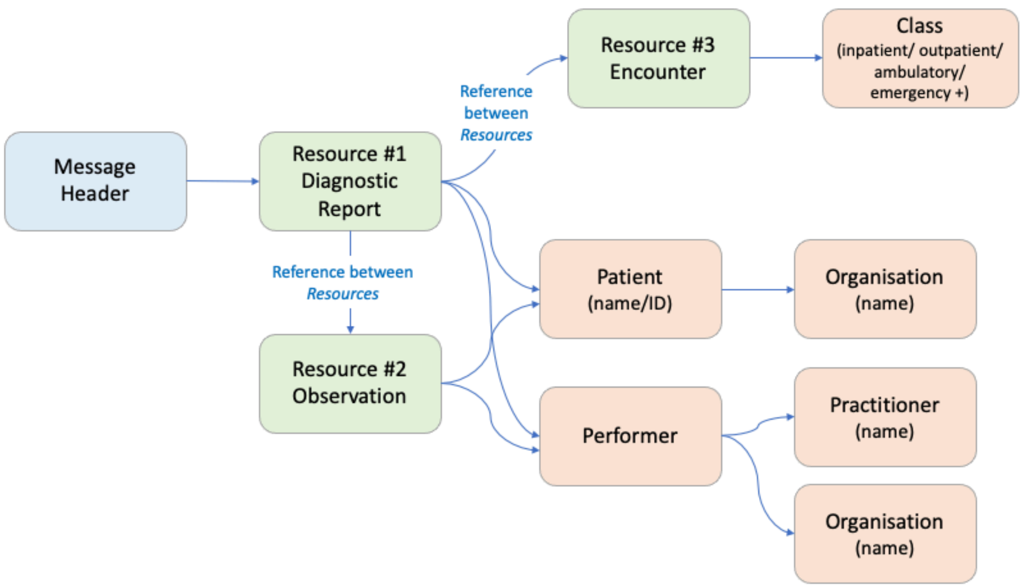


Figure 20 Example of FHIR message consisting of multiple *resources*

2) Documents exchange

The document interoperability paradigm is used to communicate multiple resources from source data whose content is ruled by authentications but without workflow involved. In the context of healthcare information exchange, a document refers to a set of fixed healthcare information packages that can be exchanged for storage for later use (Shown in Figure 21). Document paradigm adopts when a group of information which contains a composition and supporting resources are requested to transport with persistence and testability. However, when the data is dynamically requested by a client or is used to respond to a workflow request, the document paradigm is not fit for such scenarios.

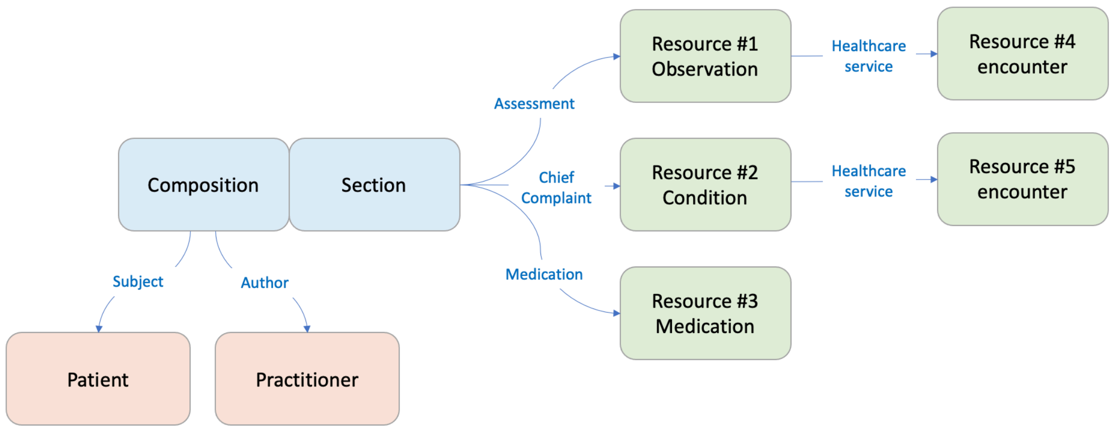


Figure 21 Example of document exchange

3) Service exchange

Service refers to a set of functions that one healthcare information system can provide externally defined based on its data and users' requirements. Especially when all previous paradigms don't fit exchange requirements, the service paradigm is invoked. For example, the workflow is more complex than a simple request/response, which is beyond the scope of the message paradigm can handle.

4.2.2.3 The Roles of FHIR in Enterprise Architectures

Enterprise architecture (EA) is widely adopted for leading enterprise responses to a dynamic business environment by carefully selecting technologies to align with the business context toward desired business vision and outcomes. EA is a discipline that provides a set of methods to define an organisation's structure and interrelationships and their interactions between critical business domains, such as business, technical, physical, or organisational (Dedić, 2021). In the context of business digital transformation, EA is employed as an architectural tool to help enterprises proactively and holistically establish a convergence strategy of business and technology. The TOGAF (The Open Group Architecture Framework) (The Open Group, 2018) and the Zachman Framework (Zachman, 1999) are the two enterprise frameworks that got massive and markable attention from academia and the industry, as shown in Figure 22. This sub-sector discusses the positions of FHIR in the two widely adopted EAs in order to understand the FHIR's facilitating in establishing IT systems to support business operation, particularly in the digital transformation.

In the left half of Figure 22, from the TOGAF perspective, FHIR specifications provide the definitions of *resources* fitted within the information systems architecture domain (circled with a red ring) to address the information models from an architecture viewpoint; the FHIR APIs for data exchange addresses aspects of application architecture. When a party in the healthcare ecosystem, for example, a hospital, has chosen to undertake a large-scale architecture transformation to adopt FHIR in the incumbent healthcare information systems, it is crucial to understand and address data management issues. A comprehensive data management approach is essential to capitalize on its competitive advantages reflected in the effective use of healthcare data. FHIR is the data management approach to define healthcare data and their relevant applications via FHIR resources and FHIR APIs.

In the right half of Figure 22, from the Zachman framework viewpoint, FHIR explains healthcare data models, how they exchange through various interoperability paradigms, and what capabilities they have through API definitions. Therefore, FHIR fits within the dimensions of *What* and *How* within the Zachman framework to address the technical and implemental concerns of architects, engineers and technicians.

The addressed FHIR positions in TOGAF and Zachman framework provide the technicians with the total consideration that hybrids the services and technology requirements when involving the digital transformation process for local healthcare information systems.

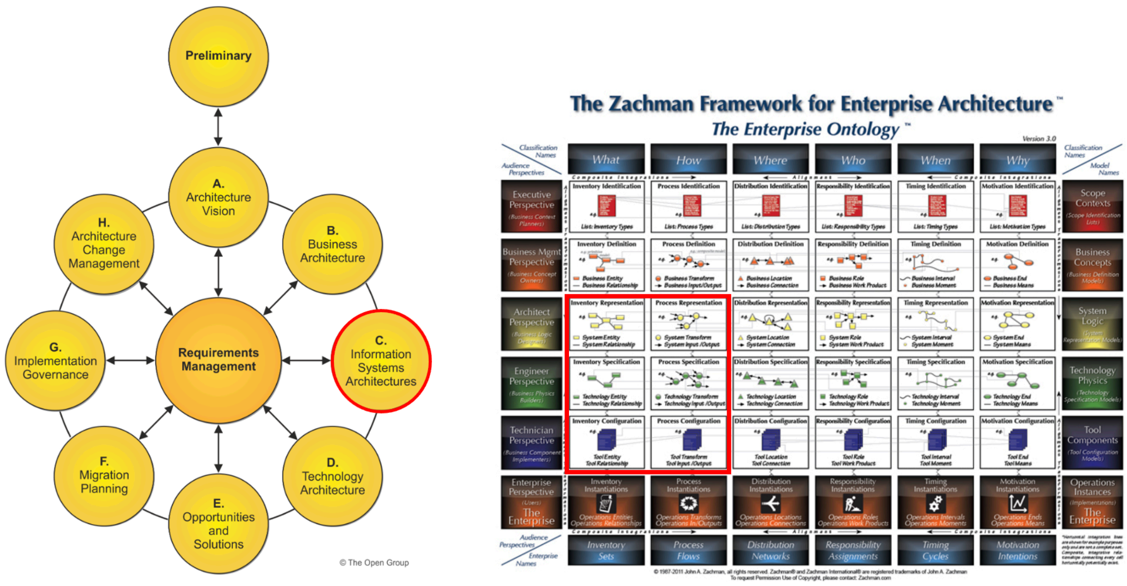


Figure 22 The positions of FHIR specification in information architectures

4.2.3 FHIR Architectural Patterns

FHIR is not only a set of healthcare data models and web-standard-based APIs to facilitate healthcare data exchange and enable healthcare services; it also provides flexible solutions to fit with various incumbent system architectures. Fundamentally, FHIR is regarded as a 'platform specification' from the perspective of constructing a health information system. FHIR offers solutions to the challenges encountered in health data usage, such as describing, sharing, and managing to respond to the numerous healthcare data needs. Because of the complex context of data applications, the developers are often required to make deliberate design decisions when adopting FHIR as a data communication standard over the incumbent health information systems. This sub-section discusses the common adopted architectural patterns for integrating FHIR with different local existing systems.

FHIR is flexible to fit different system architectures in order to serve specific purposes. For example, FHIR can be adopted as a lightweight client to deal with data requests from mobile apps in front of local health information systems or a heavyweight client for a robust analytics healthcare data solution to deal with multiple local health information systems and provide system interoperability. FHIR can either fit in a centralised server architecture or a peer-to-peer data sharing solution depending on the information systems' scale and the autonomy requirements. In general, there are three architecture patterns in the approach of FHIR-enabling an existing solution.

This first architectural pattern is the ‘interoperability wrapper’. In most implementations, FHIR is adopted as an interoperability interface, where FHIR sits on top of existing systems and interacts directly with clients. The interoperability interface architecture, shown in Figure 23, is widely adopted to wrap EHR recordings and clinical terminologies to enable their data users to access healthcare data in consensus information models (Saripalle *et al.*, 2019, Curran *et al.*, 2020, Gøeg and Hummeluhr, 2018). This architecture could be regarded as an FHIR-compliant API, which can be employed by other parties in the healthcare ecosystem to improve the data harmonies in a healthcare computing environment.

In this architecture, the data access layer explains the structures of proprietary clinical databases it connected to the requests of data accessing; the FHIR adapter translates clinical data from the proprietary database to FHIR resources; and FHIR API supports information exchange in the paradigms if RESTful, Message, Document or Services (detailed explanation of these paradigms in section 4.2.2.1). When requests need to interoperate with multiple clinical databases, the FHIR is employed to achieve code reuse and health information portability between clinical databases. In some cases, this system architecture has a proprietary API adapter parallel with the FHIR API adapter to satisfy specific business requirements.

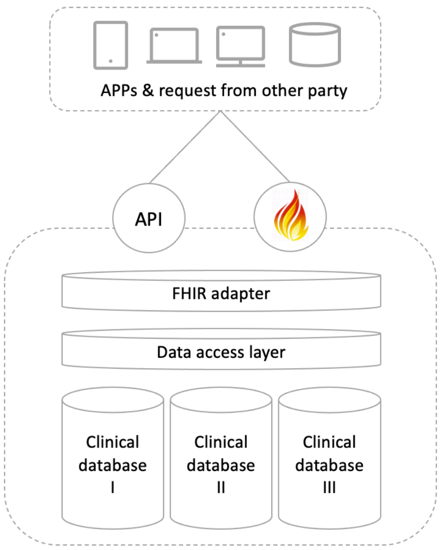


Figure 23 FHIR as an interoperability interface

The second architectural pattern is the 'broker adapter', shown in Figure 24. In the healthcare ecosystem, healthcare data resides in many different systems and are required by other organisations in a variety of formats. For example, health organisations are needed to send data to payers' systems or syndromic surveillance public health agencies for business and management purposes. In another scenario, lab results are required to integrate into other information systems; this architecture is also employed to transform the data formats.

Such format exchange requirements that improve the healthcare ecosystem's interoperability are usually fulfilled by interoperability vendors exclusively. They provide operational support and usually do not persistently store those data. To facilitate healthcare data exchange, the broker system adopting FHIR resources as the underlying domain models can greatly simplify the broker system's design (Rinner and Duftschmid, 2016, Deppenwiese *et al.*, 2021, Dolin *et al.*, 2021). The high demand for such format conversion is that the healthcare information systems are extremely complex, and multiple versions of standards coexist in the incumbent information systems. In particular, the two widely adopted standards, HL7 v2 and HL7 v3, often coexist.

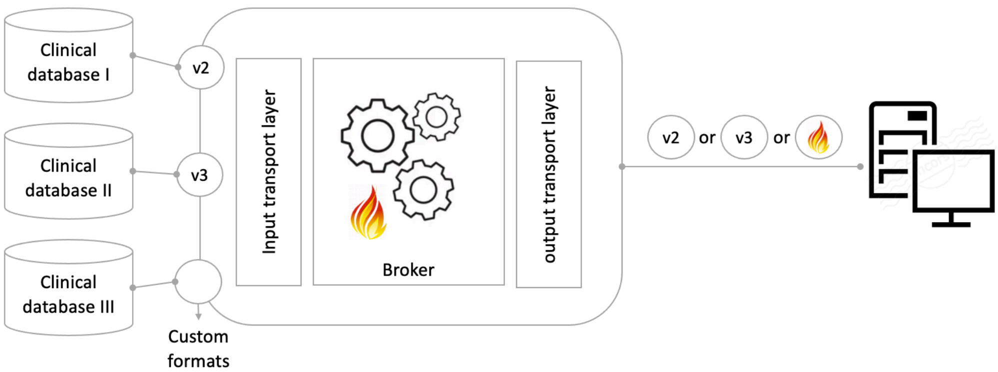


Figure 24 FHIR as a broker adapter

The third architectural pattern is the 'FHIR-based clinical data repository', shown in Figure 25. In order to promote innovations in healthcare services, the neutral health data repository is one of the commonly adopted solutions, which integrates health data from dispersed information systems and realise unified data storage and management (Ulrich *et* *al.*, 2016, Gruendner *et al.*, 2020, Shi *et al.*, 2021). The neutral clinical data repository is regarded as a health data warehouse to support computing and analytic demands. FHIR is adopted as a consensus standard to facilitate data transformation from various formats and standards to unified information models. The extension and profiling capacity of FHIR satisfy the transformation requirements.

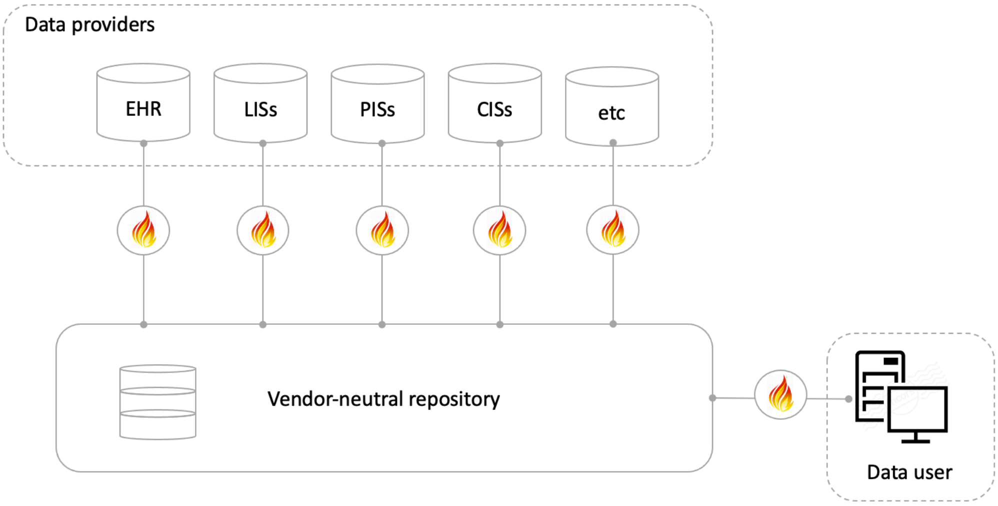


Figure 25 FHIR-based clinical data repository

In the above three architectural patterns, FHIR flexibly integrates with the incumbent information systems to support various healthcare innovations by playing different roles in health data provisioning. In contrast, the challenge the three architectures have in common is that they all require converting industry-standard healthcare data formats or proprietary formats to FHIR resource models. Due to the flexible design of FHIR itself and the fact that the FHIR resource models do not cover all healthcare data usage scenarios, the implementors are granted considerable freedom to decide how to map local data with FHIR resources. The diversity and even incompleteness of local medical data deteriorate the uncertainty of such data mapping. Even with the same local data, different implementors have inconsistent understandings of the FHIR *resource* models and their constraints, which also leads to inconsistent data mapping. Section 4.3 will discuss the inconsistent understanding of FHIR specifications in detail, and Chapter 5 will explore the fundamental cause of this issue.

4.2.4 Technical Factors Popularising FHIR

As the applications, such as telehealth, are rapidly growing in the industry, the need for information exchange based on commonly used Internet interfaces has driven the widespread adoption of FHIR in the industry. In contrast to the earlier standards of HL7 v2, v3 and CDA, FHIR is likely to rapidly gain attention from the industry because it actively embraces Internet technologies and offers advantages such as agility, fast iteration, and low learning costs (Bender and Sartipi, 2013, Zong *et al.*, 2021, Xu *et al.*, 2020, Leroux *et al.*, 2017), with additional support for mobile applications (Mandel *et al.*, 2016, Bender and Sartipi, 2013, Sayeed *et al.*, 2020). Moreover, FHIR adopts a RESTful API to enable interactions and represents data in the currently popular JSON (JavaScript Object Notation) format instead of the EDI and XML format proposed by the earlier standards. FHIR provides a set of standards with established patterns to improve interoperability among a wide range of systems and devices which transcend EHR (Electronic health record) systems. FHIR to heterogeneous healthcare information systems is akin to the TCP/IP standard to the Internet. FHIR significantly reduces the difficulty of the transformation of incumbent information systems, and its implementation compared to OpenEHR and the previous versions of HL7 (Bender and Sartipi, 2013) is significantly simplified.

In addition to the RESTful interface, FHIR resources can be exchanged through the paradigms of Document, Messaging and Services; these comprise the three types of resource collections serving different purposes (McKenzie, 2016). For current solutions, FHIR is usually adopted as a front-end server, expressing the local healthcare data with the term ‘resources’; it provides an HTTP/REST interface for applications by developers to access data (Saripalle *et al.*, 2019). Heterogeneous databases mutually communicate through their front-end servers. These FHIR servers are oriented toward each other, establishing an unimpeded network of intercommunication through RESTful API at the technical and syntactical levels. The semantic interoperability between heterogeneous databases is theoretically ensured by FHIR resources, which constitute unified information models to ensure that all agents communicate via the same discourse system.

In 2018, six Internet giants, namely Amazon, Google, IBM, Microsoft, Oracle and Salesforce, jointly committed themselves to the elimination of interoperability barriers in healthcare by adopting FHIR as an exchange standard (Information Technology Industry Council, 2018). FHIR is selected as the basis of this research because it has been adopted as the national standard across all hospitals in the United Kingdom (UK) (NHS, 2020) and has also been widely adopted in other sectors.

Due to the increasing availability of new health data and the advent of the "app" economy, consumers and physicians must now be able to share information in a way that is both lightweight and real-time, utilising current internet technologies and industry-standard protocols. In this context, FHIR is a reaction to requests for a way that can share a huge amount of health data more efficiently, promptly, and easily. As FHIR is built on Internet standards that are widely used in industries other than healthcare, by utilising existing standards and technologies that are already familiar to software developers, FHIR dramatically reduces the entry barriers for new software developers to meet healthcare demands. In conclusion, FHIR standardises the procedure through which information can be represented and sent across multiple organisations and parties.

4.3 Limitations of FHIR

The widespread adoption of FHIR has led to an increased debate on the limitations of semantic interoperability; Kubick (2016) and Kraus (2018) discuss the semantic ambiguity introduced by the implementors due to different combinations of FHIR *resources* being used to explain the same healthcare service. When FHIR is adopted as an 'interpretation wrapper' in a healthcare ecosystem for information exchange, all parties to it are able to choose FHIR *resources* to represent their healthcare services. In consequence, different institutions may not be able to interoperate due to the inconsistencies in the use of FHIR *resources*; semantic ambiguity in communication is introduced and amplified by the interactive process. This chapter will explore semantic ambiguity from two perspectives. The first concerns the rigour of FHIR specifications, and the second explores the semantic ambiguity due to FHIR's flexible architecture.

4.3.1 Insufficient Consistency of FHIR Specification Description

Three types of semantic ambiguity have been identified in FHIR implementation, with the first caused by the insufficient rigour of the FHIR specification. Beale (2019) contends that the inconsistency in definition of FHIR produces semantic ambiguity. The following examples have been found in FHIR v4.3.0:

- The same semantic with the different lexical name

Dosage in Medication Statement has the same meaning as Dosage Instruction in Medication Dispense (Figure 26). The three elements, Location.hoursOfOperation, HealthcareService.availabletime, and Slot.start, are different names although they appear to designate the same thing (Figure 27).

MedicationStatement.dosage	
Element Id	MedicationStatement.dosage
Definition	Indicates how the medication is/was or should be taken by the patient.
Cardinality	0..*
Type	Dosage
Summary	false
Comments	The dates included in the dosage on a Medication Statement reflect the dates for a given dose. For example, "from November 1, 2016 to November 3, 2016, take one tablet daily and from November 4, 2016 to November 7, 2016, take two tablets daily." It is expected that this specificity may only be populated where the patient brings in their labeled container or where the Medication Statement is derived from a MedicationRequest.

MedicationDispense.dosageInstruction	
Element Id	MedicationDispense.dosageInstruction
Definition	Indicates how the medication is to be used by the patient.
Cardinality	0..*
Type	Dosage
Summary	false
Comments	When the dose or rate is intended to change over the entire administration period (e.g. Tapering dose prescriptions), multiple instances of dosage instructions will need to be supplied to convey the different doses/rates. The pharmacist reviews the medication order prior to dispense and updates the dosageInstruction based on the actual product being dispensed.

Figure 26 Dosage vs Dosage instruction

Location.hoursOfOperation	
Element Id	Location.hoursOfOperation
Definition	What days/times during a week is this location usually open.
Cardinality	0..*
Summary	false
Comments	This type of information is commonly found published in directories and on websites informing customers when the facility is available. Specific services within the location may have their own hours which could be shorter (or longer) than the locations hours.

HealthcareService.availableTime	
Element Id	HealthcareService.availableTime
Definition	A collection of times that the Service Site is available.
Cardinality	0..*
Summary	false
Comments	More detailed availability information may be provided in associated Schedule/Slot resources.

Slot.start	
Element Id	Slot.start
Definition	Date/Time that the slot is to begin.
Cardinality	1..1
Type	instant
Summary	true

Figure 27 Hours of operation vs available time vs start

- The same lexical name with the different semantic

The ‘substitution’ in Medication Request and Medication Dispense describes two different actions (Figure 28).

MedicationRequest.substitution	
Element Id	MedicationRequest.substitution
Definition	Indicates whether or not substitution can or should be part of the dispense. In some cases, substitution must happen, in other cases substitution must not happen. This block explains the prescriber's intent. If nothing is specified substitution may be done.
Cardinality	0..1
Summary	false

substitution	0..1	BackboneElement	Any restrictions on medication substitution
allowed[x]	1..1		Whether substitution is allowed or not ActSubstanceAdminSubstitutionCode (Example)
allowedBoolean		boolean	
allowedCodeableConcept		CodeableConcept	

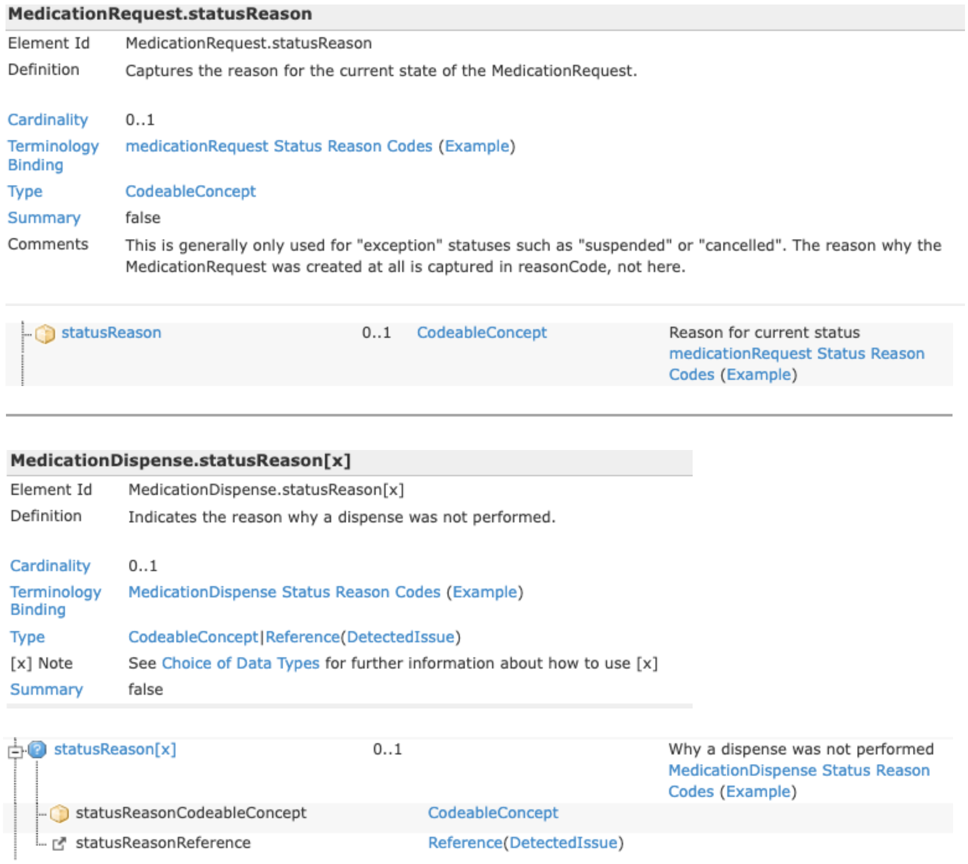
MedicationDispense.substitution	
Element Id	MedicationDispense.substitution
Definition	Indicates whether or not substitution was made as part of the dispense. In some cases, substitution will be expected but does not happen, in other cases substitution is not expected but does happen. This block explains what substitution did or did not happen and why. If nothing is specified, substitution was not done.
Cardinality	0..1
Summary	false

substitution	0..1	BackboneElement	Whether a substitution was performed on the dispense
wasSubstituted	1..1	boolean	Whether a substitution was or was not performed on the dispense
type	0..1	CodeableConcept	Code signifying whether a different drug was dispensed from what was prescribed ActSubstanceAdminSubstitutionCode (Example)
reason	0..*	CodeableConcept	Why was substitution made SubstanceAdminSubstitutionReason (Example)
responsibleParty	0..*	Reference(Practitioner | PractitionerRole)	Who is responsible for the substitution

Figure 28 Different meanings of same lexical name

- The same lexical name and semantics but different data structure

The ‘status reason’ in Medication Request is defined as a Single-valued attribute; in Medication Administration, it is a container attribute, and in Medication Dispense, it includes two sub-elements.



MedicationRequest.statusReason	
Element Id	MedicationRequest.statusReason
Definition	Captures the reason for the current state of the MedicationRequest.
Cardinality	0..1
Terminology Binding	medicationRequest Status Reason Codes (Example)
Type	CodeableConcept
Summary	false
Comments	This is generally only used for "exception" statuses such as "suspended" or "cancelled". The reason why the MedicationRequest was created at all is captured in reasonCode, not here.

statusReason	0..1	CodeableConcept	Reason for current status medicationRequest Status Reason Codes (Example)

MedicationDispense.statusReason[x]	
Element Id	MedicationDispense.statusReason[x]
Definition	Indicates the reason why a dispense was not performed.
Cardinality	0..1
Terminology Binding	MedicationDispense Status Reason Codes (Example)
Type	CodeableConcept|Reference(DetectedIssue)
[x] Note	See Choice of Data Types for further information about how to use [x]
Summary	false

statusReason[x]	0..1		Why a dispense was not performed MedicationDispense Status Reason Codes (Example)
statusReasonCodeableConcept		CodeableConcept	
statusReasonReference		Reference(DetectedIssue)	

Figure 29 Same lexical name and semantics but different data structure

These imprecise definitions inevitably lead to misuse or inconsistency in the implementation of FHIR; further, the FHIR specification involves terminology in a variety of fields and is relatively complicated. The lexical definitions do not have the capacity to guide implementers to precisely match FHIR resources to idiosyncratic local databases because it is impossible for the FHIR specification to describe all mapping scenarios; this is an inherent flaw of the lexical approach, which is discussed in Chapter 5 will from a theoretical perspective.

4.3.2 Semantic Ambiguity Introduced by FHIR Profiling

The second type of semantic ambiguity is introduced by FHIR extensions. As the 80/20 rule of FHIR resources (HL7 International, 2019a) is adopted to avoid overlap and redundant definition of resources, lesser-used terms can be freely defined by implementors in the format of extension of resources, which accounts for 20% of clinical terminologies. Because

FHIR unifies healthcare resources but lacks explicit contextual constraints, the 80/20 rule enables an institution to define its own extensions for the same healthcare service. An issue of this nature both causes barriers to information exchange and also obstructs medical discoveries based on cross-institutional data analysis (Dolin *et al.*, 2018). Semantic interoperability particularly deteriorates when extensions of resources are used to deal with specialist health data.

4.3.3 Semantic Ambiguity Introduced by FHIR Flexibility

The third type of semantic ambiguity is due to the freedom and flexibility FHIR offers to implementers; they can employ FHIR resources, or combinations of them, in order to interpret healthcare services, even though some may not be mature and/or stable, which leads to semantic ambiguity. FHIR v4.3.0 defines 139 resources, of which 100 belong to non-clinical categories. The number, which increases with every release, grants a substantial degree of freedom to implementers to use these resources. For example, in Figure 30, the resources of 'observation' can be combined with other resources to represent laboratory results, imaging study findings, diagnostic test results, vital signs, and other physical examination findings. These are the combinations of FHIR resources suggested by HL7 (https://www.hl7.org/fhir/resourceguide.html). The potential misuse of resources occurs when healthcare data is beyond the scope of HL7-suggested combinations. Additionally, FHIR implementers can tailor FHIR integrations to specific business needs, as shown in **Appendix-1**, resulting in multiple customised resource collections occurring between different systems (Dolin *et al.*, 2018, Jiang *et al.*, 2016). The interoperability issue caused by diversified FHIR collections is recognised by HL7 International (2022).

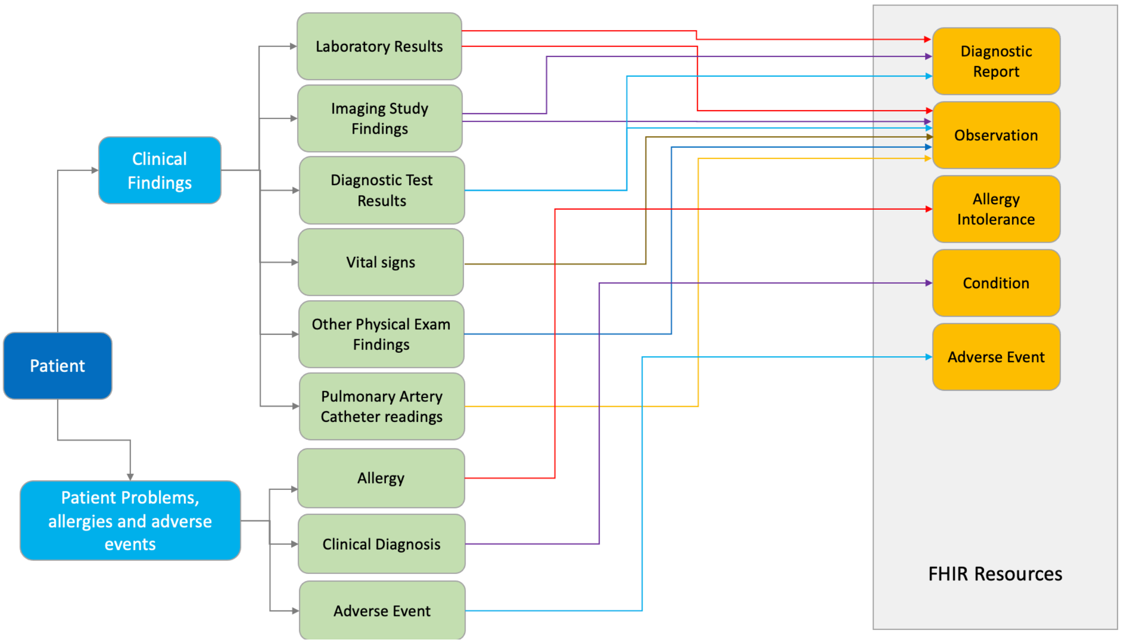


Figure 30 An example of correspondence between clinical actions and FHIR resources

HL7 also addressed such issues on the webpage of the FHIR specification (HL7, 2022b). Hl7 recognised that the lack of rules to prohibit resources from duplicating used leads to the resource inconsistency issue. The variations in size, complexity and comprehensiveness of the existing FHLR resources result in the resource granularity issue. Regarding these issues, the implementer's safety check list (HL7, 2022d) and the FHIR resource considerations (Lloyd McKenzie, 2020) could help for achieving the consistency of both resource content and granularity to some extent, however, these actions are not enforced to implementors and the criteria for test the consistency are unclear.

An irreconcilable relationship always exists between the design and the practice, particularly when the designer pursues the architectural virtue and flexibility of the design architecture. The data discrepancies, inconsistencies, and the different context constraints of data bring a severe difficulty to the design of describing all the scenarios in practice. Furthermore, the inconsistency is amplified when FHIR is applied in the practice of healthcare across different organisations and practitioners because various organisations adopt different granularity of concepts in the healthcare ecosystem.

Since FHIR is still an emerging standard, HL7 subjectively avoids excessive constraints and rigour to maximise initial adoption rates; and HL7 considers further improving the consistency when FHIR approaches a final normative standard because the resource consistency and granularity catch considerable attention in practice as well as in academia.

4.4 Recent Development in FHIR Compliance Solutions

Regarding the semantic ambiguity introduced by FHIR profiling (Section 4.3.2), HL7 launched an office website (http://hl7.org/fhir/registry/) to manage the extension publication. Once an extension defined by implementers is approved by HL7, it is shared with all FHIR users through an official channel; this centralised management method effectively unifies and standardises custom extensions. However, the use of extensions nevertheless faces the problem of interoperability caused by implementers' contrasting understandings of lexical definitions. McClure *et al.* (2020) propose a framework to guide harmonisation among multiple FHIR users in terms of terminology, data elements, measure clauses, and measure concepts. Tute *et al.* (2021) take a similar approach, proposing a data quality assessment method for the support of collaborative governance.

The approach of ensuring FHIR conformity through review processes is usually costly in terms of time and labour. Sayeed *et al.* (2020) take an alternative approach, proposing an application which automatically merges patient-generated health data, represented by FHIR resources, into EHR. This approach is effort-effective but its scope is limited to patient-generated health data and it does not cover electronic health records, which is the most complex aspect of healthcare ecosystems. Pfaff *et al.* (2019) contribute mapping scripts for the interpretation of medical data with FHIR resources; in their study, the data from the Integrating Biology & the Bedside (i2b2), the Patient-Centred Outcomes Research Network (PCORnet), and the Observational Medical Outcomes Partnership are automatically encapsulated by the FHIR. However, the script compatibility issues caused by the idiosyncrasies of local data sources persist. In their framework, the adaption of a local database to the mapping script is allocated at the local database layer, therefore, the inconsistency of using FHIR resource caused by different implementers remains unresolved.

Another approach is to leverage the national effort to harmonise the FHIR resources for medical data representation across hospitals. The Medical Informatics Initiative (MII) and local data integration centres (DICs) in Germany collaborate to standardise COVID-19 data in FHIR profiles through another set of models, i.e., the German Corona Consensus Dataset (GECCO). Using FHIR, GECCO defines 83 data elements, and has been extended to all hospitals nationwide (Rosenau *et al.*, 2022). The United Kingdom adopts the same approach and proposes FHIR UK core (NHS Digital, 2021) to enable consistent information flows across borders. However, the disadvantage of this approach is the lack of agility and the high cost of upgrading.

In industry, a more straightforward approach is adopted; a developer collaboration and publishing platform (Firely, 2015) plays the role of coordinator and facilitator among developers to improve the conformance of FHIR, constituting a loosely-regulated approach. The above approaches have their own advantages and disadvantages; this research seeks to develop a low-cost and high-efficiency method by which to ensure FHIR conformity.

Table 10 summarises the benefits and constraints of existing FHIR compliance solutions in terms of cost, efficiency, implementation difficulty, and application breadth. The suggested ostensive information architecture has clear advantages over current solutions.

Table 10: Comparison between FHIR Compliance Solutions

Solution	Time and cost	Efficiency to decrease ambiguity	Easy to Implement	Scope of application
HL7 official website (http://hl7.org/fhir/registry/)	Low	***Low***	Easy	Wide
A framework for harmonisation (McClure *et al.*, 2020, Tute *et al.*, 2021)	***High***	High	***Hard***	Wide
Automatic tools (Pfaff *et al.*, 2019, Sayeed *et al.*, 2020)	Low	High	Easy	***Narrow***
FHIR resources harmonisation national wide (NHS Digital, 2021, Rosenau *et al.*, 2022)	***High***	High	Easy	Wide
A developer collaboration and publishing platform (Firely, 2015)	Low	***Low***	Easy	Wide
An ostensive information architecture	Low	High	Easy	Wide

In view of the necessity for FHIR to use a consensus-based approach, this study considers the related work of ontology used as an artefact by which to promote the harmonisation of health information systems. Ontology artefacts play a critical role in the fields of medical terminology unification, cross-medical protocol interoperability, and information exchange between heterogeneous systems for healthcare services.

4.5 Ontology in Healthcare Information Systems

The term ‘ontology’ originates from philosophy, which is defined as ‘an account of being in the abstract’ (Bailey, 1969). “Ontology is the study of the categories of things that exist or may exist in some domain” (Sowa, 1983). With knowledge representation and reasoning becoming one of the promising domains of Artificial Intelligence (AI), ontology is widely adopted as its sophisticated method for knowledge acquisition, sharing and reusing (Gruber, 2009, Liu *et al.*, 2008) and decision support (Gua and Liu, 2019). From the recent work in the fields of AI and information systems, an ontology is employed as content-specific agreements in knowledge engineering for sharing and reuse or adopted as a tool to solve the interoperability issue between systems (Schriml and Mitraka, 2015). Similar cases of using ontology can be found in the domain of Health Information Systems (HIS).

In HIS, multiple systems such as hospitals, clinics, laboratories, pharmacies, telemedicine, and biosensors have disparate objectives and processes presented by different granular medical information. That results in semantic heterogeneity (Ahmadi *et al.*, 2019). Therefore, semantic interoperability is complex and fundamental in health information system integration. As the requirements of domain knowledge representation, ontology is recognised as a promising approach to building interoperable information systems (Daraio *et al.*, 2016). In general, two types of ontology-based methods are usually practised in handling interoperability. Sharing a consensus terminological ontology among all applications is the first type of solution, which is normally used in a specific domain with clear boundaries. Ontology matching techniques are used to align the difference between the internally used terminology of applications and the terminology of the standard. The second type aims at multiple terminology systems. A logic-based ontology is constructed as an interlingua to enable these heterogeneous systems to communicate in an unambiguous fashion. The semantic relationships of the dispersed systems are specified by the logic-based ontology; therefore, the two-way semantic mapping can be automatically calculated.

According to the survey conducted by Wache *et al.* (2001)on how ontology is used in the integration of information systems, Wache et al. concluded three types of approaches based on the role ontology played in system integration. They are the single ontology approach, the multiple ontology approach and the hybrid ontology approach, which depend on the degree of semantic heterogeneity. Diaz *et al.* (2017) implement a comprehensive review on the purposes of adopting ontologies in health information system integration and categorises them into four types: information share, diagnostic processes improvement, clinical systems integration, and management process optimisation, which systematically summarised the functionalities of ontology from information acquisition, information exchange, to information integration. It provides readers with a comprehensive perspective to understand the role of ontology in HIS. Ontology mapping techniques used for enabling interoperability are systematically reviewed by Choi *et al.* (2006), and these techniques have been widely employed in various works (Blobel, 2011, Dhombres and Bodenreider, 2016).

Considering the characteristics of medical data and the incumbent clinical standards, this research examines the ontology used in HIS from the information system architecture point of view. Gradually upward from the leaf nodes, the healthcare ontology can be regarded as consisting of medical terminology, clinical concept and healthcare services. Accordingly, ontology plays the role from the unification of medical terminology to enable inter-exchange between systems with different standards, where a cross-system clinical concept is born, up to the top level to underpin the information exchanged among heterogeneous systems with the purpose of supporting healthcare services.

- Ontology for terminology: representing terminological and taxonomic aspects of medical knowledge

  Ontology has been used to unify the medical professional interpretation in order to establish an agreement on medical terminologies in diverse clinical systems, such as

LOINC and SNOMED-CT. From this perspective, terminology ontologies are the pre-defined agreements designed to standardise the language of a domain, providing each term with a precise meaning and a specific granularity. Therefore, terminology ontologies lay the foundation for the exchange of medical information.

- Ontology as a bridge between two systems: representing the relationships between terminology ontologies

  Regarding the clinical concepts that comprise the medical terminologies defined by different standards, 'bridge' ontologies are adopted to facilitate information exchanges between terminology ontologies, and are employed to unify the definitions of clinical concepts. Ryan (2006) interconnects HL7 v3 and SNOMED-CT through ontology matching, and Bodenreider (2008) uses the same method to enable SNOMED-CT to understand the laboratory test result coded in LOINC. Those similar operations facilitate semantic interoperability between heterogeneous coding systems and also support the integration of dispersed health information systems (Plastiras *et al.*, 2014).

- Domain ontology: providing a common knowledge base for healthcare ecosystems

  As the clinical historical data, latest laboratory results and demographic data are all dispersedly stored, a call for a unified repository for all relevant data is urgent to obtain a comprehensive picture of a patient's health status. There is an enormous amount of work to be done in this field. In general, the relevant works can be divided into two categories.

  The first category is constructing a novel domain ontology to enable the information exchange between heterogeneous health systems. Kataria and Juric (2010) propose a hierarchical ontology, which elaborates the categorization of taxonomy and relevant assertions from the perspective of the use of data to support healthcare services to enable interoperability of clinical data structured in different formats and stored in dispersed repositories. This method can only interpret ontological concepts and their relationships at a high abstract level. It is difficult to cover the ontological concepts at a lower level, for example, clinical concepts. Thus, this solution is not universal and requires multiple ontology mapping calculations when it is used in different systems.

The second category is leveraging the published international standards in healthcare domain, such as FHIR. Compared to the healthcare domain ontology proposed by individual researchers or national institutions, the FHIR-based ontology has evident international influence and the advantage of wide promotion. To encourage FHIR's adoption, the following studies propose methods for the transformation of healthcare data into the corresponding HL7 FHIR structure (Kiourtis *et al.*, 2019, Jiang *et al.*, 2017b). A significant amount of research effort has been devoted to the improvement of FHIR coverage scenarios. Beredimas *et al.* (2015) propose an OWL (Web Ontology Language) ontology that defines the primitive and complex data types of the FHIR framework and the validation rules to enable FHIR to express data information externally to traditional medical databases.

El-Sappagh *et al.* (2019) extend FHIR to the telehealth scenario, introducing real-time sensor data into the historical EHR medical data, with the aim of providing more comprehensive patient data to clinical decision support systems. Similar works have been carried out by Peng and Goswami (2019), combining data generated from the Internet of Things (IoT)-empowered smart home devices to EHR; meanwhile, Mavrogiorgou *et al.* (2019) collect multi-dimensional data reflecting patients' health. This type of research (Wagholikar *et al.*, 2017, Moreira *et al.*, 2018) extends the application of FHIR to a broader range of medical data, promoting the wider adoption of FHIR

Although there is so much research work to promote FHIR as a common standard, especially the use of FHIR-ontology to strengthen the ability of semantic interoperability. It has been difficult to find a similar work that deals with the limitations of FHIR (Kubick, 2016, Kraus, 2018).

This section provides evidence that ontology artefacts are widely adopted, with the aim of improving data harmonisation and accessibility, and FHIR-based ontology is a mainstream approach to contend with the ever-increasing complexity of healthcare ecosystems.

Thereby, this research can explore the FHIR conformity solution on the basis of FHIR ontology artefact. The next section delves into the root cause of semantic ambiguity in FHIR implementation.

Chapter 5 An Ostensive Approach to Elaborate Semantics

Concerning the semantic ambiguity in FHIR implementation, this research investigates a novel information architecture that satisfies the data and process requirements of healthcare data stakeholders and adapts to incumbent health information systems. This chapter begins by examining how information interacts within information systems. To appreciate the fundamental reasons of semantic ambiguity, semiotics gives a methodology for understanding the link between sign and the meaning it expresses. Then, the development of the ostensive information architecture will be described in depth. In Section 5.2, micro-services and service-oriented architecture are discussed, establishing the groundwork for the Semantic Engine. The design concepts of these two architectures provide theoretical support for the deployment of healthcare terminologies of various granularity to satisfy the needs of data analysis and healthcare applications. After introducing the federated architecture, Section 5.3 elaborates on the design philosophy used to create the ostensive information architecture. Section 5.4 describes the suggested ostensive information architecture, whereas Section 5.5 describes the development process for an ostensive information architecture based on FHIR.

5.1 Information Interaction through Lexical and Ostensive Approach

This section examines the semantic ambiguity generated by FHIR implementation from the perspective of the interaction between information systems and their users. This section first discusses the use of signals in human communication and then compares lexical and ostensive approaches for reducing semantic ambiguity.

5.1.1 Signs in Communication

By investigating the practice of information management and human communication, this research recognises that the fundamental cause of semantic ambiguity lies in the use of signs to represent objects. When implementers want to apply the FHIR standard to their own complex and heterogeneous local databases, their understanding of the FHIR standard will naturally be affected by the characteristics of the local data, resulting in possible inconsistencies in the application of FHIR. Thus, this chapter will examine the problems with the theory of the using signs in communication.

A sign can be anything which is taken as substitution for something else' (Eco, 1979), particularly in human communication. In semiotics, researchers examine information interaction through the study of signs and their effect on the human actors involved. Multiple semiotic theories hold different stances of epistemology and have laid different cornerstones in communication. They profoundly impact the fields of informatics (Liu *et al.*, 2010, Liu and Li, 2015, Liu and Tan, 2014), information systems(Baxter *et al.*, 2018, Brödner, 2019, Liu *et al.*, 2014, Liu, 2005), knowledge management (Holzinger *et al.*, 2014) and artificial intelligence (Targon, 2018, Chartier *et al.*, 2019, Staab, 2019).

Saussure's theory of sign was initiated from the thought of a dichotomy (Figure 31). He believed that a sign links signifier and signified, which may exist in a material form and concept. The signifier in his theory is something that explicitly exists and can be distinguished by the human senses (Leeds-Hurwitz, 1993). In comparison, Peirce reckoned that the existence of an interpretant is critical and must be introduced in the process of making sense of a sign, which he terms as a semiosis. An interpretant directly connects a sign and an object, while the sign and the object are dotted linked (Figure 32) in Peirce's triadic model. The dotted line in the figure indicates that the correspondence between the sign and the object is not objectively determined but dependent on the context and purpose of the communication and hence subject to personal interpretation. The interpretant can be regarded as the effect of such a sense-making process (Chandler, 2017) through the use of signs in different contexts or for different purposes (Liszka, 1990, Savan, 1987). Therefore, between a sign and an object, there is no strict one-to-one correspondence as suggested in Saussure's model; although most specifications of information sharing adopt the Saussurean model of static mapping between lexicons and objects, including FHIR.

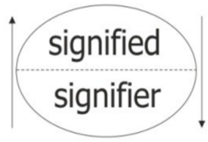


Figure 31 Saussurian dyadic model of sign (Chandler, 1994)

Peircean semiotics emphasises the effect of using signs in the context (Staab, 2019). By emphasizing the subjectivity in the mapping between the sign and the business context in which the sign is used, the triadic model of semiosis offers a theoretical basis for an ostensive approach to pinpoint the meaning of the sign (i.e., semantics) and its effect of the sense-making of the sign (i.e., pragmatics).

Peirce's triadic model is particularly useful in examining the reason for semantic ambiguity which may be generated in the implementation. When each implementor intends to apply FHIR to his local health information system, the sense-making process, using FHRI resources to explain healthcare data in a heterogeneous local database, may differ.

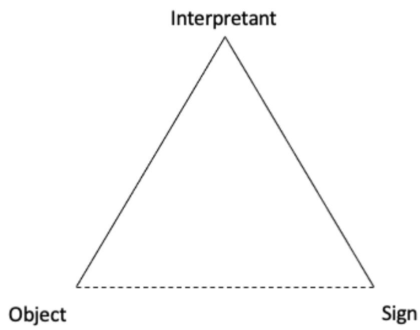


Figure 32 Peirce's triadic model (Peirce, 1958)

FHIR specification uses the lexical approach to explain the definitions of resources. In other words, FHIR interprets the meanings of resources in language, which can be understood as the 'sign' (as illustrated in Figure 32). In the context of FHIR implementation, different implementers may have contrasting understanding of FHIR definition, leading to the same FHIR resource being used to explain different clinical data. Semantic ambiguity is created when many interpreters illustrate the same concept with contradicting signs. The ambiguity of "same semantic with different lexical names" is depicted in **Figure 33**. Similarly, the ambiguity of 'same lexical name with different semantics' occurs when the same sign is mapped to multiple objects by different interpretants. Just as Dolin *et al.* (2018) addressed, the primary challenge of FHIR adoption is to transform multiple distributed local datasets into consistent FHIR formats.

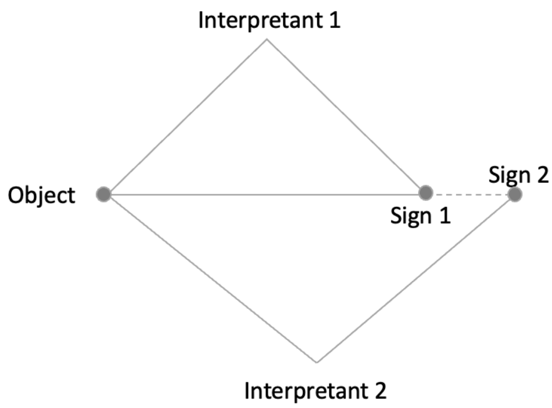


Figure 33 Peirce's triadic model to explain semantic ambiuity

The semantic ambiguity in FHIR implementation addresses the limitation of the lexical approach in communication. This problem severely hinders the exchange of healthcare data across organisations. The following section will explore the improved semantic interoperability brought by an ostensive approach to supplement the lexical approach.

5.1.2 Lexical and Ostensive Approach to Information Exchange

An ostensive approach in philosophy refers to denoting the meaning of a concept by offering examples of things to which the concept applies (Wittgenstein, 2019). When a semantic meaning passes between people, the concept can be defined verbally and also can be demonstrated by an example. It is more intuitive and effective to directly denote an example for things that are abstract or not easy to describe in language, such as sapphire blue.

In a case where one FHIR resource is mapped to different clinical data, revealing the mapping relationships helps implementors determine the meaning of localised semantics. For example, the mapping relationship shows which attribute in which local database corresponds to which parameter of FHIR. From such an approach, the semantic ambiguity caused by a symbol corresponding to multiple objects can be resolved. This approach, which shows the correspondences between FHIR and local databases as examples to facilitate semantic communications, is named an ostensive approach in this research. The examples further clarify the understanding of FHIR specifications by implementors and reduce semantic ambiguity, especially when complex signs such as FHIR are involved.

Based on the above exploration, this research study recognises the limitations of the lexical approach to the resource as defined in FHIR and extended with an ostensive treatment in the use of the resources. As the JSON format of FHIR naturally gains the advantages of graph-based knowledge representation, this research considers building a centralised FHIR-based ontology to illustrate the meaning of resources. When FHIR is employed as a healthcare ontology, the medical terminology, clinical concept, and healthcare services defined by FHIR are represented by nodes and their relationships in healthcare ontology.

In order to further clarify the understanding of implementors for the FHIR resource, the healthcare data in local databases can be retrieved as examples to show the sign-object correspondence. Thus, the FHIR-based ontology and mapping data can simultaneously provide users with semantic explanations and data examples.

The following chapter will focus on building an ostensive architecture for dispersed healthcare information systems using the FHIR knowledge graph.

5.2 MSA Inspired FHIR Knowledge Graph

HIS are complex because they are highly dependent on the context of domain knowledge (Mainzer, 2007), surrounded by subjective norms (Kannampallil *et al.*, 2011), and employed by multi-stakeholders for multiple purposes (Plsek and Greenhalgh, 2001). More and more personal health analysis devices, such as wearable sensors and telerehabilitation, are gradually accepted as a new paradigm for personal healthcare. In such a ubiquitous environment, multiple HIS users require an information architecture to support massive data exchange and various data analytic tasks. In the healthcare ecosystem (Figure 34), numerous stakeholders collaborate as accounts of both data users and providers. The data they provided describe the healthcare status from different dimensions, and the information they requested serves various analytics purposes, leading to the data being abstracted at different levels (Blobel *et al.*, 2006).

The requests for data at different levels of abstraction desire cooperation between the various entities in the healthcare ecosystem. As the challenge of increasing complexity in the healthcare ecosystem, the architecture of healthcare information systems needs to be evolved.

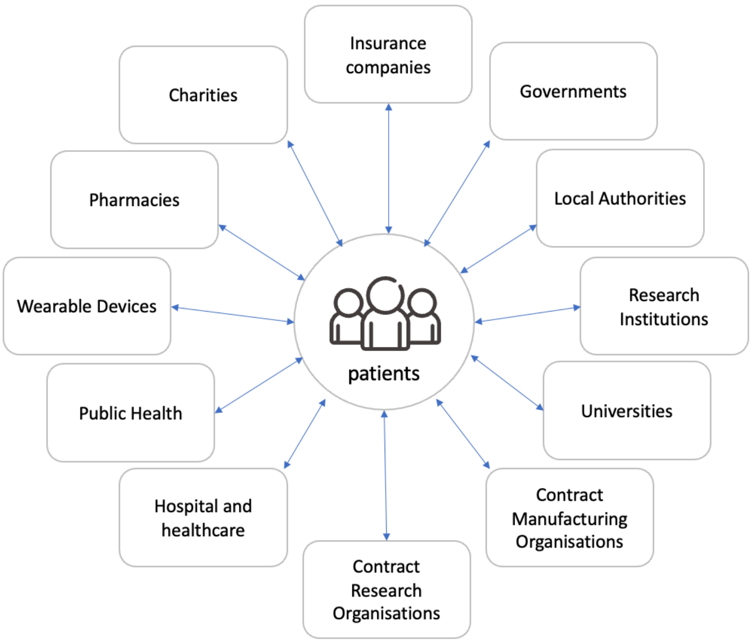


Figure 34 Multi stakeholders in healthcare ecosystem

Appendix-2 provides a comprehensive review of the evolution of information architecture for large-scale software systems. Through the review, it is certain that micro-services architecture (MSA) MSA is an effective solution for large-scale systems that can quickly process external requests. The essence of Microservices design is dividing a whole service provisioning into multiple individually deployable services. The single responsibility principle (Dijkstra, 1974), performing a single task as a suite of small services, enables task distribution in large-scale systems; the '**loosely coupled, highly cohesive**' design principle (Stevens *et al.*, 1974, Myers, 1978) provides feasible guidance for the division of service modules. Thus, MSA makes it possible for large-scale software systems to respond quickly to business processing needs.

This research adopts FHIR to facilitate message exchange at the semantic level. When applying MSA design philosophy to FHIR, FHIR specification can be regarded as a set of healthcare terminologies composed of different level granularities. The following will discuss the multi-hierarchical nature of FHIR fitting with MSA, including how healthcare concepts are organised in FHIR and the FHIR knowledge graph inspired by MSA.

In healthcare information systems, raw data about patients' health are collected in the format of event logs. In order to support the secondary uses, raw data need to be compiled and synthesised. As shown in Figure 35, healthcare concepts are located at different levels, and the hierarchical abstraction of words' meanings can be used for different purposes of data use.

In Figure 35, the concept of PCR (polymerase Chain Reaction) test is a laboratory technique to identify whether a patient is infected with the Covid-19 virus. This concept is usually adopted by patients or doctors to understand the Covid positive or negative. At the lower level, there are three concepts that connect the PCR test. They are DNA polymerase, Taq polymerase and PCR primers. Hospitals or research institutions need the values of DNA polymerase, Taq polymerase and PCR primers to identify the validation of the PCR test. The combined result of these three values is the test result of a PCR test. In the context of doctor-patient communication, all three concepts are shielded. While at the top level, the public health centres only care about the statistical results of sampling PCR tests in a certain area, and the individual test results are not vital for disease control. Therefore, the abstraction level of data for disease surveillance and strategic decision-making services is higher than that of data used for individuals' health diagnoses.

This example shows semantics for different purposes mapping to various levels of data abstraction. Engaging data at varying levels of abstraction masks irrelevant details and promote data reuse. Thus, data use is related to the level of data abstraction. The right side of Figure 35 indicates that entities with different purposes use the data at different levels of abstraction.

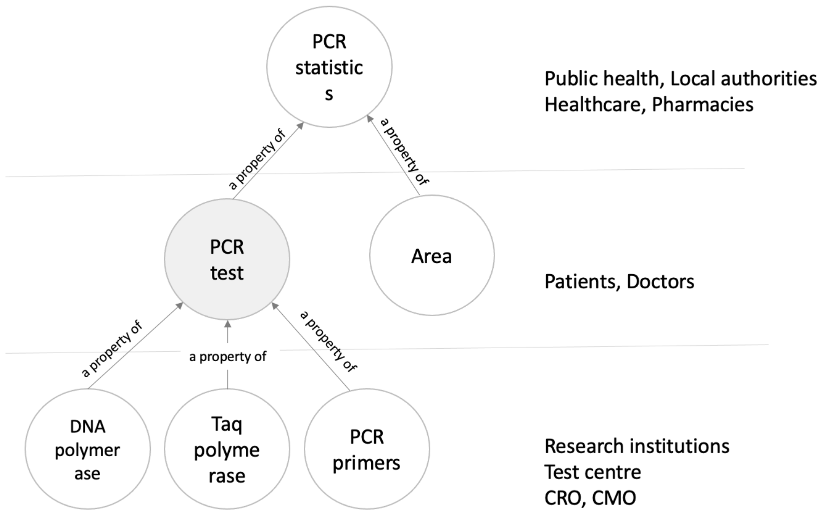


Figure 35 Semantic structure of PCR test

In this example, atomic concepts, such as DNA polymerase, Taqpolyme rase and PCR primers, are the basic nodes that cannot be subdivided further without losing their senses and functionalities. The nodes establish the semantic structure at various levels of granularity and the composition rules for medical information models and healthcare services.

This research adopted the FHIR specification to explain healthcare concepts, so that the healthcare concepts defined by FHIR can be represented in hieratical relationships for flexible data requirements. This research divides healthcare concepts into three categories: *medical terminologies*, *clinical concepts*, and *healthcare services* from the bottom to the top level. Each healthcare concept can be represented as a node, and the links between nodes reflect the hieratical relationships. Each note is **loosely coupled** with others, but its semantics are **highly cohesive** as its relationships with its subordinating nodes are relatively stable. Therefore, each node can be regarded as a *Microservice* entity to meet the demands of different data-use purposes. Besides the basic node, all nodes can be self-evolving in semantics by adding, deleting, or changing their subordinating nodes. The basic nodes have to change the name of the node in order to change their semantics.

Based on this already formed expression of nodes and their relationships, the Knowledge Graph (Ehrlinger and Wöß, 2016) is adopted to illustrate *medical terminologies*, *clinical* *concepts*, and *healthcare services* at various levels. In order to better fulfil the data access requirements, this research employs Neo4j to build an FHIR knowledge graph because its rich built-in functions can support semantic computing and reasoning, which can be leveraged by the semantic queries at different granularities. In summary, aiming for semantic queries from various stakeholders in the healthcare domain, the FHIR knowledge graph is selected to respond dynamically to the queries of healthcare concepts by showing the node-edge topology.

The FHIR knowledge graph only contains healthcare concepts and is used to handle semantic queries; therefore, the FHIR knowledge graph can be regarded as a consensus ontology in the healthcare domain. As discussed in Section 4.5, the FHIR domain ontology provides the knowledge base for a healthcare ecosystem and can be used to integrate dispersed information systems. The next will discuss the federated architecture, which connects the FHIR knowledge and multiple geographically isolated health information systems.

5.3 Federated Architecture in Healthcare Domain

A healthcare ecosystem is complex, which has multiple stakeholders and dimensional healthcare data dispersed in various autonomous or semi-autonomous information systems. Sharing data across institutions is a significant challenge because of the concerns regarding data privacy, confidentiality, and safety, as well as the distribution, heterogeneity, and autonomy of these health information systems.

The most straightforward solution for the effective sharing of data is to create a vendor-neutral archive (VNA) system (Figure 25) to allow different users to access the shared data repository. The benefits of NVA are significant, particularly in supporting patient-centred care (Sirota-Cohen *et al.*, 2019, Pantanowitz *et al.*, 2018) or evidence-based medicine (Sox and Greenfield, 2009, VanLare *et al.*, 2010) from a data provisioning perspective. However, the challenges faced by NVA are enormous. Concerns about data privacy and VNA management issues lead to numerous challenges and problems in VNA implementation because each entity exhibits varying levels of autonomy and heterogeneity. Sirota-Cohen *et* *al.* (2019) summarised ten points to address these problems, which indicates that VNA architecture is feasible for small-scale data sharing within an individual organisation, such as sharing medical image data across systems. Still, it is difficult to implement large-scale systems across organisations.

Considering the limitations of NVA architecture, this research explores federated architecture (FA) as a solution to take both data sharing and data privacy needs into account. FA is adopted extensively where multiple heterogenous systems need to be integrated while respecting the security and compliance requirements of each distinct system (Sheth and Larson, 1990).

Federated architecture is a pattern in enterprise architecture that aims to reconcile the needs for data integration among decentralised databases. FA usually is adopted to cooperate with a collection of autonomous and possibly heterogeneous databases, balancing organisational autonomy with application needs (Heimbigner and McLeod, 1985). The architecture of a federated database system is shown in Figure 36.

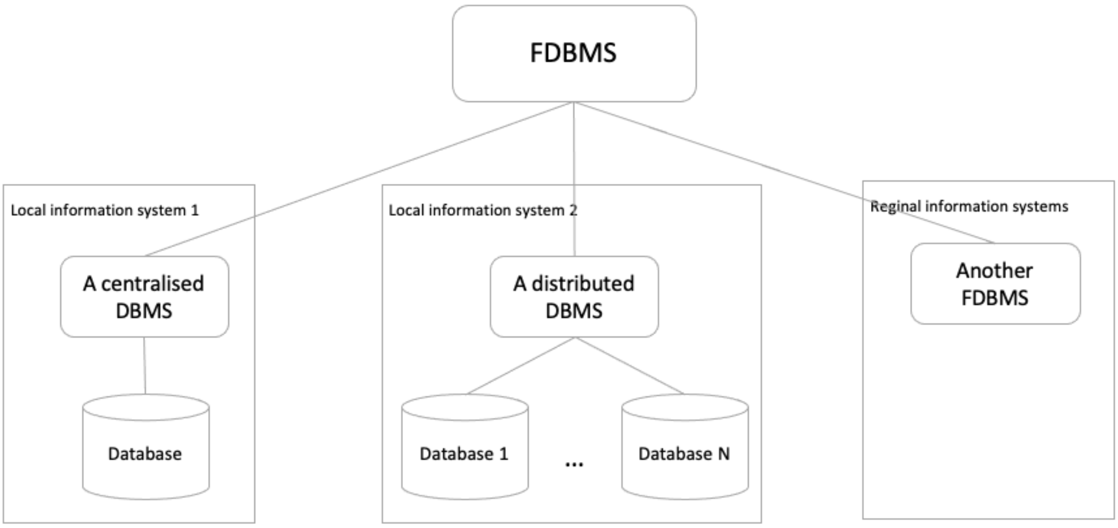


Figure 36 A federated database systems (Sheth and Larson, 1990)

FA aims to achieve cooperation among existing heterogeneous and autonomous information systems. As shown in Figure 36, an FDBMS (federated database management system) connects with dispersed local database systems. Healthcare data are stored on geographically distributed information systems in a healthcare ecosystem. Each of them is heterogeneous from the other mainly due to technical differences, for example, the difference in hardware or software solutions. Therefore, data structure, contextual constraints and even the data query languages in each information system are diverse. Each one exhibits different levels of autonomy to fit with its service provisioning and system management demands. In order to enable these systems to share data with each other while retaining their respective existing data management models, and to reduce the cost of system integration systems, FA is an ideal choice. FDMBS is added on the top of the logical structure and connects with the dispersed local information systems with different levels of data schema.

The schema architectures include three-level (Tsichritzis and Klug, 1978), four-level (Templeton *et al.*, 1987) and five-level (Sheth and Larson, 1990) data description architectures addressing the understanding of logical data layers with various emphasises. This section draws on a five-layer model to understand how to connect multiple existing health information systems.

Figure 37 exhibits the method of integrating local databases through layer-by-layer data schema abstraction. The bottom layer is the local schema expressed in the native and customised data model. The component schema translates the local schema into a common data model to facilitate upper-level data integration. The export schema is a subset of the component schema to reconcile the component schema to the federated schema. Introducing the component schema is to adapt the accessing demands for specific federation users and facilitate control of association autonomy. Thus, as shown in Figure 37, the mapping relationship between the export schema and the federation schema is M to N. The federated schema integrates multiple export schemas to support the needs of various groups of users or applications. The top layer is the external schema, defined as the data schema from a class of users' or applications' points of view. To better facilitate user/application access to data in specific contexts, the external schema can be flexibly designed to cater to the needs of adding additional constraints and access control.

Figure 37 shows the M-to-N relationships among the export schema, the federated schema, and the external schema, reflecting the architectural adjustments made by the databases to suit different users and applications. As addressed by Sheth and Larson (1990), the redundancy between external and federated schemas usually happens in practices. External schemas can be merged into federated schema to simplify the implementation. In the context of this research, FHIR is the consensus of health data representation, which reconciles the various needs of users and can be regarded as the combination of the federated and external schema. Accordingly, the dispersed local health information systems contain the customised export, component, and local schema, respectively.

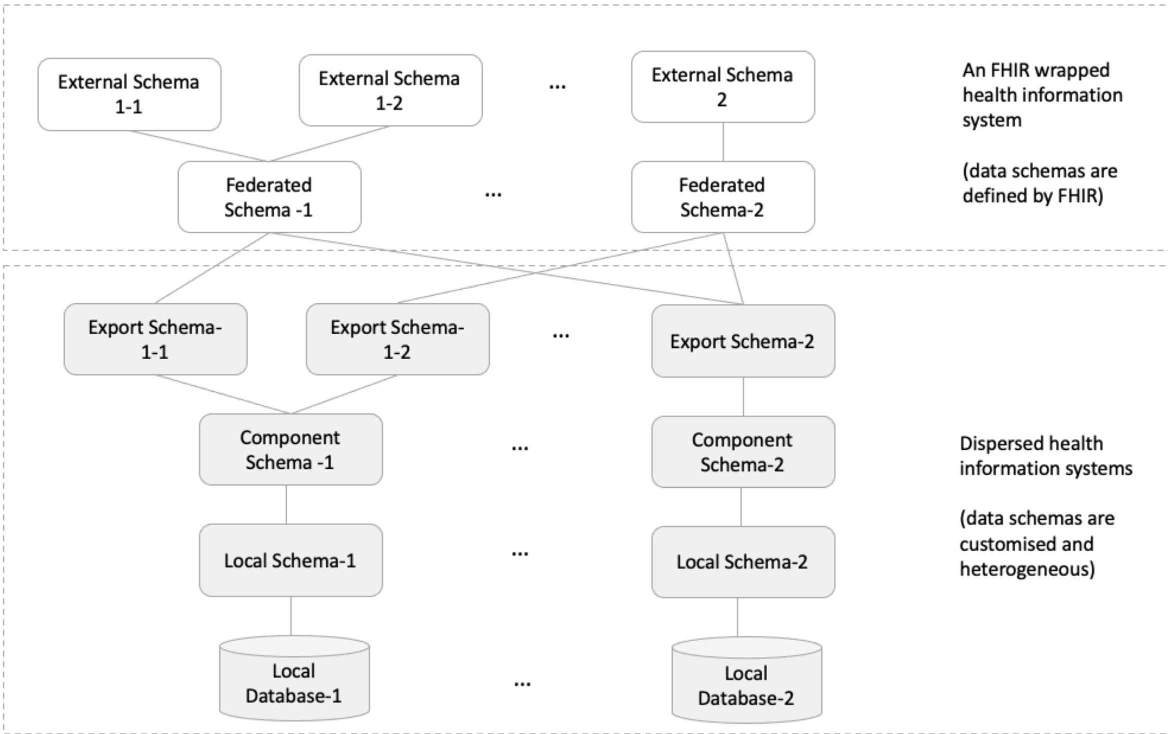


Figure 37 A five-level schema architecture (Sheth and Larson, 1990)

The Five-level schema architecture provides the methodology to develop a federated architecture either adopting a bottom-up or top-down approach (Sheth, 1988) which includes the design from local databases and gradually integrated by different granularity data schemas up to the top level. Another approach to achieving the federated architecture is to simply add a layer of software above existing databases.

This research will adopt the extra-layer approach to maximise the preservation of existing medical systems and reduce integration costs. FHIR will be adopted as the combination of federated and external schema to integrate dispersed health information systems in this research. Therefore, FHIR knowledge graph is the top layer in a federated architecture to provide the domain knowledge in terms of unified health concepts definitions. The dispersed heterogeneous health information systems exchange information with FHIR defined formats. By now, the full picture of the federated architecture has been revealed.

In summary, the FHIR knowledge graph is the top logical layer to explain the semantics of healthcare concepts through the node-edge format in this federated architecture. All semantic queries are processed by the FHIR knowledge graph. The healthcare data and the mapping relationships between local data and FHIR resources can be retrieved from the dispersed local information systems as an example when further denoting the semantic meaning is needed. Therefore, the semantic ambiguity introduced by FHIR implementors can be further clarified by mapping relationships and data through this ostensive approach. When the implementor’s understanding of FHIR specification is shown through the mapping relationships and local data, the semantic ambiguity introduced by the different understanding of applying FHIR to local information systems is reduced. The next section will introduce the ostensive information architecture in detail.

5.4 An Ostensive Information Architecture

This research study adopts FHIR as the grounds on which to explain the concepts in the healthcare domain. In response to the limitations of FHIR, an ostensive approach is proposed, which provides clinical data as examples to further explain the semantics defined by FHIR, along with the understandings of different implementors.

In this research, a knowledge graph is constructed on the basis of FHIR, and the FHIR knowledge graph is extended to connect attributes stored in local databases; this is termed the ‘FHIR-centric knowledge graph of the Semantic Engine’, enabling semantic elaboration and reasoning, and the clinical data stored in heterogeneous local information systems can be retrieved by the Semantic Engine. In summary, as shown in Figure 38, the core Semantic Engine is the FHIR knowledge graph; the peripheral consists of attribute nodes in the local datasets. The correspondents between the FHIR resources nodes and local attribute nodes are connected by ‘maps’ lines. The attribute nodes are defined by local implementors, which can be modified and may be immature.

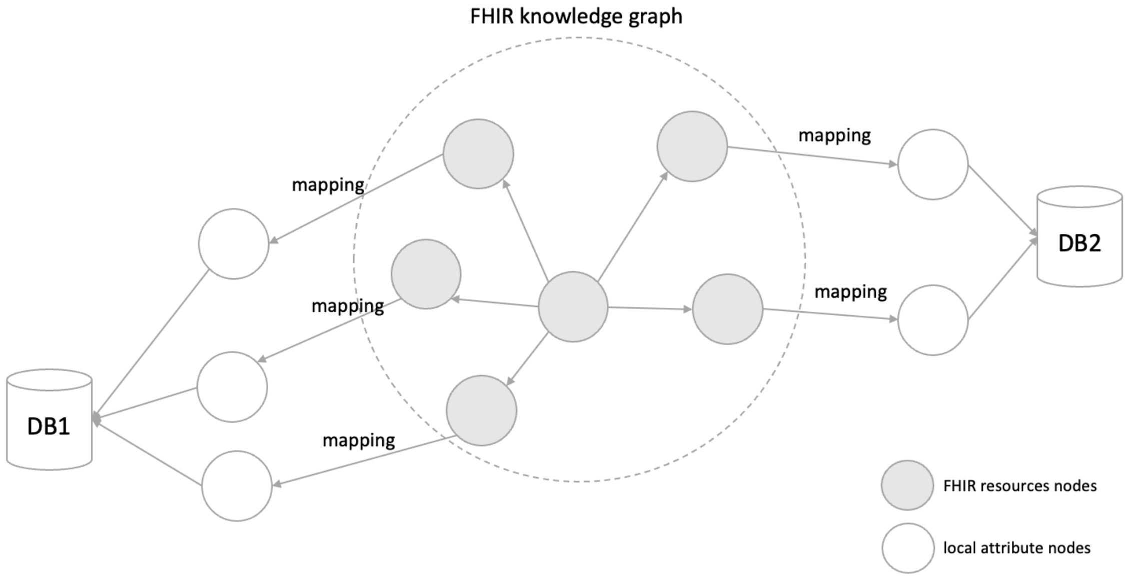


Figure 38 The components of Semantic Engine

The query statements sent by clients to the Semantic Engine reflect their understanding of the FHIR specification through the lexical approach. The data in response to the request ostensibly exhibits the data providers' understanding of FHIR specifications. If there is mismatching between the clients and data providers in terms of the understandings of FHIR, the data in the response can help the user to comprehend the gap. In summary, the proposed information architecture helps users to comprehend the semantic ambiguity produced by the lexical approach through the ostensive examples.

In general, the Semantic Engine is responsible for the processing of all semantics-related tasks. For example, the meaning of a node can be elaborated by the nodes connected with it and their relationships; effectively, the topology graph of this node discloses the node's meaning. Semantic reasoning can be conducted through analysis of the relationships between nodes, for example, the shortest path between two of them.

This paper proposes this semantics-data separated architecture for HISs to support semantic interoperability (Figure 39). The key purpose of this design is to employ FHIR in a computational method that leverages the advantages of the knowledge graph to process semantics and takes privacy and security concerns into account. The separation of semantic processing and data storage can reduce the problem of data privacy leakage caused by the unified storage of data, and access mechanisms based on authorisation further guarantee data privacy; this is discussed in Chapter 6

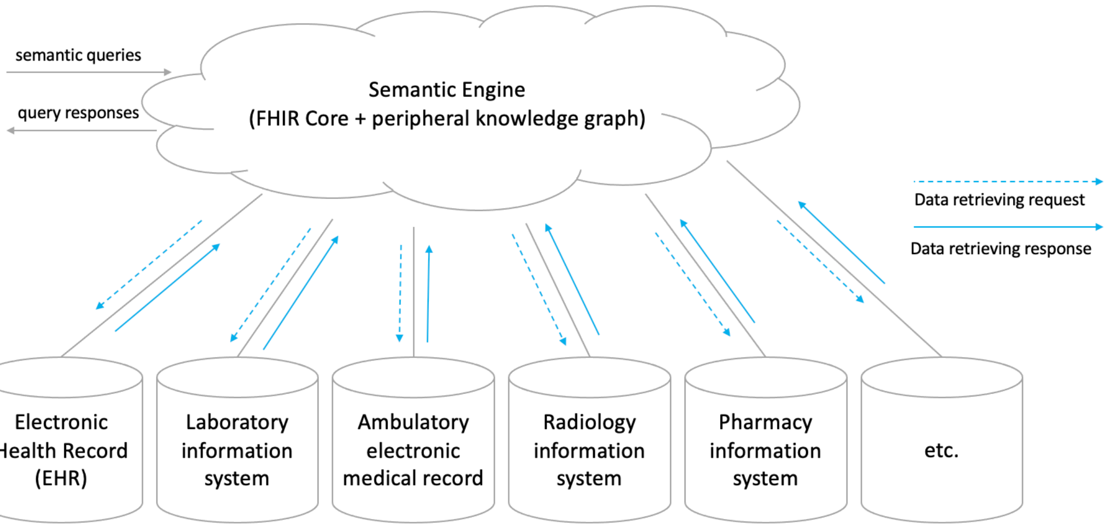


Figure 39 A semantics-data separated information architecture

In summary, the Semantic Engine has two main functions:

1) the enhancement of semantic interoperability across dispersed health information systems by feeding back the JSON file to show the semantic definitions in FHIR along with the understandings by implementors, and:

2) the elimination of semantic ambiguity by providing corresponding examples in the form of data stored in different local health information systems.

5.5 Developing an Ostensive Information Architecture

On the principle of separating semantic processing and data storage, this study positions FHIR in health information systems. In contrast to the use of FHIR as a standard protocol for the transformation of local databases for information exchange (the patterns discussed in 4.2.3), FHIR is abstracted from the front-end of each local information system and unified at the logical top level of the entire health information system (Figure 40).

This study's proposal simplifies the architecture of health information systems by centralising the semantic interpretation layer in order to avoid the ambiguity caused by different interpretants of FHIR, which means that the correspondence between the data provided by the specified primary database and FHIR *resources* is a system-wide standard, and other systems which differ from the standard definitions should follow the semantic interpretation of the primary system. For example, patients' names and addresses could come from multiple clinical systems, but the patient registration system is usually taken as the primary system. When a query requires the ICU (intensive care units) information of a patient to be provided by the Semantic Engine, this will feed back with the name and address from the patient registration system and the relevant ICU information from the intensive care information system, with the information of name and address of this patient stored in the intensive care information system in different formats being ignored. The centralised semantic interpretation layer can be deployed on the cloud to solve the problem of access bottlenecks caused by multiple requests.

Through setting the master-slave relationships between dispersed systems, the system-wide semantics are now unified; in other words, regarding a piece of data to describe a certain patient attribute, there is only one mapping relationship between FHIR and local clinical information systems within an entire ecosystem, even though there are multiple databases storing the same patient attributes. When two peer hospitals make inconsistent use of *resources,* the two different mapping methods are represented as two external graphs to the FHIR knowledge graph. The local implementers of both hospitals can establish a consensus by comparing and selecting.

The architecture of the proposed HISs is shown in Figure 40. To support data retrieval, the Semantic Engine comprises two main elements: FHIR knowledge graph and transformation components.

This architecture contains three layers, in order to respond to the FHIR queries; the semantic interpretation layer is a FHIR knowledge graph which provides an explanation of semantics in lexical definition by nodes and their relationships. The transformation layer works with local health information systems to provide the semantics by example, constituting the data stored in heterogeneous local systems. The **mapping connector** in the transformation layer matches data with FHIR resources, generating conflict alerts if and when data inconsistencies are detected. For example, an alert occurs if a date of birth has been assigned to two data sources through schema matching (Section 5.5.3), or the same concept has been interpreted by different FHIR resources. Therefore, the mapping connector consists of several sub-components.

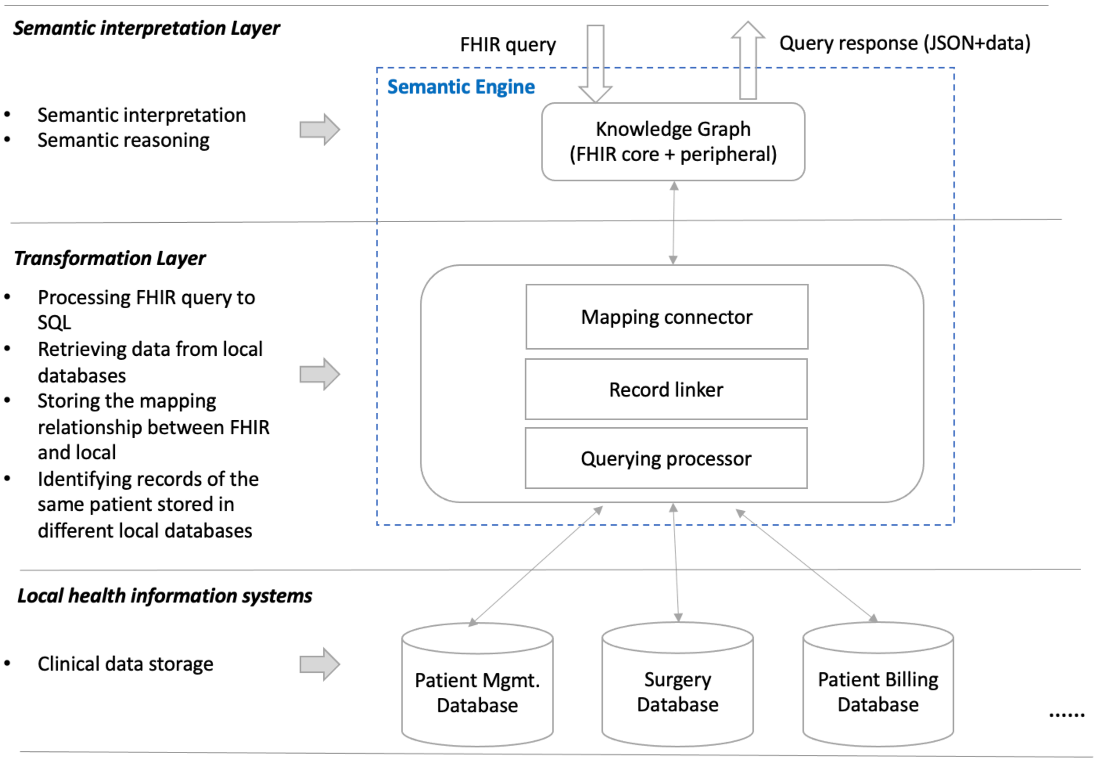


Figure 40 An ostensive architecture of healthcare information systems

An explanation is provided in Section 6.2.2. of the functions that convert the data from two data sources into a unified FHIR-defined format in the **record linker**, which combines the records of the same patient from different databases. For example, the record linker can recognise the records for a patient in a hospital's billing systems and the claim management system of an insurance company, associating the two records. The **querying processor** translates queries from the Semantic Engine and obtains data from local databases. The bottom layer represents the local healthcare information systems where the clinical data are stored.

Figure 41 illustrates the processes of a semantic engine dealing with the semantic query. The FHIR knowledge graph plays a critical role as a user interface and semantic interpreter. Ten internal steps (shown in Figure 41) include redirecting user queries to different local databases, generating query statements, and collecting query results, and merging them to return responses to users. Nie and Roantree (2019) address the question of how to merge the records of different aspects of the same object when they are stored in multiple databases. In this study, the patient profile can be taken as a key variable by which to conduct the record linkage.

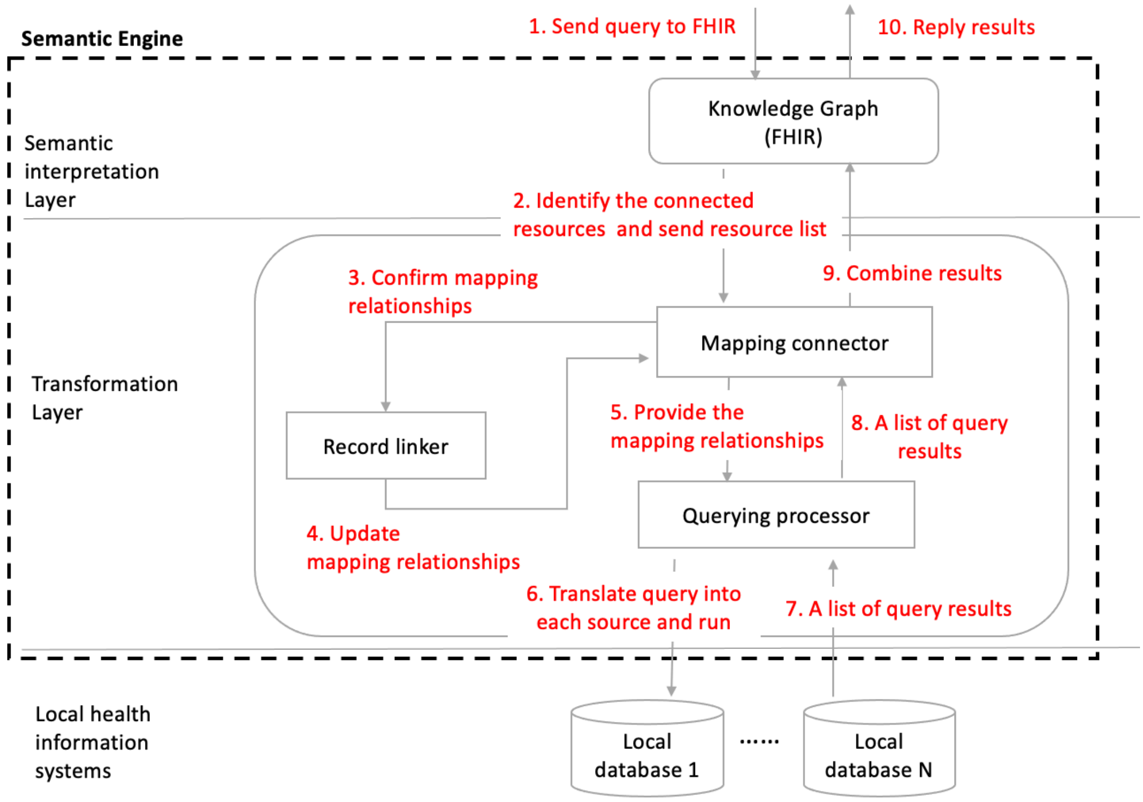


Figure 41 Semantic query processing flow

5.5.1 Semantic Engine

As previously mentioned, the core of the Semantic Engine is an FHIR knowledge graph; this study uses Neo4j for its development. In order to facilitate data exchange between dispersed information systems, the local data require connection to the Semantic Engine. This research transforms the properties of local data into these property nodes; values of local data are retrieved as examples to further render the semantics explicit. This Semantic Engine can support semantic interpretation, semantic computing, and semantic reasoning. This research study focuses on the function of semantic interpretation, which explains concepts to the queries. The details of how the Semantic Engine is structured based on FHIR schema are shown below, along with how local data connections to the Semantic Engine are implemented.

5.5.2 The Construction of an FHIR Knowledge Graph

The JSON representation of an FHIR schema is used to construct the knowledge graph with each defined entity becoming an ***Entity*** node. Each property of the defined entities occupies ***Property*** nodes. Relationships between entities that are defined within the JSON schema become edges within the knowledge graph. Figure 42 details the construction of a knowledge graph from the FHIR JSON schema.

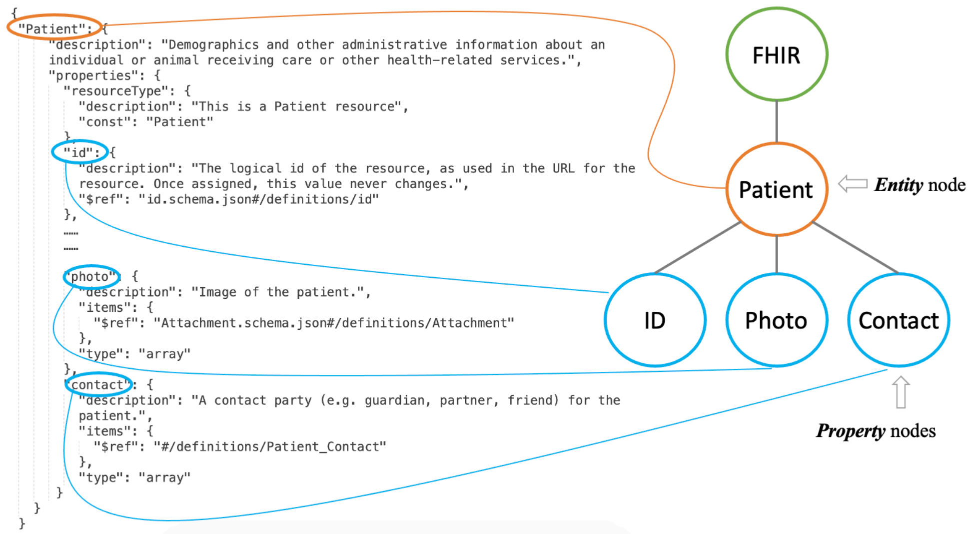


Figure 42 FHIR – Graph mapping

For example, the following Cypher command (Command box 1)can convert resources 'Patient' from JSON format (Figure 43) to node-relationship format (Figure 44). In this research study, the FHIR knowledge graph is constructed by the Cypher command (Lal, 2015) automatically. The software to generate the knowledge graph is available here (Guo, 2023).

Command box 1: Query attributes of 'Patient'

```
Neo4j $ match (a:Table {table_name: 'FHIR.Patient' }) – [r:PROPERTY_OF]-(b:Column) return a,b,r
```

```
{
  "resourceType" : "Patient",
  // from Resource: id, meta, implicitRules, and language
  // from DomainResource: text, contained, extension, and modifierExtension
  "identifier" : [{ Identifier }], // An identifier for this patient
  "active" : <boolean>, // Whether this patient's record is in active use
  "name" : [{ HumanName }], // A name associated with the patient
  "telecom" : [{ ContactPoint }], // A contact detail for the individual
  "gender" : "<code>", // male | female | other | unknown
  "birthDate" : "<date>", // The date of birth for the individual
  // deceased[x]: Indicates if the individual is deceased or not. One of these 2:
  "deceasedBoolean" : <boolean>,
  "deceasedDateTime" : "<dateTime>",
  "address" : [{ Address }], // An address for the individual
  "maritalStatus" : { CodeableConcept }, // Marital (civil) status of a patient
  // multipleBirth[x]: Whether patient is part of a multiple birth. One of these 2:
  "multipleBirthBoolean" : <boolean>,
  "multipleBirthInteger" : <integer>,
  "photo" : [{ Attachment }], // Image of the patient
  "contact" : [{ // A contact party (e.g. guardian, partner, friend) for the patient
    "relationship" : [{ CodeableConcept }], // The kind of relationship
    "name" : { HumanName }, // A name associated with the contact person
    "telecom" : [{ ContactPoint }], // A contact detail for the person
    "address" : { Address }, // Address for the contact person
    "gender" : "<code>", // male | female | other | unknown
    "organization" : { Reference(Organization) }, // C? Organization that is associated with the
contact
    "period" : { Period } // The period during which this contact person or organization is vali
d to be contacted relating to this patient
  }],
  "communication" : [{ // A language which may be used to communicate with the patient about his
or her health
    "language" : { CodeableConcept }, // R!  The language which can be used to communicate with
the patient about his or her health
    "preferred" : <boolean> // Language preference indicator
  }],
  "generalPractitioner" : [{ Reference(Organization|Practitioner|
   PractitionerRole) }], // Patient's nominated primary care provider
  "managingOrganization" : { Reference(Organization) }, // Organization that is the custodian of
the patient record
  "link" : [{ // Link to another patient resource that concerns the same actual person
    "other" : { Reference(Patient|RelatedPerson) }, // R!  The other patient or related person r
esource that the link refers to
    "type" : "<code>" // R!  replaced-by | replaces | refer | seealso
  }]
}
```

Figure 43 JSON format of patient defined by FHIR

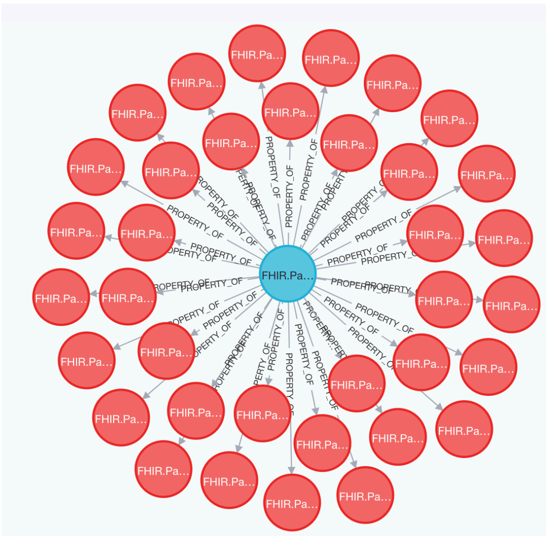


Figure 44 The knowledge graph of *patient*

Similar Cypher commands can be used to convert other *resources*. Since the *resources* are interconnected, the FHIR knowledge graph can be constructed.

5.5.3 Schema Matching

This step is designed to clarify the correspondence between FHIR resources and local data. The knowledge graph of the Semantic Engine comprises a set of nodes, N, and a series of edges, E. This knowledge graph contains not only the low-level mappings for individual data sources but further abstractions of these data providing the capacity to semantically reason. For the remainder of this section, this paper focuses on the schema-matching and schema-mapping components, which are used to provide a basis of interoperability between healthcare systems. To map data stored in dispersed systems correctly to the Semantic Engine, each individual source must be understood in detail; this requires a graph model which can capture the complexity of this individual source.

Figure 45 shows on a high level the nodes and edges required to effectively provide a means of schema mapping. Nodes in the graph represent sources, properties, and mappings while edges are used to denote relations between these. Within the graph there are four node types: source, entity, property and mapping.

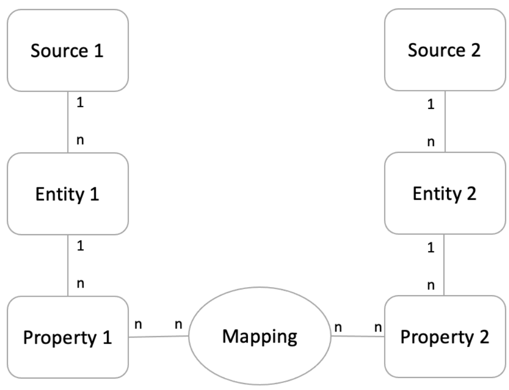


Figure 45 Graph Structure

Source denotes a particular data source, identifying the system from which the data are obtained. This node contains the connecting information for an individual source to facilitate communication with a particular mapping connector. ***Entity*** relates to a particular entity from a data source; within a DBMS (database management system), these may correspond to tables. ***Property*** refers to an entity's attribute such as the name of a patient, which corresponds to columns within an RDBMS (relational database management system); there is a 'one-to-many' relationship between a property and an entity. Finally, ***Mapping***

denotes the way in which two properties between local data sources and the semantic engine may be related.

5.5.4 Mapping Data to FHIR

This study uses MIMIC III (https://mimic.physionet.org/about/mimic/) and a diabetes dataset (https://archive.ics.uci.edu/ml/datasets/Diabetes) as two local health information databases. In the process of mapping the data with corresponding FHIR *resources*, it is often the case that a concept defined by FHIR requires data from multiple MIMIC data tables for ostensive interpretation. Because MIMIC datasets are focused on intensive care medicine, many concepts defined by FHIR cannot be fully explained by MIMIC data, although it can be the case that two data records need to be amalgamated to match an attribute of an FHIR *resource*, or a data record needs to be split into two segments to match the attributes of the FHIR resource. There is also a conflict between the index relationship between the MIMIC database and the FHIR resource, which occurs when querying the health information of an individual patient. Restrictions such as data types should follow the definition of FHIR and be guaranteed by the implementer.

When all datasets within a health ecosystem are matched with FHIR *resources*, it can be said that the health information relating to patients has been semantically connected. By this stage, any stakeholder of health ecosystems can theoretically access all health information relating to a specific patient, therefore patient-centred diagnosis, evidence-based medical research, medical insurance services, public health policy development, and such other associated healthcare related services can be supported by this system.

In order to map data to FHIR, the *structural mapping information* of the data source, a set of *contextual mappings*, and a series of *transformation functions* are all required.

Structural information links entities and their properties within the graph, with each entity and property representing a node. The structural information is either derived from a supplied schema such as an RDBMS or, for flat files, manually supplied by a user. Once the structural information is converted into the graphical format, it can be mapped to the FHIR knowledge graph using the 'Cypher' command for batching processing. This is required in order to overcome differences in terminology and structural differences where one entity in FHIR may be composed of two or more entities within the local data source. This challenge resulted from the semantic ambiguity described in Section 4.3 and is the reason why this study sought to expose the inconsistency of the use of FHIR between its implementers.

The *contextual mappings* denote the context in which a specific data source is to be used; for example, the FHIR schema contains the concept of an "*observation*", referring to medical observation, such as body weight or bone density. While this entity has wide usage due to its generic nature, specific data sources may focus only on a specific measurement. For the diabetes dataset, while it is an *observation* within the FHIR schema, it should only be queried if the user is requesting blood glucose levels.

This requires a mapping which can determine context; it can be achieved by embedding the semantics of the mapping within a *mapping* node. When mapping across data sources, data may require semantic augmentation in order to ensure accuracy. An example is data which provide values for the same entity, such as blood glucose levels, but are represented by differing units of measurement. These inconsistencies are overcome by the use of transformative functions embedded within the mapping nodes linking two properties.

In this research, the MIMIC data was firstly converted into a graphical format using the relational schema; it was then supplemented with manual mappings to FHIR supplied in CSV format for batch processing with Cypher. The diabetes data are a series of flat-files; this representation therefore does not contain the necessary structural information, which was provided by a domain expert. In addition, the diabetes dataset has low dimensionality, requiring the provision of additional *contextual* mappings in order to accurately map the data to FHIR. The schema- and data mapping are performed manually in this research, whereas in industry, developers can use tools to convert local data into the FHIR format in batches (Kiourtis *et al.*, 2019). Regardless of the method used by the implementer, the purpose of this step is to illustrate the corresponding relationships between FHIR and local data in node-edge format.

Chapter 6 Application and Validation of the Semantic Engine

In this section, two case studies are conducted in order to validate that the proposed ostensive information architecture can: (1) decrease semantic ambiguity by showing the data value and its context, and (2) synthesise data from disparate systems with the aim of achieving the patient-centred diagnosis. In section 6.1, there is semantic ambiguity due to the citing of data sources with similar interpretations. This semantic uncertainty can be cleared out by describing the detailed reference paths stored by the Semantic Engine. Section 6.2 demonstrates how the Semantic Engine translates data from two datasets with distinct formats into the FHIR-defined format, thereby facilitating data interchange.

The applications of the Semantic Engine are discussed in Section 6.3, which shows the benefits of the knowledge-graph-based Semantic Engine to support data analysis and management. Section 6.4 discusses the latest solution for FHIR compliance and the summary of the advantages of the proposed information architecture.

6.1 Ostensive Approach to the Enhancement of Semantic Interpretation

FHIR v4 defines 146 types of resources to describe the concepts within the healthcare domain; all resources are represented in JSON format and have naturally have sufficient feasibility to be presented by a knowledge graph. Because Neo4j enables semantic searching and reasoning, the meaning of a concept such as ‘Patient’ can be easily understood through the property node and its relationships. For this reason, the FHIR knowledge graph is termed a ‘core Semantic Engine’. The lexical definition can be searched on the Semantic Engine, while the ostensive examples can also be retrieved by it. The following example illustrates the way in which the ostensive approach supports the reduction of semantic ambiguity.

In FHIR, for example, one of the properties of 'patient' is 'DeceaseDateTime'. Because FHIR has not clearly defined the concept of date of death with context, the possibility of the introduction of semantic ambiguity occurs. In MIMIC datasets, two tables reflect the content of 'death time'. There are three relevant columns in the patient table (Figure 46): DOD, DOD_HOSP and DOD_SSN.

DOD_HOSP indicates the date of death stored in the hospital database, and DOD_SSN refers to the date of death in a social security database. The screenshot to the right of Figure 46 shows that the values of DOD_HOSP and DOD_SSN are different. From the screenshot to the left of Figure 46 it can be deduced that DOD is the combination of records of DOD_HOSP and DOD_SS, and DOD_HOSP has a higher priority for adoption if both values exist.

There is also a DEATHTIME in the Admission Table (Figure 47). Comparison demonstrates that records of death times in the two tables are inconsistent; for example, in *Patient* table, the death time of HADM_ID=9 is 11/14/49 0:00; while in *Admission* table, the record is 11 /14/49 10:15. As the times in all records in the *Patient* table are 0:00, it is assumed that the record in the *Admission* table is more accurate.

Thus, on the basis of the above observations, semantic ambiguity is generated if the data source has not been shown to data users; this leads to misjudgements during data analysis. Semantic ambiguity, a typical type of data quality problem that occurs often, has been identified as the cause of such issues because FHIR specification cannot enumerate all matching situations for local databases.

The ostensive approach has the capacity to reduce this type of ambiguity by providing the sources of data. The source of DOD_HOSP, DOD_SSN and DEATHTIME can be found by retrieving the *table*, *property,* and *source* attributes from the knowledge graph, as a similar process to a 'reverse lookup'.

The following query would return all sources and tables by initially searching for all mappings which link to the FHIR Patient attribute 'deceasedDateTime'

Command box 2: Query the links to "deceasedDateTime"

```
MATCH (src:Source)-(tab:Table)-(prop:Property)-(map:MAP)-(eprop:Property)
WHERE eprop = "deceasedDateTime" RETURN src, tab;
```

In order to identify the source, a query must be run on that source dataset to identify matching values. Such operation is for data users to figure out how the corresponds designed by FHIR implementer, which is beneficial for the data users to use data correctly. For example, for a datetime x and patient ID y this would be converted into the following queries.

Command box 3: Query the sources of "deceasedDateTime"

```
SELECT * FROM Patient where SUBJECT_ID = y and DOD_HOSP = x;
SELECT * FROM Patient where SUBJECT_ID = y and DOD_SSN = x;
SELECT * FROM Admission where SUBJECT_ID = y and DEATHTIME = x;
```

In summary, the Semantic Engine performs the lexical- and ostensive approach through semantic searching facilitated by the FHIR knowledge graph, retrieving data and its context as examples from local clinical systems. Furthermore, specifying primary data sources in local datasets through the construction of mapping relationships can avoid data conflicts in data exchange processes.

SUBJECT_ID	GENDER	DOB	DOD	DOD_HOSP	DOD_SSN	EXPIRE_FLAG
2	M	7/17/38 0:00				0
3	M	4/11/25 0:00	6/14/02 0:00		6/14/02 0:00	1
4	F	5/12/43 0:00				0
5	M	2/2/03 0:00				0
6	F	6/21/09 0:00				0
7	F	5/23/21 0:00				0
8	M	11/20/17 0:00				0
9	M	1/26/08 0:00	11/14/49 0:00	11/14/49 0:00	11/14/49 0:00	1
10	F	6/28/03 0:00				0
11	F	2/22/28 0:00	11/14/78 0:00		11/14/78 0:00	1
12	M	3/24/32 0:00	8/20/04 0:00	8/20/04 0:00	8/20/04 0:00	1
13	F	2/27/27 0:00				0
16	M	2/3/78 0:00				0
17	F	7/14/87 0:00				0
18	M	11/29/16 0:00				0
19	M	1808-08-05 00:00:00	8/18/09 0:00		8/18/09 0:00	1
20	F	6/13/07 0:00				0

SUBJECT_ID	GENDER	DOB	DOD	DOD_HOSP	DOD_SSN
670	M	9/30/80 0:00	2/15/61 0:00	2/15/61 0:00	2/22/61 0:00
834	M	6/15/90 0:00	9/12/66 0:00	9/12/66 0:00	9/20/66 0:00
115	F	1/20/19 0:00	3/24/95 0:00	3/24/95 0:00	3/23/95 0:00
281	F	10/12/41 0:00	10/25/01 0:00	10/25/01 0:00	10/24/01 0:00
294	M	5/21/39 0:00	1/27/19 0:00	1/27/19 0:00	1/28/19 0:00
545	F	3/27/96 0:00	10/25/81 0:00	10/25/81 0:00	10/19/81 0:00
584	F	1814-10-21 00:00:00	4/1/15 0:00	4/1/15 0:00	3/26/15 0:00
1339	F	3/16/71 0:00	2/27/40 0:00	2/27/40 0:00	2/28/40 0:00
937	M	9/24/87 0:00	1/26/63 0:00	1/26/63 0:00	1/25/63 0:00
2452	M	10/16/54 0:00	1/5/13 0:00	1/5/13 0:00	1/6/13 0:00
3612	F	8/30/03 0:00	9/15/86 0:00	9/15/86 0:00	9/16/86 0:00

Figure 46 Patient table in MIMIC data sets

SUBJECT_ID	HADM_ID	ADMITTIME	DISCHTIME	DEATHTIME	ADMISSION_TYPE	ADMISSION_LOCATION	DISCHARGE_LOCATION	INSURANCE
2	163353	7/17/38 19:04	7/21/38 15:48		NEWBORN	PHYS REFERRAL/NORMAL DELI	HOME	Private
3	145834	10/20/01 19:08	10/31/01 13:58		EMERGENCY	EMERGENCY ROOM ADMIT	SNF	Medicare
4	185777	3/16/91 0:28	3/23/91 18:41		EMERGENCY	EMERGENCY ROOM ADMIT	HOME WITH HOME IV PROVIDR	Private
5	178980	2/2/03 4:31	2/4/03 12:15		NEWBORN	PHYS REFERRAL/NORMAL DELI	HOME	Private
6	107064	5/30/75 7:15	6/15/75 16:00		ELECTIVE	PHYS REFERRAL/NORMAL DELI	HOME HEALTH CARE	Medicare
7	118037	5/23/21 15:05	5/27/21 11:57		NEWBORN	PHYS REFERRAL/NORMAL DELI	HOME	Private
8	159514	11/20/17 10:22	11/24/17 14:20		NEWBORN	PHYS REFERRAL/NORMAL DELI	HOME	Private
9	150750	11/9/49 13:06	11/14/49 10:15	11/14/49 10:15	EMERGENCY	EMERGENCY ROOM ADMIT	DEAD/EXPIRED	Medicaid
10	184167	6/28/03 11:36	7/6/03 12:10		NEWBORN	PHYS REFERRAL/NORMAL DELI	SHORT TERM HOSPITAL	Medicaid
11	194540	4/16/78 6:18	5/11/78 19:00		EMERGENCY	EMERGENCY ROOM ADMIT	HOME HEALTH CARE	Private
12	112213	8/7/04 10:15	8/20/04 2:57	8/20/04 2:57	ELECTIVE	PHYS REFERRAL/NORMAL DELI	DEAD/EXPIRED	Medicare
13	143045	1/8/67 18:43	1/15/67 15:15		EMERGENCY	TRANSFER FROM HOSP/EXTRAM	HOME HEALTH CARE	Medicaid
16	103251	2/3/78 6:35	2/5/78 10:51		NEWBORN	PHYS REFERRAL/NORMAL DELI	HOME	Private
17	194023	12/27/34 7:15	12/31/34 16:05		ELECTIVE	PHYS REFERRAL/NORMAL DELI	HOME HEALTH CARE	Private
17	161087	5/9/35 14:11	5/13/35 14:40		EMERGENCY	EMERGENCY ROOM ADMIT	HOME HEALTH CARE	Private
18	188822	10/2/67 11:18	10/4/67 16:15		EMERGENCY	PHYS REFERRAL/NORMAL DELI	HOME	Private
19	109235	8/5/08 16:25	8/11/08 11:29		EMERGENCY	EMERGENCY ROOM ADMIT	REHAB/DISTINCT PART HOSP	Medicare
20	157681	4/28/83 9:45	5/3/83 14:45		ELECTIVE	PHYS REFERRAL/NORMAL DELI	HOME	Medicare
21	109451	9/11/34 12:17	9/24/34 16:15		EMERGENCY	EMERGENCY ROOM ADMIT	REHAB/DISTINCT PART HOSP	Medicare
21	111970	1/30/35 20:50	2/8/35 2:08	2/8/35 2:08	EMERGENCY	EMERGENCY ROOM ADMIT	DEAD/EXPIRED	Medicare
22	165315	4/9/96 12:26	4/10/96 15:54		EMERGENCY	EMERGENCY ROOM ADMIT	DISC-TRAN CANCER/CHLDRN H	Private
23	152223	9/3/53 7:15	9/8/53 19:10		ELECTIVE	PHYS REFERRAL/NORMAL DELI	HOME HEALTH CARE	Medicare
23	124321	10/18/57 19:34	10/25/57 14:00		EMERGENCY	TRANSFER FROM HOSP/EXTRAM	HOME HEALTH CARE	Medicare
24	161859	6/6/39 16:14	6/9/39 12:48		EMERGENCY	TRANSFER FROM HOSP/EXTRAM	HOME	Private
25	129635	11/2/60 2:06	11/5/60 14:55		EMERGENCY	EMERGENCY ROOM ADMIT	HOME	Private

Figure 47 Admission table in MIMIC data sets

6.2 Querying Blood Glucose Levels in FIHR Defined Format

In this section, an example is used to explain how data can be retrieved from multiple institutional EHRs in FHIR format.

This case study queries blood glucose levels from MIMIC and diabetes datasets in FHIR format by use of the Semantic Engine; the query is posed to the system using FHIR terminology. In this instance, all *observations* related to a patient which are blood glucose measurements are the object of the research**.** Command box 4 details the query in SQL format. The *observation* in FHIR is used to model the result of medical observations, while the *coding* property of FHIR is used to denote the type of test. For this example, it is assumed that LOINC codes (McDonald *et al.*, 2003) are used to code medical observations. Within an observation, 'subject.reference' refers to the patient with ID 1.

Command box 4: Query to view all blood glucose levels for patient with ID=1

```
SELECT *
FROM Observation
WHERE coding.code = "2339-0" and subject.reference =1
```

Using the query SQL format, the next step is to query the Semantic Engine in order to determine what sources are required in order to fulfil the query. This is achieved by examining all mapping nodes connected to an FHIR *observation*.

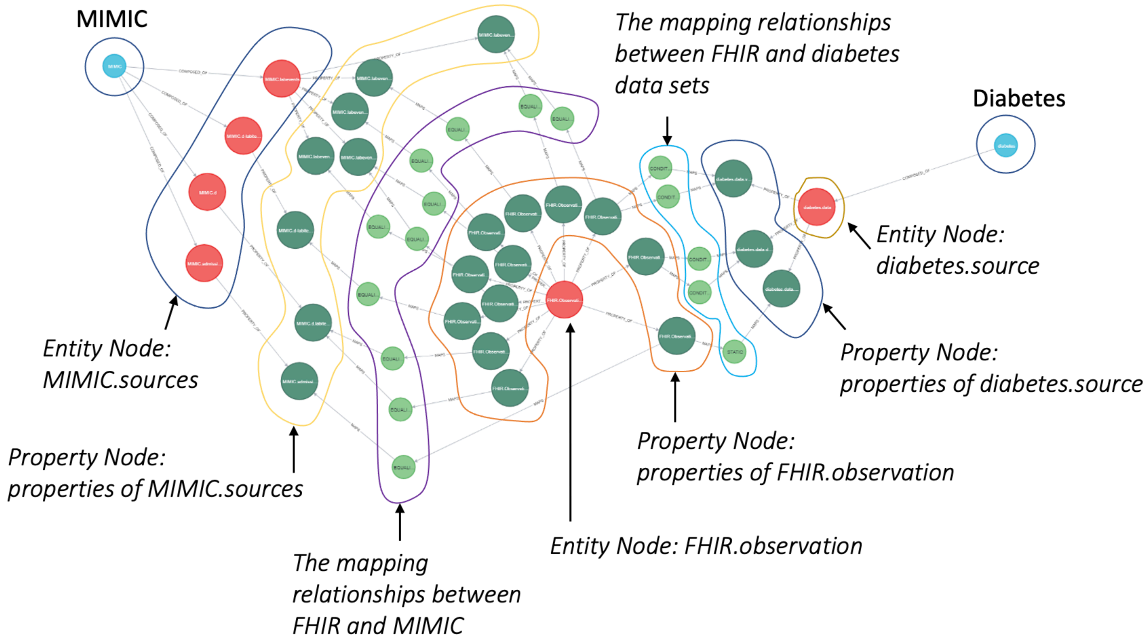


Figure 48 Mappings for FHIR Observation

Figure 48 details the mappings for an FHIR *observation* of both the MIMIC data and the diabetes datasets. The FHIR knowledge graph sits at the centre of the Semantic Engine and remains stable unless FHIR evolves to a new version. The MIMIC and diabetes datasets are connected to the FHIR knowledge graph through schema mapping (Section 5.5.3) and data mapping (Section 5.5.4). When a new data source is connected to the FHIR knowledge graph, the Semantic Engine is updated.

6.2.1 Query Processing

After identification of what source(s) are required to fulfil the query, in this case the diabetes data and the MIMIC dataset, the next step is to translate the query into a format that can be read by each mapping connector.

The diabetes data are a single-source dataset, thus in this instance manual mappings provided by domain expert are required to match FHIR entities to the dimensions within the diabetes dataset schema. This is achieved by examination of the mapping nodes between the diabetes dataset and FHIR shown in Table 11.

Table 11 Mappings for the diabetes dataset

column	map
{"column_name": diabetes.data.Patent1}	{"condition":STATIC, "from":FHIR.Observation-1.identifier, "map_condition": code="2339-0" AND subject.reference = "1"}lines
{"column_name": diabetes.data.value}	{"condition":STATIC, "from":FHIR.Observation-1.identifier, "map_condition": code="2339-0" AND subject.reference = "1"}
{"column_name": diabetes.data.datetime}	{"condition":STATIC, "from":FHIR.Observation-1.identifier, "map_condition": code="2339-0" AND subject.reference = "1"}

From these mappings it can been observed that the 'patient' attribute is a STATIC value embedded within the mappings, the attribute 'value' from 'observation' maps directly to the column 'value' within the diabetes dataset and that the FHIR attribute 'issued' maps to the 'datetime' attribute within the diabetes dataset. The STATIC value represents an annotation to the source data to supply required semantics in order to achieve integration. In this instance as the source dataset only contains the dimensions datetime and value annotated information such as the Patient is required as a static annotation to the source dataset. An example of the dataset with semantic annotations can be seen in Table 15.

Examining these mappings which describe the source data, a comparative SQL query to extract the data from its respective source can be seen in Command box 5.

Command box 5: Translated query for the diabetes dataset

```
SELECT datetime,value

FROM diabetes.data;
```

The MIMIC dataset shows that an *observation* in FHIR is represented by the tables Lab events, Admissions, Patients and D_Labitems within MIMIC. A query for this data requires a join across these tables necessitating knowledge of the MIMIC schema. The knowledge graph can be queried to identify these internal joins to translate the query (Figure 49); a tabular representation of these joins is presented in Table 12 where *entity* refers to FHIR entity, and *property* relates to property of the entity. Each row of the table represents a mapping across entities within FHIR and the properties which join the above-mentioned entities.

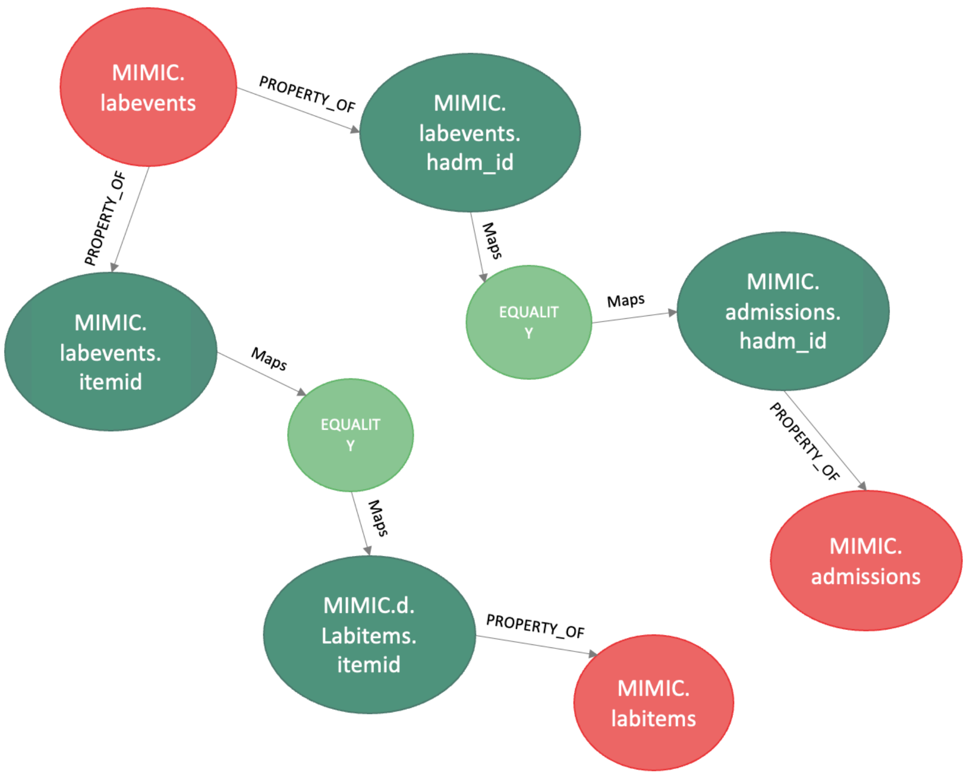


Figure 49 MIMIC internal mappings

Table 12 Tabular representation of MINIC internal mappings presented in Figure 49

Entity	Property
Labevents	MIMIC.labevents,itemid
Labevents	MIMIC.labevents,hadm_id
Labitems	MIMIC.labitems.itemid
Admissions	MIMIC.admissions.hadm_id

By examining these relationships, the query can be translated into a similar one which can query data in the MIMIC schema (Command box 6).

Command box 6: Translated MIMIC query

```
SELECT *
FROM labevents
   JOIN d_labitems ON labevents.itemid = d_labitems.itemid
   Join admissions ON labevents.hadm_id = admissions.hadm_id
Where d_labitems.loinc_code = "2339-0" AND admissions.subject_id = 1
```

The query is now translated and can be passed to the local clinical database for the retrieval of data according to the mapping relationships.

6.2.2 Combining Data from Multiple Data Sources

Each query passed to the local clinical database returns a csv file. The data returned from the MIMIC and diabetes can be seen in Table 13 and Table 14 respectively. The next steps are to convert these files into FHIR format and to integrate them in order to return a unified view.

Table 13 Data returned from MIMIC

Subject _ID	HADM _ID	ITEM ID	CHARTT IME	VAL UE	VALUEN UM	VALUEU OM	FLAG	LABE L	FLUD ID	CATEG ORY	LOINC_C ODE
1	1	5080 9	20/10/20 18 20:04	265	265	mg/dL	abnor mal	Gluc ose	Blood	Blood Gas	2339-0
1	1	5080 9	20/10/20 18 21:51	267	267	mg/dL	abnor mal	Gluc ose	Blood	Blood Gas	2339-0
1	1	5080 9	21/10/20 18 00:42	299	299	mg/dL	abnor mal	Gluc ose	Blood	Blood Gas	2339-0
1	1	5080 9	21/10/20 18 01:46	294	294	mg/dL	abnor mal	Gluc ose	Blood	Blood Gas	2339-0

In the case of sparse data sources, these may **not contain** sufficient information to correctly integrate data into FHIR. For example, the diabetes data (Table 14) contains only two columns, i.e., datetime and value. These data require annotation with static semantic data in order to be integrated with FHIR; these semantic annotations are embedded within the mapping nodes with the static identifier. These STATIC mappings are used in conjunction with the mappings derived from the source to facilitate semantic integration within the knowledge graph.

Table 14 Data returned from diabetes

DATETIME	VALUE
23/10/2018 08:00	354
23/10/2018 18:00	275

For the diabetes dataset, the static data requiring annotation are patient id, the LOINC code, and the unit of measurement. This produces an intermediate csv file, as shown in Table 15.

Table 15 Diabetes data after annotation

DATETIME	VALUE	FHIR.Observation.subject.reference	FHIR.Observation.coding.code	FHIR.Observation.Unit
23/10/2018 08:00	354	1	2339-0	mg/dL
23/10/2018 18:00	275	1	2339-0	mg/dL

The next step of the process is the re-examination of the mappings in order to transform each attribute into FHIR. This is achieved by re-examining the mappings to determine how attributes returned map to FHIR, and application of any transformations embedded within the mapping nodes. Any attributes which contain no mappings are disregarded. The MIMIC data and diabetes data after this re-mapping are shown in Table 16 and

Table 17 respectively.

Table 16 MIMIC data after re-mapping

FHIR.Observation.subject.reference	FHIR.Observation.issued	FHIR.Observation.value	FHIR.Observation.Unit	FHIR.Observation.coding.code
1	20/10/2018 20:04	265	mg/dL	2339-0
1	20/10/2018 21:51	267	mg/dL	2339-0
1	20/10/2018 00:42	299	mg/dL	2339-0
1	20/10/2018 01:46	294	mg/dL	2339-0

Table 17 Diabetes data after re-mapping

FHIR.Observation.issued	FHIR.Observation.value	FHIR.Observation.subject.reference	FHIR.Observation.coding.code	FHIR.Observation.Unit
23/10/2018 08:00	354	1	2339-0	mg/dL
23/10/2018 18:00	275	1	2339-0	mg/dL

Finally, the two datasets require integration. A previous work (Scriney *et al.*, 2019) proposes a methodology for the determination of an integration strategy by examining the common datasets for each source in order to design a common data model, although for this study the common data model is the FHIR JSON schema. Using this methodology, the row-append method is selected to produce the unified data mart shown in Table 18.

Table 18 FHIR Observation response of diabetes data

FHIR.Observation.subject.reference	FHIR.Observation.issued	FHIR.Observation.value	FHIR.Observation.Unit	FHIR.Observation.coding.code
1	20/10/2018 20:04	265	mg/dL	2339-0
1	20/10/2018 21:51	267	mg/dL	2339-0
1	20/10/2018 00:42	299	mg/dL	2339-0
1	20/10/2018 01:46	294	mg/dL	2339-0
1	23/10/2018 08:00	354	mg/dL	2339-0
1	23/10/2018 18:00	275	mg/dL	2339-0

6.3 The Applications of the Semantic Engine

The main functionalities and applications of the semantic engine can be summarised in the following four points:

1. Semantic reasoning

Underpinned by Neo4j, the concepts or the resources defined in FHIR are explained through the connected properties of nodes and their relationships. This schematised FHIR data naturally develops the capacity for semantic reasoning between clinical concepts.

The steps for general semantic reasoning are summarised as follows.

Data acquisition:

A query enters the system in FHIR format. From this query, a list of entities is obtained, represented as **e** and their properties as $\mathbf{e_p}$ which are required to deliver the query.

Data source (**src**) which can satisfy this query are discovered through a traversal of the knowledge graph identifying mapping nodes (**m**) which link to these properties:

$$\boldsymbol{src} = \exists \boldsymbol{m} \forall \boldsymbol{e_p} \in \boldsymbol{e}$$

The following query (Command box 7) identifies any mappings for a given property ($\mathbf{e_p}$) and returns the data source (**src**), the relevant entity (**tab**) and the property required (**prop**).

Command box 7: Query the relationships of a given property

```
MATCH (src:Source)-(tab:Table)-(prop:Property)-(map:MAP)-
(eprop:Property) WHERE eprop = ep RETURN src, tab, prop, eprop;
```

For each mapping the data source (**src**) is queried in order to return the defined properties (**prop**). The data obtained from the source systems are then converted back into FHIR format (as specified in Section 6.2.2).

2. Patient-centric data organisation

The organisation of data into a patient-centric approach is the premise of the achievement of patient-centric care. From the perspective of the stakeholders of the healthcare ecosystem ranging from clinicians to carers, legal practitioners, and taxpayers, a wide range of individuals seek to obtain a holistic view of every individual patient's case. The benefits and the challenges of this have been addressed by academics from many different fields (Pelzang, 2010), with the organisation of patient data used in a patient-centric approach representing the first step towards the elimination of the silos between the health systems.

To facilitate semantic interoperability, the Semantic Engine organises the circulation of clinical data relating to patients and reflects the logic of diagnosis and treatment (as shown in Figure 50). Therefore, in addition to the provision of holistic patient health information which can be presented via the Semantic Engine, the patient him/herself can be empowered to authorise which data can be accessed and used by which organisations and agencies. This function can be achieved by the use of an extra module of data authority management, which is not discuss in detail in this paper. Moreover, patients may not be aware of the consequences of their own choices, which is an issue worthy of further exploration from the perspective of healthcare management.

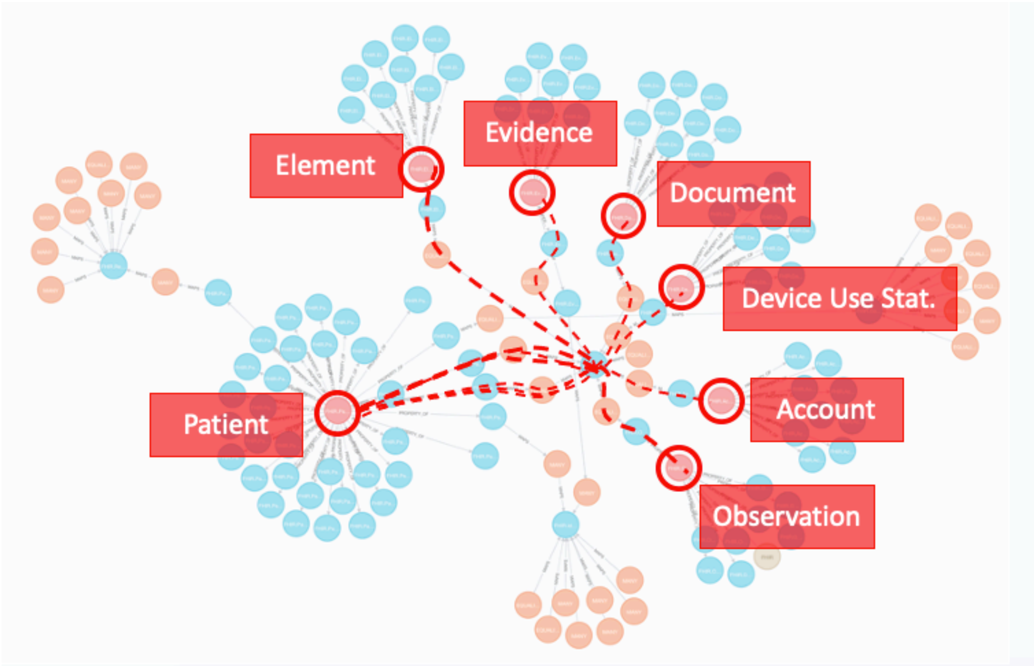


Figure 50 Patient-centric data organisation

3. Enhancing semantic interoperability

In addition to the definition and interpretation of medical terms, the Semantic Engine can retrieve data from disparate local databases to further clarify the meaning of definitions by provision of examples. Through this ostensive approach, the ambiguity caused by lexical definition can be minimised.

In the process of designing the verification scenarios, this study identifies a problem with unclear data sources, which potentially poses challenges for subsequent data analysis processes. The same FHIR resource, observation, has been used to interpret data collected from patient-worn monitoring equipment and clinical equipment in hospital settings. On consideration of the level of data reliability to support patient-centred diagnosis, it is clear that patient-worn monitoring equipment is less reliable than clinical equipment used in a clinical setting. Therefore, in practice, physicians should carefully review the laboratory reports and only use the data provided by the monitoring equipment for reference. In order to provide a firm foundation for an information-assisted clinical diagnosis system, the limitations of this study and suggestions for future research are discussed.

4. Applicability in other fields

This proposed information architecture processes semantics and data separately to avoid privacy and security issues arising from centralised data storage, while support information can be exchanged across heterogenous databases. This architecture can be applied in other domains which require information exchange and communication between dispersed systems. The construction of a consensus knowledge graph is the premise of the application of this semantics-data separated architecture.

6.4 Discussion and Summary

Regarding FHIR compliance issues, the current solutions (discussed in Section 4.4) have some limitations on implementation cost or effectiveness. Interestingly, during the thesis writing up, FHIR released the latest version (FHIR v4B released on 28th May 2022), which discusses the issue of conformality. In the new release, FHIR v4.3.0 introduces a conformance layer (HL7 International, 2022) to mitigate the interoperability problem caused by the inconsistent use of FHIR specifications by different applications, which aims to the third type of semantic ambiguity discussed in Section 4.3.1. The conformance layer is a statement provided by implementers about how the *resources* and their exchange paradigms are used to solve particular use cases, comprising a value set, a structure definition, a capability statement, and an implementation guide. The conformance layer is similar to the *extension* publishing management, which can improve the FHIR conformality, but challenges nevertheless remain.

Different from the existing solutions that ensure FHIR consistency through working processes or computer tools, this study targets large-scale health information systems and reconciles manual workloads with computer-automated workloads, avoiding high implementation costs and effectively addressing inconsistency problems. This research broadens the scope of the application of FHIR in healthcare ecosystems. The data from heterogeneous sources, such as smart devices, can be interchanged with clinical data via the Semantic Engine.

The proposed ostensive architecture is validated by the prototype of the Semantic Engine to enable data exchange and improve semantic interoperability. The work has been partially tested in a project supported by the Government of the Republic of Ireland in 2021, which involved multiple data sources for COVID-19 data analytics; a relational database was constructed to interpret semantics, and acts as the Semantic Engine.

In summary, in order to enhance the semantic interoperability of FHIR, and also to consider the data privacy issues and regulatory requirements for data sharing, an ostensive information architecture is proposed, which separates semantic processing and clinical data storage. There is a deliberate separation of semantics schema and underlying data, with the aim of improving flexibility and scalability. The centralised FHIR knowledge has the capacity to reduce the cost of the application of FHIR to multiple disparate clinical systems and is also flexible in its evolution. This study summarises the benefits of the semantics-data separated architecture into three principal points, as follows:

1. The centralised deployment of FHIR can reduce the costs incurred by its separate deployment in individual local systems, alleviating the impact of its evolution. The proposed ostensive information system is a federated architecture where queries are first to run on their respective sources and the data returned is mapped and integrated using the Semantic Engine returning a unified view of the data. As we do not envisage incremental updates to the Semantic Engine, it is possible to alleviate potential time costs within the integration and mapping steps by hosting multiple instances of the FHIR knowledge graph within the cloud.

2. This architecture acquires horizontal scalability through the maintenance of the distributed storage of clinical data and deployment of the centralised FHIR knowledge graph layer in the cloud cluster. This architecture supports vertical scalability in terms of handling the complex semantic reasoning and the evolution of FHIR.

3. The abstract semantic layer provides patients with the capacity to gain a complete view of their healthcare from dispersed data sources, enabling them to precisely decide the degree and extent of information exposure by managing the access permissions which can be embedded in the Semantic Engine. The Semantic Engine

executes the role-based accessed management tasks without exposing the FHIR knowledge graph to patients.

Chapter 7 Discussion

This chapter responds to the four research questions by reviewing the study and discussing its contributions. Section 7.1 discusses the findings and results in response to the study's research questions. The contributions from theoretical, methodological, and practical perspectives are discussed in Section 7.2 .

7.1 Responses to the Four Research Question

The ever-increasing complexity of the medical ecosystem and the rapidly growing demand for health data usage have placed urgent capacity requirements on the health information system architecture. One of the most vital requirements is to increase interoperability between heterogeneous information systems in order to create healthcare data accessibility for all stakeholders and to improve the data sharing quality for healthcare services in healthcare ecosystems. System interoperability has a broad reach, including empiric, syntactic, semantic, and pragmatic aspects. This research focus on the semantic interoperability and formulates four research questions: (1) what is the current state of health information systems and the interoperability challenges they face? (2) how to improve the semantic interoperability regarding to the multidisciplinary and cross-organisational healthcare delivery? (3) how to enhance semantic interoperability of FHIR? (4) How can FHIR-underpinned healthcare information platform integrate data from heterogeneous local systems with a unified schema for multiple purposes? The findings and outcomes of this research now readily answer this research questions.

To be able to successfully carry out this research, one must first have an understanding of the existing health information systems (research question 1) in terms of what factors make the health ecosystem so complex, what types of data are collected and made available, what are the expected capacities the health ecosystem should support, and what interoperability challenges exist. **Section 2.1** examines the complexity of the health ecosystem from the structure of technology network and further to explore the data availability for quality-of-care services. **Section 2.2** investigates the goals of HIS from stakeholders' perspective and the difficulties HIS faces from technological and social aspects. **Section 2.3** identifies the cost of interoperability difficulties from a financial perspective; the annual cost of the lack of medical device interoperability is $35 billion. In addition, the technical and operational obstacles of transforming traditional healthcare into patient-centre healthcare.

Section 2.4 describes semantic interoperability before investigating its solution (research question 2). To enable semantic interoperability, agents, services, and applications must share the same mutually agreed vocabulary or create correspondences/mappings between their different vocabularies. International standard organisations have released several standards for the healthcare domain. **Section 2.4** reviews these existing global standards through the lens of semiotics theory and classifies them according to their interoperability levels. OpenEHR, FHIR, HL7 CDA, and HL7 v3 fall into the category of semantic interoperability. As FHIR is the evolved version of H7 CDA and HL7 v3, a comparison between OpenEHR and FHIR has been made. FHIR offers a considerable advantage over Open HER in terms of technology architecture, openness, scalability, flexibility, portability, as well as adoption rate. Therefore, FHIR is a viable choice for enabling semantic interoperability in interdisciplinary and cross-organisational healthcare delivery.

The next step is to improve semantic interoperability of FHIR, such that **Chapter 4** reviews four interoperability paradigms to comprehend the capacity of FHIR to enable information exchange between heterogeneous systems. FHIR actively embraces Internet technologies and delivers benefits such as agility, fast iteration, and low learning costs; as a result, FHIR archives widespread industry adoption. However, semantic ambiguity that occurred in the FHIR implementation has garnered more attention and has been addressed in a number of studies. After exploring the semantic ambiguity introduced by FHIR profiling and the inherent features of FHIR, **Chapter 5** investigates the fundamental source of these semantic ambiguities and identifies the use of signs to represent objects as the cause. Peirce's triadic model illustrates the interpretant's role in the semantics between signs and objects.

Person-to-person variation in sign interpretation may result from diverse contexts and purposes of using signs. Consequently, during FHIR implementation, different implementers may have divergent interpretations of FHIR definition, resulting in the usage of the same FHIR resource to explain distinct clinical data. An ostensive approach, specifying an example to further elucidate the semantics, is particularly useful when complex signs are involved in reducing such contrasting sense-making. On the basis of the effectiveness of the ostensive approach, **Chapter 6** provides an ostensive information architecture to enhance semantic interoperability by introducing an ostensive approach during FHIR-based information exchange to answer the third research question.

Chapter 6 builds the Semantic Engine with a federated architecture in order to implement FHIR in a large-scale healthcare ecosystem and to integrate data from heterogeneous local systems for supporting multiple objectives (research question 4). The Semantic Engine adopts the MSA design philosophy to satisfy the need for dynamic semantic queries from diverse agents with various purposes of using healthcare data. In a federated architecture, heterogeneous and autonomous dispersed information systems collaborate with the Semantic Engine to communicate with one another. **Chapter 7** provides a study case to integrate MIMIC III data and diabetes data in the format of the FHIR re*source* to respond to the patient-centred data query, which demonstrates the FHIR-based Semantic Engine can integrate data from heterogeneous local systems with a unified schema for multiple purposes.

7.2 Research Contributions

This research explores the information architecture from semiotic perspective and recognises the reason of semantic ambiguity by the theory of semiotics. The contributions of this research can be judged from developing research methodology, applying an existing theory in a new domain of research, and bringing insights for practice. The following sections summaries the contributions from the theoretical, methodological and practical perspectives.

7.2.1 Theoretical Contributions

This research makes theoretical contributions to the field of information system architecture, which applies the theory of semiotics in the field of information architecture design.

In order to increase the interoperability of health information systems, international organisations have released many standards, some of which claim to enhance semantic interactivity. However, most standards target syntactic rather than semantic interoperability from the information systems point of view. This research reviews the existing standards and categorises them from low to high interoperability by labelling empirics, syntax, semantics, and pragmatics. This research further points out that only OpenEHR and FHIR provide semantic interoperability from the information exchange perspective.

In addition, regarding the most adopted international standard, this research explores the FHIR in implementation and identifies the issues of FHIR compliance. After reviewing current solutions for FHIR compliance, this research points out their limitations in terms of the practical scope and implementation cost. This thesis provides another research perspective to scrutinise the underlying causes of FHIR conformity. Semiotics, a study of sign processes (semiosis) and meaning making (Liu and Li, 2015), is adopted in this research to understand the communication between user and system, and identifies the possible semantic ambiguity caused by the lexical approach. Then this research proposes an ostensive approach as a supplementary method to reduce semantic ambiguity.

Lastly, in this research, an ostensive information architecture has been proposed based on the federation architecture. The proposed architecture can be regarded as a collaboration between MSA-inspired Semantic Engine and heterogeneous and autonomous local health information systems to achieve information sharing with a unified format defined by FHIR. The ostensive information architecture reflects the essence of information exchange by combining the lexical- and ostensive approaches to illustrate the semantics of the information exchange process. Specifically, in this research, the Semantic Engine elaborates the lexical definitions of concepts; local data are requested as examples to decrease the semantic ambiguity led by lexical definitions. This combination of lexical- and ostensive approaches alleviates the semantic ambiguity that occurred in the information exchange.

In theory, the proposed ostensive architecture provides a new perspective to understanding FHIR implementing issues and provides a cost-effective and implementable architecture for large-scale healthcare ecosystems.

7.2.2 Methodological Contributions

This research contributes methodologically by asserting the falsification research process (Magee, 1973) can be used as a logical reasoning approach in developing information architecture for digital healthcare ecosystems. The four phrases of falsification reasoning (see Section 3.4.2) encourages researchers to break through the limitations of existing theories.

Following the falsification research process, this research starts by addressing the limitations of the current solution for semantic exchange between healthcare information systems (see Chapter 4). Regarding the identified limitations, this research proposes a new solution to enhance the semantic interoperability of FHIR (see Chapter 6) on the basis of investigating the causes of semantic ambiguity (see Chapter 5). The proposed architecture is examined to prove the capacity to decrease semantic ambiguity and facilitate the exchange between dispersed local information systems. The process of falsification research has significant implications to extend the boundaries of existing theories/frameworks/specifications. In addition to the field of information architecture design in this research, the falsification research process is also helpful for scientific research in other fields. Based on existing theories, the latest research can focus more on the flaws or limitations in order to extend the established boundaries and better explain the observed world.

The realisation of system architecture design ideas by transforming software tools is another contribution of this study to research methodology. In the design of the Semantic Engine, Microservice philosophy from information systems has been applied to construct the FHIR ontology with the format of node-edge. Furthermore, drawing on the idea of a successful large-scale system, federated architecture has been used in the Semantic Engine design to gain the processing capability for a large volume of semantic queries. Moreover, instead of using common ontology languages, such as OWL and RDF, this research adopts Neo4j to implement the FHIR knowledge graph, which further leverage the benefits of existing tools in another domain. The built-in functions of Neo4j make the realisation of semantic reasoning more convenient; the developer communities of Neo4j provide flexible implementations of the FHIR knowledge graph.

7.2.3 Practical Contributions

This research has significant practical implications as the research question arises from the FHIR implementation and the research objective is to resolve these issues.

With the increasing adoption of FHIR in the industry, primarily when FHIR is used as a unified terminology ontology across systems, FHIR compliance has become an increasingly prominent issue in practice. The ostensive information architecture exhibits the correspondents between FHIR resources and local healthcare data attributes to further clarify the implementors' understandings of FHIR resources, thus harmonising the FHIR understanding across information systems. The proposed information architecture is easy to implement and has the capacity to handle large-scale systems with low implementing costs. In this research, MIMIC III datasets are employed to verify the effectiveness of the ostensive information architecture and a good demonstration effect on the use of other databases (see Chapter 7). MIMIC III datasets are de-identified health-related data in critical care units of the Beth Israel Deaconess Medical Centre between 2001 and 2012.

Another practical contribution posited in this research is the "semantics and data separation" design. The proposed ostensive architecture separates semantic explanation and reasoning from the healthcare data and echoes the requirements of data privacy and security management. This design avoids the legal and ethical issues (Ashok *et al.*, 2022) associated with sharing healthcare data across systems in practice. Individual healthcare systems can share data with others through a federal architecture while remaining autonomous, which is what these healthcare organisations need.

Furthermore, this research describes constructing a Semantic Engine in detail, providing implementors with an instruction guide. Particularly in Chapter 7, the integration of MIMIC III and diabetes datasets demonstrates how heterogeneous information systems collaborate through the Semantic Engine. These detailed descriptions bring implications to FHIR implementation and potentially drive the FHIR adoption rate.

Lastly, this research deeply considers the potential difficulties of FHIR implementation in practice and the negative impact of its implementation costs on adopting FHIR. The ostensive information architecture follows the design idea of the federated architecture, which does not just consider the autonomy needs of each organisation, but also considers how to integrate with heterogeneously incumbent databases at low cost.

In summary, the most important aspect of the proposed ostensive architecture lies in that the new architecture inherits the essence of FHIR's design and alleviates the limitation of the semantic ambiguity that occurred in the FHIR implementation.

Chapter 8 Conclusion

This chapter's focus is on drawing conclusions., as contrast to the previous chapter's emphasis on discussions. This chapter also examines the study's limitations and their implications for future research.

8.1 New Knowledge Gained from this Research

The primary acquired knowledge is the ostensive approach derived from this research. The ostensive information architecture improves the semantic interoperability of FHIR by providing a comprehensive understanding of the information interaction between users and systems. The separation of semantic and data can increase the flexibility of FHIR's evolution as well as protect the confidentiality of patient data. In addition, the ostensive information architecture is easy to implement and can improve FHIR compliance, hence facilitating a greater rate of adoption. This research demonstrates the effectiveness of ostensive approach in IS research.

In addition, the significance of falsificationism in the field of information systems is confirmed by this study, which is another contribution to the body of knowledge. Due to the fact that an information system is a medium that reflects and varies in accordance with social and cultural diversity. Consequently, there is no universally applicable solution in the design of information systems. Continuous development based on the best solution available could be a more feasible strategy for IS continuous improvement. This study demonstrates the efficacy of falsificationism for continual improvement in the realm of IS research.

8.2 The Limitations of the Study

This research adopts FHIR as domain knowledge to construct a Semantic Engine for the interpretation of the meanings of clinical concepts. It is observed that FHIR does not distinguish between the different levels of data reliability. When this ostensive information architecture is brought into use, consideration should be given to the level of data reliability and the conflicts caused by of multiple data sources being used for the same indicator. This problem can potentially be solved by specification of the primary database, although due to the limited availability of medical data, this study does not provide an in-depth discussion of this issue.

This research focuses on semantic interoperability but does not explore the relationship between semantics and operational process, such as patient pathways or clinical procedures. The context of the data is an extremely important factor concerning its semantics, and this may vary in the context of different processes, which is not explored in depth by this study.

The ostensive information architecture proposed by this study is applicable to the entire medical ecosystem, therefore, it is evident that there is a serious problem of record linkage, specifically in terms of detecting, identifying, matching, and merging records across heterogeneous databases which related to the same patient (Reyes-Galaviz *et al.*, 2017). For example, two systems may refer to the same patient but use different identity codes to correctly identify a patient across systems. Manual data entry compounds this problem when spelling errors and formatting inconsistencies limit an automatic means of identification of patient records across systems. To overcome these issues, this study proposes a model of **record linker** in the transformation layer; this is similar to the method proposed by Nie and Roantree (Nie and Roantree, 2019) which seeks to produce a probabilistic means of identifying patients during the re-mapping process. The difficulty for this research study of obtaining data on patients' profiles from multiple systems precludes the conduct of a case study to demonstrate how the record linker works.

8.3 Implications for Future Research

This proposal of the ostensive information architecture simply represents the first step towards the achievement of patient-centric diagnosis. The following are directions worthy of further exploration:

- In the current Semantic Engine, the properties for edges are limited, which indicates the affiliation between nodes. In further research, richer semantic properties could be added to the edges. For example, data from a wearable device could be given low priority if there is data for the same measurement from a medical device. The richer semantic properties can support the Semantic Engine in the construction of a diagnostic graph which has the capacity to reason and prioritise the level of data reliability according to its sources.
- Breakthroughs in the medical field and the discovery of new diseases mean that the definitions of clinical concepts are in constant evolution; this underlines the fact that a gap between the Semantic Engine and the data examples is likely to persist. Therefore, the function of tracing and managing the changes of FHIR resources becomes essential to ensure the rigour in the mapping of relationships. Blockchain technology offers an optional solution to this challenge; it can be used to record the evolutionary history of FHIR whilst also tracking the changes in patients' medical history records (Zhang *et al.*, 2018). The use of blockchain technology in healthcare information systems has many potential application scenarios and is of high practical value (Mettler, 2016). For example, blockchain can be used to provide access to medical data (Azaria *et al.*, 2016) and privacy control (Yue *et al.*, 2016). Overall, the proposed ostensive information architecture provides a foundation for HISs; additional research work, including organisation, patient pathway and clinical processes mapping to the Semantic Engine, is required prior to the completion of a comprehensive HIS proposal.

8.4 Reflection of Research

This is my first experience conducting scientific research. This extensive path towards scientific discoveries has taught me the rigours of scientific investigation. Beginning with a comprehensive literature review and gradually reducing the area of the research, I was able to identify a research gap. My supervisor exhibits considerable patience in permitting me to explore different academic fields in the interim. While my efforts and time were not wasted, they did expand my knowledge my understanding in the vertical and horizontal dimensions of my studies. Thank you to my supervisor for assigning me this 'difficult' literature review process, which has transformed me into a thorough scientist.

The main benefit of this scientific research is that I now comprehend the distinction between a software project and scientific research in the field of software engineering. My supervisor's theoretical guidance from a lofty vantage point has illuminated the direction of my scientific research in the subject of information systems. Only solid research logic and theoretical support can confer academic merit on software engineering advancements. I now recognise the importance of multidisciplinary studies and the significance of scientific theories.

Research data accessibility is the most significant feature of academic research. Due of the difficulty of obtaining data for scientific research, I altered the industry of my investigation. This was made possible by the unselfish sharing of the scientific community and the industry's support for scientific research. I am fortunate to have gotten MIMIC III datasets managed by the MIT Laboratory for Computational Physiology, which enabled my research to proceed smoothly.

This research experience to me is priceless to me. It changed my perspective and prepared me to go from an industry expert to a scientific researcher.

References

Agbo, C. C.; Mahmoud, Q. H. & Eklund, J. M. Blockchain Technology in Healthcare: A Systematic Review. Healthcare, 2019. Multidisciplinary Digital Publishing Institute, 56.

Ahearn, L. M. 2021. *Living language: An introduction to linguistic anthropology*, John Wiley & Sons.

Ahmadi, H.; Arji, G.; Shahmoradi, L.; Safdari, R.; Nilashi, M. & Alizadeh, M. 2019. The application of internet of things in healthcare: a systematic literature review and classification. *Universal Access in the Information Society,* 18**,** 837-869.

Al-Turjman, F.; Nawaz, M. H. & Ulusar, U. D. 2020. Intelligence in the Internet of Medical Things era: A systematic review of current and future trends. *Computer Communications,* 150**,** 644-660.

ALCIDION. 2022 *Interoperability with Open Standards* [Online]. Available: https://www.alcidion.com/news/interoperability-with-open-standards/ [Accessed].

Alemdar, H. & Ersoy, C. 2010. Wireless sensor networks for healthcare: A survey. *Computer networks,* 54**,** 2688-2710.

Aloini, D.; Benevento, E.; Stefanini, A. & Zerbino, P. 2022. Transforming healthcare ecosystems through blockchain: Opportunities and capabilities for business process innovation. *Technovation***,** 102557.

Alterovitz, G.; Warner, J.; Zhang, P.; Chen, Y.; Ullman-Cullere, M.; Kreda, D. & Kohane, I. S. 2015. SMART on FHIR Genomics: facilitating standardized clinico-genomic apps. *Journal of the American Medical Informatics Association,* 22**,** 1173-1178.

Andreewsky, E. & Bourcier, D. 2000. Abduction in language interpretation and law making. *Kybernetes*.

Arsanjani, A. 2004. Service-oriented modeling and architecture. *IBM developer works,* 1**,** 15.

Ashok, M.; Madan, R.; Joha, A. & Sivarajah, U. 2022. Ethical framework for Artificial Intelligence and Digital technologies. *International Journal of Information Management,* 62**,** 102433.

Ashrafi, N.; Kuilboer, J.-P. & Stull, T. Semantic Interoperability in Healthcare: Challenges and Roadblocks. STPIS@ CAiSE, 2018. 119-122.

Atalag, K.; Beale, T.; Chen, R.; Gornik, T.; Heard, S. & McNicoll, I. 2016. openEHR—a semantically enabled, vendor-independent health computing platform. *URL: http://www. openehr. org/resources/white_paper_docs/openEHR_vendor_independent_platform. pdf.*

Aungst, T. D. & Patel, R. 2020. Integrating digital health into the curriculum—considerations on the current landscape and future developments. *Journal of medical education and curricular development,* 7**,** 2382120519901275.

Azaria, A.; Ekblaw, A.; Vieira, T. & Lippman, A. Medrec: Using blockchain for medical data access and permission management. 2016 2nd international conference on open and big data (OBD), 2016. IEEE, 25-30.

Bailey, N. 1969. 1721. An universal etymological English dictionary. *London: Bell*.

Baker, A. 2001. Crossing the quality chasm: a new health system for the 21st century. British Medical Journal Publishing Group.

Bassham, G. 2011. *Critical thinking: A student's introduction*.

Baxter, J. S.; Gibson, E.; Eagleson, R. & Peters, T. M. 2018. The semiotics of medical image segmentation. *Medical image analysis,* 44**,** 54-71.

Beale, T. Archetypes: Constraint-based domain models for future-proof information systems. OOPSLA 2002 workshop on behavioural semantics, 2002.

Beale, T. 2019. *A FHIR experience: consistently inconsistent* [Online]. Available: https://wolandscat.net/2019/05/05/a-fhir-experience-consistently-inconsistent/ [Accessed].

Bender, D. & Sartipi, K. HL7 FHIR: An Agile and RESTful approach to healthcare information exchange. Proceedings of the 26th IEEE international symposium on computer-based medical systems, 2013. IEEE, 326-331.

Benson, T. & Grieve, G. 2016. *Principles of health interoperability: SNOMED CT, HL7 and FHIR*, Springer.

Beredimas, N.; Kilintzis, V.; Chouvarda, I. & Maglaveras, N. A reusable ontology for primitive and complex HL7 FHIR data types. 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015. IEEE, 2547-2550.

Berger, P. L. 1967. *The social construction of reality*, Anchor books.

Blobel, B. 2011. Ontology driven health information systems architectures enable pHealth for empowered patients. *International journal of medical informatics,* 80**,** e17-e25.

Blobel, B. 2019. Challenges and solutions for designing and managing pHealth ecosystems. *Frontiers in medicine,* 6**,** 83.

Blobel, B.; Engel, K. & Pharowe, P. 2006. Semantic interoperability. *Methods of information in medicine,* 45**,** 343-353.

Bloomfield Jr, R. A.; Polo-Wood, F.; Mandel, J. C. & Mandl, K. D. 2017. Opening the Duke electronic health record to apps: implementing SMART on FHIR. *International journal of medical informatics,* 99**,** 1-10.

Blumenthal, D. & Tavenner, M. 2010. The "meaningful use" regulation for electronic health records. *New England Journal of Medicine,* 363**,** 501-504.

Blumer, H. 1986. *Symbolic interactionism: Perspective and method*, Univ of California Press.

Bodenreider, O. Issues in mapping LOINC laboratory tests to SNOMED CT. AMIA Annual Symposium Proceedings, 2008. American Medical Informatics Association, 51.

Boland Jr, R. J. 1979. Control, causality and information system requirements. *Accounting, Organizations and Society,* 4**,** 259-272.

Boland Jr, R. J. 1985. 10 phenomenology: a preferred approach to research in information systems. *Research Methods in Information Systems, Amsterdam: North-Holland***,** 181-90.

Braa, K. & Vidgen, R. 1999. Interpretation, intervention, and reduction in the organizational laboratory: a framework for in-context information system research. *Accounting, Management and Information Technologies,* 9**,** 25-47.

Braunstein, M. L. 2018. *Health Informatics on FHIR: How HL7's New API is Transforming Healthcare*, Springer.

Brödner, P. 2019. Coping with Descartes' error in information systems. *AI & SOCIETY,* 34**,** 203-213.

Brown, A. W.; Delbaere, M.; Eeles, P.; Johnston, S. & Weaver, R. 2005. Realizing service-oriented solutions with the IBM rational software development platform. *IBM systems journal,* 44**,** 727-752.

Burrell, G. & Morgan, G. 2017. *Sociological paradigms and organisational analysis: Elements of the sociology of corporate life*, Routledge.

Buttery, E. A. & Buttery, E. M. 1991. Design of a marketing information system: useful paradigms. *European Journal of marketing,* 25**,** 26-39.

Cerny, T.; Donahoo, M. J. & Pechanec, J. Disambiguation and comparison of soa, microservices and self-contained systems. Proceedings of the International Conference on Research in Adaptive and Convergent Systems, 2017. 228-235.

Chamola, V.; Hassija, V.; Gupta, V. & Guizani, M. 2020. A comprehensive review of the COVID-19 pandemic and the role of IoT, drones, AI, blockchain, and 5G in managing its impact. *Ieee access,* 8**,** 90225-90265.

Chandler, D. 1994. Semiotics for beginners. Aberystwyth University.

Chandler, D. 2017. *Semiotics: the basics*, Taylor & Francis.

Chartier, J.-F.; Pulizzotto, D.; Chartrand, L. & Meunier, J.-G. 2019. A data-driven computational semiotics: The semantic vector space of Magritte's artworks. *Semiotica,* 2019**,** 19-69.

Chartier, M. R. 1981. Clarity of expression in interpersonal communication. *The Journal of nursing administration,* 11**,** 42-46.

Chaudhry, B.; Wang, J.; Wu, S.; Maglione, M.; Mojica, W.; Roth, E.; Morton, S. C. & Shekelle, P. G. 2006. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Annals of internal medicine,* 144**,** 742-752.

Chen, M.; Yang, J.; Hao, Y.; Mao, S. & Hwang, K. 2017. A 5G cognitive system for healthcare. *Big Data and Cognitive Computing,* 1**,** 2.

Chen, W. & Hirschheim, R. 2004. A paradigmatic and methodological examination of information systems research from 1991 to 2001. *Information systems journal,* 14**,** 197-235.

Chettri, L. & Bera, R. 2019. A comprehensive survey on Internet of Things (IoT) toward 5G wireless systems. *IEEE Internet of Things Journal,* 7**,** 16-32.

Choi, N.; Song, I.-Y. & Han, H. 2006. A survey on ontology mapping. *ACM Sigmod Record,* 35**,** 34-41.

Christensen, B. & Ellingsen, G. 2016. Evaluating model-driven development for large-scale EHRs through the openEHR approach. *International journal of medical informatics,* 89**,** 43-54.

Curran, R. L.; Kukhareva, P. V.; Taft, T.; Weir, C. R.; Reese, T. J.; Nanjo, C.; Rodriguez-Loya, S.; Martin, D. K.; Warner, P. B. & Shields, D. E. 2020. Integrated displays to improve chronic disease management in ambulatory care: A SMART on FHIR application informed by mixed-methods user testing. *Journal of the American Medical Informatics Association,* 27**,** 1225-1234.

Currie, W. L. & Finnegan, D. J. 2011. The policy-practice nexus of electronic health records adoption in the UK NHS: an institutional analysis. *Journal of enterprise information management*.

Danermark, B.; Ekström, M. & Karlsson, J. C. 2019. *Explaining society: Critical realism in the social sciences*, Routledge.

Daraio, C.; Lenzerini, M.; Leporelli, C.; Naggar, P.; Bonaccorsi, A. & Bartolucci, A. 2016. The advantages of an Ontology-Based Data Management approach: openness, interoperability and data quality. *Scientometrics,* 108**,** 441-455.

Data Bridge Market Research. 2022. *Internet of Medical Things (IoMT) Market Industry Trends, Share, Size, Demand, Application, Growth and Future Scope* [Online]. Available: https://www.designerwomen.co.uk/internet-of-medical-things-iomt-market-industry-trends-share-size-demand-application-growth-and-future-scope/ [Accessed].

De Saussure, F. 2011. *Course in general linguistics*, Columbia University Press.

Dedić, N. 2021. EAFP: Enterprise Architecture Fusion Process. *Journal of Information and Organizational Sciences,* 45**,** 223-241.

Denzin, N. K. & Lincoln, Y. S. 1995. Transforming qualitative research methods: Is it a revolution? *Journal of Contemporary Ethnography,* 24**,** 349-358.

Deppenwiese, N.; Delpy, P.; Lambarki, M. & Lablans, M. 2021. ADT2FHIR–A Tool for Converting ADT/GEKID Oncology Data to HL7 FHIR Resources. *German Medical Data Sciences 2021: Digital Medicine: Recognize–Understand–Heal.* IOS Press.

Dhombres, F. & Bodenreider, O. 2016. Interoperability between phenotypes in research and healthcare terminologies—Investigating partial mappings between HPO and SNOMED CT. *Journal of biomedical semantics,* 7**,** 3.

Diaz, M. A. C.; Antonelli, L. & Sanchez, L. E. 2017. Health ontology and information systems: a systematic review. *IEEE Latin America Transactions,* 15**,** 103-120.

Dijkstra, E. W. 1974. On the role of scientific thought. *Online accessible under http://www. cs. utexas. edu/users/EWD/ewd04xx/EWD447. PDF***,** 12-04.

Dills, C. R. & Romiszowski, A. J. 1997. *Instructional development paradigms*, Educational Technology.

Dolin, R.; Boxwala, A. & Shalaby, J. 2018. A pharmacogenomics clinical decision support service based on FHIR and CDS Hooks. *Methods of Information in Medicine,* 57**,** e115-e123.

Dolin, R. H.; Alschuler, L.; Beebe, C.; Biron, P. V.; Boyer, S. L.; Essin, D.; Kimber, E.; Lincoln, T. & Mattison, J. E. 2001. The HL7 clinical document architecture. *Journal of the American Medical Informatics Association,* 8**,** 552-569.

Dolin, R. H.; Gothi, S. R.; Boxwala, A.; Heale, B. S.; Husami, A.; Jones, J.; Khangar, H.; Londhe, S.; Naeymi-Rad, F. & Rao, S. 2021. vcf2fhir: a utility to convert VCF files into HL7 FHIR format for genomics-EHR integration. *BMC bioinformatics,* 22**,** 1-11.

Dubois, A. & Gadde, L.-E. 2002. Systematic combining: an abductive approach to case research. *Journal of business research,* 55**,** 553-560.

Eco, U. 1979. *A theory of semiotics*, Indiana University Press.

Ehrlinger, L. & Wöß, W. 2016. Towards a Definition of Knowledge Graphs. *SEMANTiCS (Posters, Demos, SuCCESS),* 48**,** 2.

El-Sappagh, S.; Ali, F.; El-Masri, S.; Kim, K.; Ali, A. & Kwak, K.-S. 2018. Mobile health technologies for diabetes mellitus: Current state and future challenges. *IEEE Access,* 7**,** 21917-21947.

El-Sappagh, S.; Ali, F.; Hendawi, A.; Jang, J.-H. & Kwak, K.-S. 2019. A mobile health monitoring-and-treatment system based on integration of the SSN sensor ontology and the HL7 FHIR standard. *BMC medical informatics and decision making,* 19**,** 97.

Ellis, J. 1986. The superstring: theory of everything, or of nothing? *Nature,* 323**,** 595-598.

Erl, T. 2005. *Service-oriented architecture : concepts, technology, and design,* Upper Saddle River, NJ, Prentice Hall Professional Technical Reference.

Farhoomand, A. F. 1987. Scientific progress of management information systems. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems,* 18**,** 48-56.

Fielding, R. T. 2000. *Architectural styles and the design of network-based software architectures*, University of California, Irvine.

Firely. 2015. *The FHIR registry* [Online]. Available: https://simplifier.net/ [Accessed].

Firouzi, F.; Rahmani, A. M.; Mankodiya, K.; Badaroglu, M.; Merrett, G. V.; Wong, P. & Farahani, B. 2018. Internet-of-Things and big data for smarter healthcare: from device to architecture, applications and analytics. Elsevier.

Ganzha, M.; Paprzycki, M.; Pawłowski, W.; Szmeja, P. & Wasielewska, K. 2017. Semantic interoperability in the Internet of Things: An overview from the INTER-IoT perspective. *Journal of Network and Computer Applications,* 81**,** 111-124.

Garlan, D.; Allen, R. & Ockerbloom, J. 1995. Architectural mismatch: Why reuse is so hard. *IEEE software,* 12**,** 17-26.

Gauthier, R. L. & Ponto, S. D. 1970. *Designing Systems Programs*, Prentice-Hall.

Geertz, C. 2008. *Thick description: Toward an interpretive theory of culture*, Routledge.

Geraci, A. 1990. IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries. 1990. *Institute of Electrical and Electronics Engineers Inc., New York*.

Geraci, A.; Katki, F.; McMonegal, L.; Meyer, B.; Lane, J.; Wilson, P.; Radatz, J.; Yee, M.; Porteous, H. & Springsteel, F. 1991. *IEEE standard computer dictionary: Compilation of IEEE standard computer glossaries*, IEEE Press.

Gøeg, K. R. & Hummeluhr, M. An Empirical Approach to Enhancing Terminology Binding-An HL7 FHIR SNOMED CT Example. MIE, 2018. 206-210.

Goguen, J. A. & Burstall, R. M. 1992. Institutions: Abstract model theory for specification and programming. *Journal of the ACM (JACM),* 39**,** 95-146.

Goldkuhl, G. 2012. Pragmatism vs interpretivism in qualitative information systems research. *European journal of information systems,* 21**,** 135-146.

Goldschmidt, P. G. 2005. HIT and MIS: implications of health information technology and medical information systems. *Communications of the ACM,* 48**,** 68-74.

Gregor, S. 2006. The nature of theory in information systems. *MIS quarterly***,** 611-642.

Gregor, S.; Müller, O. & Seidel, S. 2013. Reflection, abstraction and theorizing in design and development research.

Grimson, J.; Grimson, W. & Hasselbring, W. 2000. The SI challenge in health care. *Communications of the ACM,* 43**,** 48-55.

Gruber, T. 2009. Ontology. *Encyclopedia of database systems,* 1**,** 1963-1965.

Gruendner, J.; Wolf, N.; Tögel, L.; Haller, F.; Prokosch, H.-U. & Christoph, J. 2020. Integrating genomics and clinical data for statistical analysis by using GEnome MINIng (GEMINI) and Fast Healthcare Interoperability Resources (FHIR): System design and implementation. *Journal of medical Internet research,* 22**,** e19879.

Gua, H. & Liu, K. 2019. Hierarchical ontology graph for solving semantic issues in decision support systems.

Guba, E. G. The paradigm dialog. Alternative paradigms conference, mar, 1989, indiana u, school of education, san francisco, ca, us, 1990. Sage Publications, Inc.

Gubbi, J.; Buyya, R.; Marusic, S. & Palaniswami, M. 2013. Internet of Things (IoT): A vision, architectural elements, and future directions. *Future generation computer systems,* 29**,** 1645-1660.

Guo, H., Scriney, M., Liu, K.,. 2023. *FHIR Knowledge Graph Generation (v1.0)* [Online]. Available: https://zenodo.org/record/7545834#.Y8hL9ezMLAM [Accessed].

Håkansson Eklund, J.; Holmström, I. K.; Kumlin, T.; Kaminsky, E.; Skoglund, K.; Höglander, J.; Sundler, A. J.; Condén, E. & Summer Merenius, M. 2019. Same same or different? A review of reviews of person-centred and patient-centred care. *Patient education and counseling***,** 3-11.

Halliday, M. A. K. 1973. Explorations in the functions of language.

Halliday, M. A. K. 2014. Language as social semiotic. *The Discourse Studies Reader. Amsterdam: John Benjamins***,** 263-272.

Haux, R. 2006. Health information systems–past, present, future. *International journal of medical informatics,* 75**,** 268-281.

He, J.; Baxter, S. L.; Xu, J.; Xu, J.; Zhou, X. & Zhang, K. 2019. The practical implementation of artificial intelligence technologies in medicine. *Nature medicine,* 25**,** 30-36.

Heeks, R. 2006. Health information systems: Failure, success and improvisation. *International journal of medical informatics,* 75**,** 125-137.

Heimbigner, D. & McLeod, D. 1985. A federated architecture for information management. *ACM Transactions on Information Systems (TOIS),* 3**,** 253-278.

Heller, N.; Isensee, F.; Maier-Hein, K. H.; Hou, X.; Xie, C.; Li, F.; Nan, Y.; Mu, G.; Lin, Z. & Han, M. 2021. The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: Results of the KiTS19 challenge. *Medical image analysis,* 67**,** 101821.

Henderson, K. A. 2011. Post-positivism and the pragmatics of leisure research. *Leisure Sciences,* 33**,** 341-346.

HIMSS. 2014. *What is Interoperability?* [Online]. Available: https://www.himss.org/what-interoperability [Accessed].

HIMSS. 2022. *Interoperability in Healthcare* [Online]. Available: https://www.himss.org/resources/interoperability-healthcare [Accessed].

Hirschheim, R.; Welke, R. & Schwarz, A. 2010. Service-oriented architecture: Myths, realities, and a maturity model. *MIS Quarterly Executive,* 9.

HL7. 2022a. *FHIR Code Systems* [Online]. Available: https://hl7.org/fhir/terminologies-systems.html [Accessed].

HL7. 2022b. *HL7 FHIR release 5 draft ballot* [Online]. Available: https://build.fhir.org/summary.html [Accessed].

HL7. 2022c. *HL7 Version 2* [Online]. Available: https://www.hl7.org.uk/standards/hl7-standards/hl7-version-2/ [Accessed].

HL7. 2022d. *Implementer's Safety Check List* [Online]. Available: https://www.hl7.org/fhir/safety.html [Accessed].

HL7 International. 1987. *Health level 7* [Online]. Available: http://www.hl7.org/ [Accessed].

HL7 International. 2011a. *FHIR Overview - Architects* [Online]. Available: https://www.hl7.org/fhir/overview-arch.html [Accessed].

HL7 International. 2011b. *RESTful API* [Online]. Available: https://www.hl7.org/fhir/http.html [Accessed].

HL7 International. 2019a. *impedance differences between FHIR and ontologies* [Online]. Available: https://www.hl7.org/fhir/rdf.html [Accessed].

HL7 International. 2019b. *Open Source FHIR implementations* [Online]. Available: https://wiki.hl7.org/Open_Source_FHIR_implementations [Accessed].

HL7 International. 2022. *FHIR Conformance* [Online]. Available: https://www.hl7.org/fhir/conformance-rules.html [Accessed].

Hoang, D. B. & Chen, L. Mobile cloud for assistive healthcare (MoCAsH). 2010 IEEE Asia-Pacific Services Computing Conference, 2010. IEEE, 325-332.

Holdcroft, D. 1991. *Saussure: signs, system and arbitrariness*, Cambridge University Press.

Holzinger, A.; Dehmer, M. & Jurisica, I. 2014. Knowledge discovery and interactive data mining in bioinformatics-state-of-the-art, future challenges and research directions. *BMC bioinformatics,* 15**,** 1-9.

Horrobin, D. F. 1969. The Nature of Scientific Research. *Science is God.* Dordrecht: Springer Netherlands.

Houser, N. 2020. Semiotics and philosophy. *The American Journal of Semiotics,* 36**,** 135-164.

Howe, K. R. 1988. Against the quantitative-qualitative incompatibility thesis or dogmas die hard. *Educational researcher,* 17**,** 10-16.

Hyde, K. F. 2000. Recognising deductive processes in qualitative research. *Qualitative market research: An international journal*.

Imhoff, M. 2002. Health informatics. *Evaluating Critical Care***,** 255-269.

Information Technology Industry Council. 2018. *Cloud Healthcare Pledge* [Online]. Available: https://www.itic.org/public-policy/CloudHealthcarePledge.pdf [Accessed].

International Organization for Standardization. 1994. *Information technology — Open Systems Interconnection — Basic Reference Model: The Basic Model* [Online]. Available: https://www.iso.org/standard/20269.html [Accessed].

INTEROP NoE. 2007. *Deliverable DI.3: Enterprise Interoperability Framework and Knowledge Corpus* [Online]. Available: http://interop-vlab.eu/interop/ [Accessed].

Iroju, O.; Soriyan, A.; Gambo, I. & Olaleke, J. 2013. Interoperability in healthcare: benefits, challenges and resolutions. *International Journal of Innovation and Applied Studies,* 3**,** 262-270.

ISO. *ISO 13606* [Online]. Available: http://www.en13606.org/ [Accessed].

Iyawa, G. E.; Herselman, M. & Botha, A. 2016. Digital health innovation ecosystems: From systematic literature review to conceptual framework. *Procedia Computer Science,* 100**,** 244-252.

Jabbar, S.; Ullah, F.; Khalid, S.; Khan, M. & Han, K. 2017. Semantic interoperability in heterogeneous IoT infrastructure for healthcare. *Wireless Communications and Mobile Computing,* 2017.

Janett, R. S. & Yeracaris, P. P. 2020. Electronic Medical Records in the American Health System: challenges and lessons learned. *Ciencia & saude coletiva,* 25**,** 1293-1304.

Jennett, P. A.; Hall, L. A.; Hailey, D.; Ohinmaa, A.; Anderson, C.; Thomas, R.; Young, B.; Lorenzetti, D. & Scott, R. E. 2003. The socio-economic impact of telehealth: a systematic review. *Journal of telemedicine and telecare,* 9**,** 311-320.

Jiang, F.; Jiang, Y.; Zhi, H.; Dong, Y.; Li, H.; Ma, S.; Wang, Y.; Dong, Q.; Shen, H. & Wang, Y. 2017a. Artificial intelligence in healthcare: past, present and future. *Stroke and vascular neurology,* 2.

Jiang, G.; Kiefer, R. C.; Rasmussen, L. V.; Solbrig, H. R.; Mo, H.; Pacheco, J. A.; Xu, J.; Montague, E.; Thompson, W. K. & Denny, J. C. 2016. Developing a data element repository to support EHR-driven phenotype algorithm authoring and execution. *Journal of biomedical informatics,* 62**,** 232-242.

Jiang, G.; Kiefer, R. C.; Sharma, D. K.; Prud'hommeaux, E. & Solbrig, H. R. 2017b. A consensus-based approach for harmonizing the OHDSI common data model with HL7 FHIR. *Studies in health technology and informatics,* 245**,** 887.

Jiang, G.; Solbrig, H. R.; Kiefer, R. C.; Rasmussen, L. V.; Mo, H.; Pacheco, J. A.; Montague, E. N.; Xu, J.; Speltz, P. & Thompson, W. K. Harmonization of Quality Data Model with HL7 FHIR to Support EHR-driven Phenotype Authoring and Execution: A Pilot Study. AMIA, 2015.

Johnson, R. B. & Christensen, L. 2019. *Educational research: Quantitative, qualitative, and mixed approaches*, Sage publications.

Joshi, R.; Didier, P.; Jimenez, J. & Carey, T. 2017. The industrial internet of things volume G5: connectivity framework. *Industrial Internet Consortium Report*.

Kankanhalli, A.; Hahn, J.; Tan, S. & Gao, G. 2016. Big data and analytics in healthcare: introduction to the special section. *Information Systems Frontiers,* 18**,** 233-235.

Kannampallil, T. G.; Schauer, G. F.; Cohen, T. & Patel, V. L. 2011. Considering complexity in healthcare systems. *Journal of biomedical informatics,* 44**,** 943-947.

Karimi, D.; Dou, H.; Warfield, S. K. & Gholipour, A. 2020. Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis. *Medical Image Analysis,* 65**,** 101759.

Kataria, P. & Juric, R. Sharing e-health information through ontological layering. 2010 43rd Hawaii International Conference on System Sciences, 2010. IEEE, 1-10.

Kaushik, H.; Singh, D.; Tiwari, S.; Kaur, M.; Jeong, C.-W.; Nam, Y. & Khan, M. A. 2021. Screening of COVID-19 patients using deep learning and IoT framework. *Cmc-Computers Materials & Continua***,** 3459-3475.

Keat, R. & Urry, J. 2011. *Social Theory as Science (Routledge Revivals)*, Routledge.

Khan, I. 2010. *SOA Tips - Address Scalability Bottlenecks with Distributed Caching* [Online]. Available: https://docs.microsoft.com/en-us/archive/msdn-magazine/2010/june/msdn-magazine-soa-tips-address-scalability-bottlenecks-with-distributed-caching [Accessed].

Kidholm, K.; Svendsen, I. W.; Yderstræde, K.; Ølholm, A. M.; Rayce, K. & Kjølhede, T. 2021. The hospital telemedicine TELEMED database: Providing information on evidence-based telemedicine services to hospital managers and healthcare professionals. *Journal of telemedicine and telecare,* 27**,** 280-287.

Kiourtis, A.; Nifakos, S.; Mavrogiorgou, A. & Kyriazis, D. 2019. Aggregating the syntactic and semantic similarity of healthcare data towards their transformation to HL7 FHIR through ontology matching. *International Journal of Medical Informatics,* 132**,** 104002.

Kissick, W. L. 1994. *Medicine's dilemmas: infinite needs versus finite resources*, Yale University Press.

Kneebone, R. 2003. Simulation in surgical training: educational issues and practical implications. *Medical education,* 37**,** 267-277.

Knight, S. R.; Ng, N.; Tsanas, A.; Mclean, K.; Pagliari, C. & Harrison, E. M. 2021. Mobile devices and wearable technology for measuring patient outcomes after surgery: a systematic review. *NPJ digital medicine,* 4**,** 1-14.

Kooi, T.; Litjens, G.; Van Ginneken, B.; Gubern-Mérida, A.; Sánchez, C. I.; Mann, R.; den Heeten, A. & Karssemeijer, N. 2017. Large scale deep learning for computer aided detection of mammographic lesions. *Medical image analysis,* 35**,** 303-312.

Kovács, G. & Spens, K. M. 2005. Abductive reasoning in logistics research. *International journal of physical distribution & logistics management*.

Kratzke, N. & Quint, P.-C. 2017. Understanding cloud-native applications after 10 years of cloud computing-a systematic mapping study. *Journal of Systems and Software,* 126**,** 1-16.

Kraus, S. Investigating the Capabilities of FHIR Search for Clinical Trial Phenotyping. German Medical Data Sciences: A Learning Healthcare System: Proceedings of the 63rd Annual Meeting of the German Association of Medical Informatics, Biometry and Epidemiology (gmds eV) 2018 in Osnabrück, Germany–GMDS 2018, 2018. IOS Press, 3.

Kruse, C. S.; Stein, A.; Thomas, H. & Kaur, H. 2018. The use of electronic health records to support population health: a systematic review of the literature. *Journal of medical systems,* 42**,** 1-16.

Kubick, W. 2016. *CR 3.0-A Manifesto for The Next Generation of Clinical Research Data Standards* [Online]. Available: https://waynekubick.com/2016/08/30/cr-3-0-a-manifesto-for-the-next-generation-of-clinical-research-data-standards/ [Accessed].

Kuhn, T. 1962. The structure of scientific revolutions. *International Encyclopedia of Unified Science,* 2.

Kuperman, G. J. 2011. Health-information exchange: why are we doing it, and what are we doing? *Journal of the American Medical Informatics Association,* 18**,** 678-682.

Lacy, A.; Bravo, R.; Otero-Piñeiro, A.; Pena, R.; De Lacy, F.; Menchaca, R. & Balibrea, J. 2019. 5G-assisted telementored surgery. *British Journal of Surgery,* 106**,** 1576-1579.

Lal, M. 2015. *Neo4j graph data modeling*, Packt Publishing Ltd.

Leavitt, J. 2010. *Linguistic relativities: Language diversity and modern thought*, Cambridge University Press.

Lee, A. S. 1989. A scientific methodology for MIS case studies. *MIS quarterly***,** 33-50.

Lee, A. S. 2004. Thinking about social theory and philosophy for information systems. *Social theory and philosophy for information systems,* 1**,** 26.

Lee, A. S. & Hubona, G. S. 2009. A scientific basis for rigor in information systems research. *MIS quarterly***,** 237-262.

Leeds-Hurwitz, W. 1993. *Semiotics and communication: Signs, codes, cultures*, Routledge.

Lehne, M.; Luijten, S.; IMBUSCH, P. V. F. G. & Thun, S. 2019a. The Use of FHIR in Digital Health-A Review of the Scientific Literature. *Studies in health technology and informatics,* 267**,** 52-58.

Lehne, M.; Sass, J.; Essenwanger, A.; Schepers, J. & Thun, S. 2019b. Why digital medicine depends on interoperability. *NPJ Digital Medicine,* 2**,** 1-5.

Leroux, H.; Metke-Jimenez, A. & Lawley, M. J. 2017. Towards achieving semantic interoperability of clinical study data with FHIR. *Journal of biomedical semantics,* 8**,** 41.

Lewis, J. & Flower, M. 2014. *Microservices. A definition of this new architectural term* [Online]. Available: http://martinfowler.com/articles/microservices.html [Accessed].

Li, J.; Green, D. L.; Santilli, S.; Ihnat, D.; James Valentine, R. & Reed, A. B. 2022. Televascular consultation is one answer to rural vascular surgery shortage. *Journal of telemedicine and telecare,* 28**,** 291-295.

Li, S.; Da Xu, L. & Zhao, S. 2015. The internet of things: a survey. *Information Systems Frontiers,* 17**,** 243-259.

Liszka, J. J. 1990. Peirce's interpretant. *Transactions of the Charles S. Peirce Society,* 26**,** 17-62.

Little, P.; Everitt, H.; Williamson, I.; Warner, G.; Moore, M.; Gould, C.; Ferrier, K. & Payne, S. 2001. Preferences of patients for patient centred approach to consultation in primary care: observational study. *Bmj,* 322**,** 468.

Liu, K. 2005. Requirements reengineering from legacy information systems using semiotic techniques. *Syst Signs Actions Int J Commun Inf Technol Work,* 1**,** 38-61.

Liu, K. & Guo, H. 2020. Artificial intelligence and data analytics in digital business transformation before, during and post COVID-19.

Liu, K. & Guo, H. Digital innovation and transformation to business ecosystems. International Conference on Enterprise Information Systems, 2021. Springer, 793-803.

Liu, K. & Li, W. 2015. *Organisational semiotics for business informatics*, Routledge Abingdon.

Liu, K.; Nakata, K. & Harty, C. 2010. Pervasive informatics: theory, practice and future directions. *Intelligent Buildings International,* 2**,** 5-19.

Liu, K.; Sun, L. & Fu, Y. 2008. Ontological modelling of content management and provision. *Information and Software Technology,* 50**,** 1155-1164.

Liu, K. & Tan, C. Semiotics in digital visualisation. International Conference on Enterprise Information Systems, 2014. Springer, 3-13.

Liu, S.; Li, W. & Liu, K. Pragmatic oriented data interoperability for smart healthcare information systems. 2014 14th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, 2014. IEEE, 811-818.

Liu, S.; Ni, Y.; Mei, J.; Li, H.; Xie, G.; Hu, G.; Liu, H.; Hou, X. & Pan, Y. iSMART: Ontology-based Semantic query of CDA documents. AMIA Annual Symposium Proceedings, 2009. American Medical Informatics Association, 375.

Lloyd McKenzie. 2020. *Resource Considerations* [Online]. Available: https://confluence.hl7.org/display/FHIR/Resource+Considerations [Accessed].

Lopez, L. 2013. Induction And Deduction As Entry Points Into Qualitative Data Analysis: How Can Early Analytical Choices Affect The Theorizing Process?

Lukka, K. & Modell, S. 2010. Validation in interpretive management accounting research. *Accounting, organizations and society,* 35**,** 462-477.

Lyytinen, K. J. 1985. Implications of theories of language for information systems. *MIS Quarterly***,** 61-74.

Magee, B. 1973. *Popper*, London : Fontana.

Magsi, H.; Sodhro, A. H.; Chachar, F. A.; Abro, S. A. K.; Sodhro, G. H. & Pirbhulal, S. Evolution of 5G in Internet of medical things. 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), 2018. IEEE, 1-7.

Mainzer, K. 2007. *Thinking in complexity: The computational dynamics of matter, mind, and mankind*, Springer Science & Business Media.

Mamom, J. & Daovisan, H. 2022. Telenursing: How do caregivers treat and prevent pressure injury in bedridden patients during the COVID-19 pandemic in Thailand? Using an embedded approach. *Journal of Telemedicine and Telecare***,** 1357633X221078485.

Mandel, J. C.; Kreda, D. A.; Mandl, K. D.; Kohane, I. S. & Ramoni, R. B. 2016. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. *Journal of the American Medical Informatics Association,* 23**,** 899-908.

Manogaran, G.; Varatharajan, R.; Lopez, D.; Kumar, P. M.; Sundarasekar, R. & Thota, C. 2018. A new architecture of Internet of Things and big data ecosystem for secured smart healthcare monitoring and alerting system. *Future Generation Computer Systems,* 82**,** 375-387.

March, S.; Hevner, A. & Ram, S. 2000. Research commentary: An agenda for information technology research in heterogeneous and distributed environments. *Information Systems Research,* 11**,** 327-341.

Martini, C. 2021. What "evidence" in evidence-based medicine? *Topoi,* 40**,** 299-305.

Mason, J. 2019. *Precision and Accuracy in the Use of Language* [Online]. Available: https://jmason37-80878.medium.com/precision-and-accuracy-in-the-use-of-language-755e11b75778 [Accessed].

Mavrogiorgou, A.; Kiourtis, A.; Touloupou, M.; Kapassa, E. & Kyriazis, D. 2019. Internet of medical things (IoMT): Acquiring and transforming data into HL7 FHIR through 5G network slicing. *Emerging Science Journal,* 3**,** 64-77.

McClure, R. C.; Macumber, C. L.; Skapik, J. L. & Smith, A. M. 2020. Igniting harmonized digital clinical quality measurement through terminology, CQL, and FHIR. *Applied Clinical Informatics,* 11**,** 023-033.

McDonald, C. J.; Huff, S. M.; Suico, J. G.; Hill, G.; Leavelle, D.; Aller, R.; Forrey, A.; Mercer, K.; DeMoor, G. & Hook, J. 2003. LOINC, a universal standard for identifying laboratory observations: a 5-year update. *Clinical chemistry,* 49**,** 624-633.

McGovern, J.; Tyagi, S.; Stevens, M. & Mathew, S. 2003. *Java web services architecture*, Elsevier.

McKenna, S.; Richardson, J. & Manroop, L. 2011. Alternative paradigms and the study and practice of performance management and evaluation. *Human Resource Management Review,* 21**,** 148-157.

McKenzie, L. 2016. *FHIR for architects* [Online]. Available: https://www.academia.edu/28838227/FHIR_for_Architects [Accessed].

McKinsey&Company. 2019. *Digital health ecosystems: A payer perspective* [Online]. Available: https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/digital-health-ecosystems-a-payer-perspective [Accessed].

McNicoll, I.; Mehrkar, A. & Shannon, T. 2019. *FHIR and openEHR* [Online]. Available: https://www.interopen.org/wp-content/uploads/2019/03/INTEROPen-openEHR-and-FHIR_March2019.pdf [Accessed].

Mead, C. N. 2006. Data interchange standards in healthcare it-computable semantic interoperability: Now possible but still difficult. do we really need a better mousetrap? *Journal of Healthcare Information Management,* 20**,** 71.

Mettler, M. Blockchain technology in healthcare: The revolution starts here. 2016 IEEE 18th international conference on e-health networking, applications and services (Healthcom), 2016. IEEE, 1-3.

Mingers, J. 2004. Real-izing information systems: critical realism as an underpinning philosophy for information systems. *Information and organization,* 14**,** 87-103.

Minnameier, G. The logicality of abduction, deduction, and induction. Ideas in action: Proceedings of the applying peirce conference, 2010. Nordic Pragmatism Network Helsinki, 239-251.

Mohapatra, S. & Sahoo, P. K. 2022. Internet of Medical Things: Applications and Research Issues in Healthcare Monitoring. *IoT Applications for Healthcare Systems***,** 1-31.

Moreira, J.; Pires, L. F.; van Sinderen, M. & Daniele, L. SAREF4health: IoT Standard-Based Ontology-Driven Healthcare Systems. FOIS, 2018. 239-252.

Mouttham, A.; Kuziemsky, C.; Langayan, D.; Peyton, L. & Pereira, J. 2012. Interoperable support for collaborative, mobile, and accessible health care. *Information systems frontiers,* 14**,** 73-85.

Muggleton, S. H.; Lin, D.; Pahlavi, N. & Tamaddoni-Nezhad, A. 2014. Meta-interpretive learning: application to grammatical inference. *Machine learning,* 94**,** 25-49.

Myers, G. 1978. Composite/structured design(Book on computer programming). *Research supported by IBM Corp. New York, Van Nostrand Reinhold Co., 1978. 183 p.*

Nadareishvili, I.; Mitra, R.; McLarty, M. & Amundsen, M. 2016. *Microservice architecture: aligning principles, practices, and culture*, " O'Reilly Media, Inc.".

Nadri, H.; Rahimi, B.; Timpka, T. & Sedghi, S. 2017. The top 100 articles in the medical informatics: a bibliometric analysis. *Journal of medical systems,* 41**,** 1-12.

Newman, S. 2015. *Building microservices: designing fine-grained systems*, " O'Reilly Media, Inc.".

Ngwenyama, O. K. & Lee, A. S. 1997. Communication richness in electronic mail: Critical social theory and the contextuality of meaning. *MIS quarterly***,** 145-167.

NHS. 2020. *FHIR UK Core* [Online]. Available: https://digital.nhs.uk/services/fhir-uk-core [Accessed].

NHS Digital. 2021. *FHIR UK Core* [Online]. Available: https://digital.nhs.uk/services/fhir-uk-core [Accessed].

Nie, D. & Roantree, M. Detecting Multi-Relationship Links in Sparse Datasets. 2019. ICEIS.

Noon, A. K.; Aziz, O.; Zahra, I. & Anwar, M. Implementation of Blockchain in Healthcare: A Systematic Review. 2021 International Conference on Innovative Computing (ICIC), 2021. IEEE, 1-10.

open EHR 2003. Open industry specifications, models and software for e-health.

openEHR. 2002. *OpenEHR* [Online]. Available: http://www.openehr.org [Accessed].

openEHR 2007. OpenEHR Design Principles.

Orlikowski, W. J. & Baroudi, J. J. 1991. Studying information technology in organizations: Research approaches and assumptions. *Information systems research,* 2**,** 1-28.

Ouksel, A. M. & Sheth, A. 1999. Semantic interoperability in global information systems. *ACM Sigmod Record,* 28**,** 5-12.

Pantanowitz, L.; Sharma, A.; Carter, A. B.; Kurc, T.; Sussman, A. & Saltz, J. 2018. Twenty years of digital pathology: an overview of the road travelled, what is on the horizon, and the emergence of vendor-neutral archives. *Journal of pathology informatics,* 9**,** 40.

Papazoglou, M. P. & Van Den Heuvel, W.-J. 2007. Service oriented architectures: approaches, technologies and research issues. *The VLDB journal,* 16**,** 389-415.

Parnas, D. L. 1972. On the criteria to be used in decomposing systems into modules. *Pioneers and Their Contributions to Software Engineering.* Springer.

Peirce, C. S. 1931. *Collected papers of charles sanders peirce*, Harvard University Press.

Peirce, C. S. 1958. *Collected Papers: Science and philosophy and Reviews, correspondence, and bibliography*, Belknap Press of Harvard University Press.

Pelzang, R. 2010. Time to learn: understanding patient-centred care. *British journal of nursing,* 19**,** 912-917.

Peng, C. & Goswami, P. 2019. Meaningful integration of data from heterogeneous health services and home environment based on ontology. *Sensors,* 19**,** 1747.

Petty, N. J.; Thomson, O. P. & Stew, G. 2012. Ready for a paradigm shift? Part 1: Introducing the philosophy of qualitative research. *Manual therapy,* 17**,** 267-274.

Pfaff, E. R.; Champion, J.; Bradford, R. L.; Clark, M.; Xu, H.; Fecho, K.; Krishnamurthy, A.; Cox, S.; Chute, C. G. & Taylor, C. O. 2019. Fast healthcare interoperability resources (FHIR) as a meta model to integrate common data models: development of a tool and quantitative validation study. *JMIR medical informatics,* 7**,** e15199.

Piantadosi, S. T.; Tily, H. & Gibson, E. 2012. The communicative function of ambiguity in language. *Cognition,* 122**,** 280-291.

Pierce, C. 1998. Nomenclature and Divisions of Triadic Relations, as Far as They Are Determined. *Essential Peirce: Selected Philosophical Writings. Bloomington: Indiana University Press***,** 298.

Pietsch, W. 2021. *Big Data*, Cambridge University Press.

Plastiras, P.; O'Sullivan, D. & Weller, P. 2014. An ontology-driven information model for interoperability of personal and electronic health records.

Plsek, P. E. & Greenhalgh, T. 2001. The challenge of complexity in health care. *Bmj,* 323**,** 625-628.

Popper, K. 2013. *Realism and the aim of science: From the postscript to the logic of scientific discovery*, Routledge.

Pottle, J. 2019. Virtual reality and the transformation of medical education. *Future healthcare journal,* 6**,** 181.

Practice, A. P. T. F. o. E.-B. 2006. Evidence-based practice in psychology. *The American Psychologist,* 61**,** 271.

Pustokhina, I. V.; Pustokhin, D. A.; Gupta, D.; Khanna, A.; Shankar, K. & Nguyen, G. N. 2020. An effective training scheme for deep neural network in edge computing enabled Internet of medical things (IoMT) systems. *IEEE Access,* 8**,** 107112-107123.

PwC. 2015. *Realising the Benefits of Mobileenabled IoT Solutions* [Online]. Available: https://www.gsma.com/iot//wp-content/uploads/2015/04/Realising-the-benefits-of-mobile-IoTsolutions-v4_7Apr2015.pdf [Accessed].

Raghupathi, W. & Umar, A. 2008. Exploring a model-driven architecture (MDA) approach to health care information systems development. *International journal of medical informatics,* 77**,** 305-314.

Rehman, A.; Naz, S. & Razzak, I. 2022. Leveraging big data analytics in healthcare enhancement: trends, challenges and opportunities. *Multimedia Systems,* 28**,** 1339-1371.

Reyes-Galaviz, O. F.; Pedrycz, W.; He, Z. & Pizzi, N. J. 2017. A supervised gradient-based learning algorithm for optimized entity resolution. *Data & Knowledge Engineering,* 112**,** 106-129.

Richardson, C. 2018. Pattern: Monolithic architecture.

Rinner, C. & Duftschmid, G. Bridging the gap between HL7 CDA and HL7 FHIR: A JSON based mapping. eHealth, 2016. 100-106.

Ristevski, B. & Chen, M. 2018. Big data analytics in medicine and healthcare. *Journal of integrative bioinformatics,* 15.

Rodrigues, J. J.; Segundo, D. B. D. R.; Junqueira, H. A.; Sabino, M. H.; Prince, R. M.; Al-Muhtadi, J. & De Albuquerque, V. H. C. 2018. Enabling technologies for the internet of health things. *Ieee Access,* 6**,** 13129-13141.

Rosenau, L.; Majeed, R. W.; Ingenerf, J.; Kiel, A.; Kroll, B.; Köhler, T.; Prokosch, H.-U. & Gruendner, J. 2022. Generation of a Fast Healthcare Interoperability Resources (FHIR)-based Ontology for Federated Feasibility Queries in the Context of COVID-19: Feasibility Study. *JMIR Medical Informatics,* 10**,** e35789.

Russell, B. 2001. *The problems of philosophy*, OUP Oxford.

Ryan, A. Towards semantic interoperability in healthcare: ontology mapping from SNOMED-CT to HL7 version 3. Proceedings of the second Australasian workshop on Advances in ontologies-Volume 72, 2006. Citeseer, 69-74.

Sackett, D. L. Evidence-based medicine. Seminars in perinatology, 1997. Elsevier, 3-5.

Salovaara, A. & Merikivi, J. 2015. IS research progress would benefit from increased falsification of existing theories.

Saripalle, R.; Runyan, C. & Russell, M. 2019. Using HL7 FHIR to achieve interoperability in patient health record. *Journal of biomedical informatics,* 94**,** 103188.

Sarker, S.; Xiao, X. & Beaulieu, T. 2013. Guest editorial: Qualitative studies in information systems: A critical review and some guiding principles. *MIS quarterly,* 37**,** iii-xviii.

Savan, D. 1987. *An Introduction to CS Peirce's Full System of Semeiotic* Toronto Semiotic Circle, Victoria College in the University of Toronto.

Sayeed, R.; Gottlieb, D. & Mandl, K. D. 2020. SMART Markers: collecting patient-generated health data as a standardized property of health information technology. *NPJ digital medicine,* 3**,** 1-8.

Schriml, L. M. & Mitraka, E. 2015. The Disease Ontology: fostering interoperability between biological and clinical human disease-related data. *Mammalian Genome,* 26**,** 584-589.

Sciama, D. W. 1964. The physical structure of general relativity. *Reviews of Modern Physics,* 36**,** 463.

Scriney, M.; McCarthy, S.; McCarren, A.; Cappellari, P. & Roantree, M. 2019. Automating data mart construction from semi-structured data sources. *The Computer Journal,* 62**,** 394-413.

Selic, B. 2003. The pragmatics of model-driven development. *IEEE software,* 20**,** 19-25.

Sheth, A. 1988. Building federated database systems. *Newsletter,* 10**,** 50-58.

Sheth, A. P. & Larson, J. A. 1990. Federated database systems for managing distributed, heterogeneous, and autonomous databases. *ACM Computing Surveys (CSUR),* 22**,** 183-236.

Shi, W.; Giuste, F. O.; Zhu, Y.; Carpenter, A. M.; Iwinski, H. J.; Hilton, C.; Wattenbarger, J. M. & Wang, M. D. A FHIR-compliant application for multi-site and multi-modality pediatric scoliosis patient rehabilitation. 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2021. IEEE, 1524-1527.

Shortliffe, E. H. 1976. *Computer-based Medical Consultations, MYCIN*, Elsevier.

Sink, J.; Blatt, S.; Yoo, D.; Henry, M.; Yang, S. D.; Vasaiwala, R.; Ghadiali, L.; Adams, W. & Bouchard, C. S. 2022. A novel telemedicine technique for evaluation of ocular exam findings via smartphone images. *Journal of Telemedicine and Telecare,* 28**,** 197-202.

Sirota-Cohen, C.; Rosipko, B.; Forsberg, D. & Sunshine, J. L. 2019. Implementation and benefits of a vendor-neutral archive and enterprise-imaging management system in an integrated delivery network. *Journal of Digital Imaging,* 32**,** 211-220.

SMART. 2009. *SMART introduction* [Online]. Available: https://docs.smarthealthit.org/ [Accessed].

Smith, B. 2012. Ontology. *The furniture of the world.* Brill.

Sowa, J. F. 1983. Conceptual structures: information processing in mind and machine.

Sox, H. C. & Greenfield, S. 2009. Comparative effectiveness research: a report from the Institute of Medicine. *Annals of internal medicine,* 151**,** 203-205.

Staab, S. Concepts in Application Context. International Conference on Formal Concept Analysis, 2019. Springer, 45-52.

Stair, R. & Reynolds, G. 2020. *Principles of information systems*, Cengage Learning.

Stamper, R. 1991. The semiotic framework for information systems research. *Information systems research: Contemporary approaches and emergent traditions***,** 515-528.

Stevens, W. P.; Myers, G. J. & Constantine, L. L. 1974. Structured design. *IBM Systems Journal,* 13**,** 115-139.

Stewart, M. 2001. Towards a global definition of patient centred care: the patient should be the judge of patient centred care. British Medical Journal Publishing Group.

Stump, D. 1990. Introduction to classical mechanics.

Tan, C.; Abdaless, S. & Liu, K. Norm-based abduction process (NAP) in developing information architecture. International Conference on Informatics and Semiotics in Organisations, 2018. Springer, 33-42.

Targon, V. 2018. Toward semiotic artificial intelligence. *Procedia computer science,* 145**,** 555-563.

Templeton, M.; Brill, D.; Dao, S. K.; Lund, E.; Ward, P.; Chen, A. L. & MacGregor, R. 1987. Mermaid—A front-end to distributed heterogeneous databases. *Proceedings of the IEEE,* 75**,** 695-708.

The Open Group 2018. *The TOGAF® Standard, Version 9.2*, Van Haren Publishing, Zaltbommel, www.vanharen.net.

The OpenEHR foundation 2003. openEHR architecture overview.

Thornton, M. 2010. The "meaningful use" regulation for electronic health records. *n Engl j Med,* 2010**,** 501-504.

Thota, C.; Sundarasekar, R.; Manogaran, G.; Varatharajan, R. & Priyan, M. 2018. Centralized fog computing security platform for IoT and cloud in healthcare system. *Fog Computing: Breakthroughs in Research and Practice.* IGI global.

Tolk, A.; Diallo, S. Y. & Turnitsa, C. D. 2007. Applying the levels of conceptual interoperability model in support of integratability, interoperability, and composability for system-of-systems engineering. *Journal of Systems, Cybernetics, and Informatics,* 5.

Tsichritzis, D. & Klug, A. 1978. The ANSI/X3/SPARC DBMS framework report of the study group on database management systems. *Information systems,* 3**,** 173-191.

Tummers, J.; Tekinerdogan, B.; Tobi, H.; Catal, C. & Schalk, B. 2021. Obstacles and features of health information systems: A systematic literature review. *Computers in Biology and Medicine,* 137**,** 104785.

Tute, E.; Scheffner, I. & Marschollek, M. 2021. A method for interoperable knowledge-based data quality assessment. *BMC medical informatics and decision making,* 21**,** 1-14.

Ullah, F.; Habib, M. A.; Farhan, M.; Khalid, S.; Durrani, M. Y. & Jabbar, S. 2017. Semantic interoperability for big-data in heterogeneous IoT infrastructure for healthcare. *Sustainable cities and society,* 34**,** 90-96.

Ulrich, H.; Kock, A.-K.; Duhm-Harbeck, P.; Habermann, J. K. & Ingenerf, J. Metadata repository for improved data sharing and reuse based on HL7 FHIR. MIE, 2016. 162-166.

Ulrich, W. 2001. A Philisophical Staircase for Information Systems Definition, Design and Development. *Journal of Information Technology Theory and Application (JITTA),* 3**,** 9.

Vaishnavi, V. & Kuechler, W. 2004. Design research in information systems.

VanLare, J. M.; Conway, P. H. & Sox, H. C. 2010. Five next steps for a new national program for comparative-effectiveness research. *New England Journal of Medicine,* 362**,** 970-973.

Wache, H.; Voegele, T.; Visser, U.; Stuckenschmidt, H.; Schuster, G.; Neumann, H. & Hübner, S. Ontology-Based Integration of Information-A Survey of Existing Approaches. Ois@ ijcai, 2001.

Wagholikar, K. B.; Mandel, J. C.; Klann, J. G.; Wattanasin, N.; Mendis, M.; Chute, C. G.; Mandl, K. D. & Murphy, S. N. 2017. SMART-on-FHIR implemented over i2b2. *Journal of the American Medical Informatics Association,* 24**,** 398-402.

Walker, J.; Pan, E.; Johnston, D.; Adler-Milstein, J.; Bates, D. W. & Middleton, B. 2005. The Value of Health Care Information Exchange and Interoperability: There is a business case to be made for spending money on a fully standardized nationwide system. *Health affairs,* 24**,** W5-10-W5-18.

Walsham, G. 1995. Interpretive case studies in IS research: nature and method. *European Journal of information systems,* 4**,** 74-81.

Walsham, G. 2006. Doing interpretive research. *European journal of information systems,* 15**,** 320-330.

Wan, S.; Nappi, M.; Chen, C. & Berretti, S. 2022. Guest Editorial Emerging IoT-Driven Smart Health: From Cloud to Edge. *IEEE Journal of Biomedical and Health Informatics,* 26**,** 937-938.

Warner, J. L.; Rioth, M. J.; Mandl, K. D.; Mandel, J. C.; Kreda, D. A.; Kohane, I. S.; Carbone, D.; Oreto, R.; Wang, L. & Zhu, S. 2016. SMART precision cancer medicine: a FHIR-based app to provide genomic information at the point of care. *Journal of the American Medical Informatics Association,* 23**,** 701-710.

Wen, H. J. & Tan, J. The evolving face of telemedicine & e-health: opening doors and closing gaps in e-health services opportunities & challenges. 36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the, 2003. IEEE, 12 pp.

West Health Institute. 2013. *The Value of Medical Device Interoperability* [Online]. Available: https://www.westhealth.org/wp-content/uploads/2015/02/The-Value-of-Medical-Device-Interoperability.pdf [Accessed].

White, L. A. 1940. The symbol: The origin and basis of human behavior. *Philosophy of Science,* 7**,** 451-463.

Williams, M. 2000. Interpretivism and generalisation. *Sociology,* 34**,** 209-224.

Wittgenstein, L. 2019. *Philosophical Investigations*.

Wolff, E. 2016. *Microservices: flexible software architecture*, Addison-Wesley Professional.

Wood, C.; Winton, B.; Carter, K.; Benkert, S.; Dodd, L. & Bradley, J. 2016. How blockchain technology can enhance EHR operability.

Worden, R. & Scott, P. 2011. Simplifying HL7 version 3 messages. *User Centred Networked Health Care***,** 709-713.

Xiao, Z.; Wijegunaratne, I. & Qiang, X. Reflections on SOA and Microservices. 2016 4th International Conference on Enterprise Systems (ES), 2016. IEEE, 60-67.

Xu, H.; Cox, S.; Stillwell, L.; Pfaff, E.; Champion, J.; Ahalt, S. C. & Fecho, K. 2020. FHIR PIT: an open software application for spatiotemporal integration of clinical data and environmental exposures data. *BMC medical informatics and decision making,* 20**,** 1-8.

Yang, J.-J.; Li, J.; Mulder, J.; Wang, Y.; Chen, S.; Wu, H.; Wang, Q. & Pan, H. 2015. Emerging information technologies for enhanced healthcare. *Computers in industry,* 69**,** 3-11.

Yang, Q.; Liu, Y.; Chen, T. & Tong, Y. 2019. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST),* 10**,** 1-19.

Yeung, A. W. K.; Tosevska, A.; Klager, E.; Eibensteiner, F.; Laxar, D.; Stoyanov, J.; Glisic, M.; Zeiner, S.; Kulnik, S. T. & Crutzen, R. 2021. Virtual and augmented reality applications in medicine: analysis of the scientific literature. *Journal of medical internet research,* 23**,** e25499.

Yu, C. H. 1994. Abduction? Deduction? Induction? Is There a Logic of Exploratory Data Analysis?

Yu, K.-H.; Beam, A. L. & Kohane, I. S. 2018. Artificial intelligence in healthcare. *Nature biomedical engineering,* 2**,** 719-731.

Yue, X.; Wang, H.; Jin, D.; Li, M. & Jiang, W. 2016. Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control. *Journal of medical systems,* 40**,** 1-8.

Zachman, J. A. 1999. A framework for information systems architecture. *IBM systems journal,* 38**,** 454-470.

Zendejas, B.; Brydges, R.; Hamstra, S. J. & Cook, D. A. 2013. State of the evidence on simulation-based training for laparoscopic surgery: a systematic review. *Annals of surgery,* 257**,** 586-593.

Zhang, P.; White, J.; Schmidt, D. C.; Lenz, G. & Rosenbloom, S. T. 2018. FHIRChain: applying blockchain to securely and scalably share clinical data. *Computational and structural biotechnology journal,* 16**,** 267-278.

Zong, N.; Stone, D. J.; Sharma, D. K.; Wen, A.; Wang, C.; Yu, Y.; Huang, M.; Liu, S.; Liu, H. & Shi, Q. 2021. Modeling cancer clinical trials using HL7 FHIR to support downstream applications: A case study with colorectal cancer data. *International Journal of Medical Informatics,* 145**,** 104308.

Zuboff, S. 1988. In the age of the smart machine.

Appendix-1 Mapping MIMIC III Data into FHIR

This study presents some MIMIC data using the FHIR specification. As shown in the following matrixes, most of attributes defined in the FHIR cannot find corresponding values in MIMIC III. The main reason is that FHIR is dedicated to describing the whole scenarios of healthcare services and MIMIC III data is only ICU-related data; the second is that the data about user profiles in MIMIC III are masked. The flexibility of FHIR provides a freedom to implementors to using FHIR resources to describe the local data. In this practice, the two FHIR resources, *Patient* and *Encounter*, are adopted to reinterpret ICU-related data. A total of three encounters are used here, which represent the health of the patient at the time of admission, the situation in the emergency room, and the situation in the ICU ward. These three encounters correspond to three different start and end times respectively.

However, rather than employ three encounters, other implementor may nested logic to express the relationship between these three scenarios, such as using the **partOf** attribute to connect admission records and emergency records, as shown in Line 52 of FHIR.encounter-1. Therefore, in the face of flexible FHIR specifications and highly differentiated local databases, the matching work of FHIR is endowed with the subjective judgment of many implementers. Since the matching process between FHIR and local data is a black box to the data user, the data that users access according to the definition of FHIR may not be what they expect. This is a typical semantic ambiguity for data consumers introduced by different implementers.

Appendix-1 Mapping MIMIC III Data into FHIR

FHIR. patient

No	FHIR.Patient	MIMIC
1	identifier	admissions_row_id admissions_subject_id admissions_hadm_id
2	active	
3	name	
4	telecom	
5	gender	patient_gender
6	birthDate	patient_dob
7	deceasedBoolean	patient_expire_flag
8	deceasedDateTime	patient_dod + admission_deathtime
9	deceased	
10	address	
11	maritalStatus	admissions_marital_status
12	multipleBirth[x]	
13	photo	
14	relationship	
15	name	
16	telecom	
17	address	
18	gender	
19	organization	
20	period	
21	contact	
22	language	admissions_language
23	preferred	
24	communication	
25	generalPractitioner	
26	managingOrganization	
29	link	

FHIR. encounter-1

No	FHIR.Encounter	MIMIC
1	identifier	admissions_row_id+admissions_subject_id+admissions_hadm_id
2	status	
3	statusHistory	
4	status	
5	period	
6	class	admissions_admission_type
7	classHistory	
8	class	
9	period	
10	type	
11	serviceType	
12	priority	
13	subject	
14	episodeOfCare	
15	basedOn	
16	participant	
17	type	
18	period	
19	individual	
20	appointment	
21	period	
22	start	admissions_admittime
23	stop	admissions_dischtime
24	length	
25	reasonCode	
26	reasonReference	
27	reference	icustays_dbsource
28	type	
29	identifier	icustays_icustay_id
30	display	
31	diagnosis	
32	condition	
33	use	diagnoses_ICD_ICD9_code
34	rank	
35	account	
36	hospitalization	
37	preAdmissionIdentifier	
38	origin	
39	admitSource	admissions_admission__location
40	reAdmission	
41	dietPreference	
42	specialCourtesy	
43	specialArrangement	
44	destination	admissions_discharge__location
45	dischargeDisposition	
46	location	
47	location	
48	status	
49	physicalType	
50	period	
51	serviceProvider	
52	part of	FHIR.encounter-2 (if class=EMERGENCY) FHIR.encounter-3 (if reasonReference.identifier != Null)

FHIR. encounter-2

No	FHIR.Encounter	MIMIC
1	identifier	admissions_row_id+admissions_subject_id+admissions_hadm_id
2	status	
3	statusHistory	
4	status	
5	period	
6	class	admissions_admission_type
7	classHistory	
8	class	
9	period	
10	type	
11	serviceType	
12	priority	
13	subject	
14	episodeOfCare	
15	basedOn	
16	participant	
17	type	
18	period	
19	individual	
20	appointment	
21	period	
22	start	admissions_edregtime (Time that the patient was registered and discharged from the emergency department.)
23	stop	admissions_edoutime (Time that the patient was registered and discharged from the emergency department.)
24	length	
25	reasonCode	
26	reasonReference	
29	diagnosis	
28	condition	
30	use	
27	rank	
31	account	
32	hospitalization	
33	preAdmissionIdentifier	
34	origin	
35	admitSource	
36	reAdmission	
37	dietPreference	
38	specialCourtesy	
39	specialArrangement	
40	destination	
41	dischargeDisposition	
42	location	
43	location	
44	status	
45	physicalType	
46	period	
47	serviceProvider	
48	part of	

FHIR. encounter-3

No	FHIR.Encounter	MIMIC
1	identifier	admissions_row_id+admissions_subject_id+admissions_hadm_id
2	status	
3	statusHistory	
4	status	Service_prev_service
5	period	
6	start	
7	end	Service_transfertime
8	class	
9	classHistory	
10	class	
11	period	
12	type	
13	serviceType	Service_curr_service
14	priority	
15	subject	
16	episodeOfCare	
17	basedOn	
18	participant	
19	type	icustays_first_careunit & icustays_last_careunit
20	period	icustays_intime
21	start	icustays_outtime
22	end	
23	individual	
24	appointment	
25	period	
26	start	icustays_intime
27	stop	icustays_outtime
28	length	icustays_los
29	reasonCode	
30	reasonReference	
31	diagnosis	
32	condition	
33	use	
34	rank	
35	account	
36	hospitalization	
37	preAdmissionIdentifier	
38	origin	
39	admitSource	
40	reAdmission	
41	dietPreference	
42	specialCourtesy	
43	specialArrangement	
44	destination	
45	dischargeDisposition	
46	location	
47	location	
48	status	
49	physicalType	icustays_first_wardid & icustays_last_wardid
50	period	
51	start	icustays_intime
52	end	icustays_outtime
53	serviceProvider	
54	reference	
55	type	
56	identifier	
57	display	
58	part of	

Appendix-2 Information Architecture for Large-scale Systems

The major challenge in information architecture development is to tackle large-scale complex software systems. This challenge can be dated back to the 1970s when computer scientists firstly applied the ancient and proven technique, namely 'divide and conquer', into software architecture (Gauthier and Ponto, 1970) which divided a complex problem into feasible minor issues. This design paradigm allows applications to be built out of independently deployable modules, thus improving the efficiency of software development. Subsequently, the concepts of 'Modularity and Information hiding' are introduced by Parnas (1972) to describe a mechanism for decomposing a system into modules to improve flexibility and comprehensibility. Software modularity herein refers to the process of decomposing a program into smaller programs with standardised communication interfaces; information hiding is to hide design decisions that are taken inside a module. The purpose of this design is to improve the flexibility of the whole system and shorten the software development time.

Later, based on modularity, Dijkstra (1974) introduces the concept of **Separation of Concern**, emphasising that each distinct software module separated from applications can only address one particular concern. After that, the idea of software system modularisation is expanded by Stevens *et al.* (1974) and Myers (1978). They propose a pair of concepts: **Cohesion** and **Coupling**. Coupling describes the relationships between software modules, represented by the degree to which a single unit is independent of others. Cohesion describes the relationships within modules, characterised by the degree to which a part of a codebase forms a logically single unit/service function. The works of Parnas, Dijkstra, Stevens, Myers and others lead to the rise of modular software development in the 1970s and forms the fundamental design principle for large-scale and complex software systems: '**loosely coupled, highly cohesive**'. 'Loosely coupled' means that software modules in a system or a network should depend on each other to the most minor extent, and 'highly cohesive' means that one module should focus on a single functionality. This design philosophy lays the foundation of the software design paradigm shifting from monolithic to composable architecture.

In the earlier stages of software architecture development, monolithic architecture is the mainly adopted pattern, which integrates a full breadth of functionalities into a single software system. As shown in Figure 51, the user interface, business logic and database are consolidated into a single program. The monolithic design philosophy tends toward performance perfectionism because the integrated program curtails cross-cutting concerns to the most significant extent. Thus, less operational overhead is needed. Furthermore, the shared database in monolithic architecture also minimises the latency of data access by different services, which is simple to scale horizontally and supported by a load balancer. This design philosophy is suitable for small and straightforward logic applications to fully leverage the advantages of being simple to develop, test and deploy. However, the drawbacks of monolithic architecture are significant. As all counterparts are encapsulated into one signal application, services cannot be deployed separately nor scaled independently. Especially when the application becomes complex, and the organisation grows in size, the drawbacks of monolithic architecture become increasingly significant (Lewis and Flower, 2014, Richardson, 2018) :

- High complexity of software development; The logic of the application can be challenging to understand and modify. Consequently, the new requirements are fulfilled in a more extended development period. So, the monolithic modifications are slow and costly.
- Continuous deployment is complex; Lack of flexibility for changes: a tiny modification could lead to the entire monolith system being rebuilt and deployed.
- Scaling the application is difficult; As a single logical executable, each application instance accesses all of the data, which leads to low system effectiveness, high memory consumption, and CPU intensive.

- Requires a long-term commitment to a technology stack; A monolithic application can be difficult to incrementally adopt newer technology.

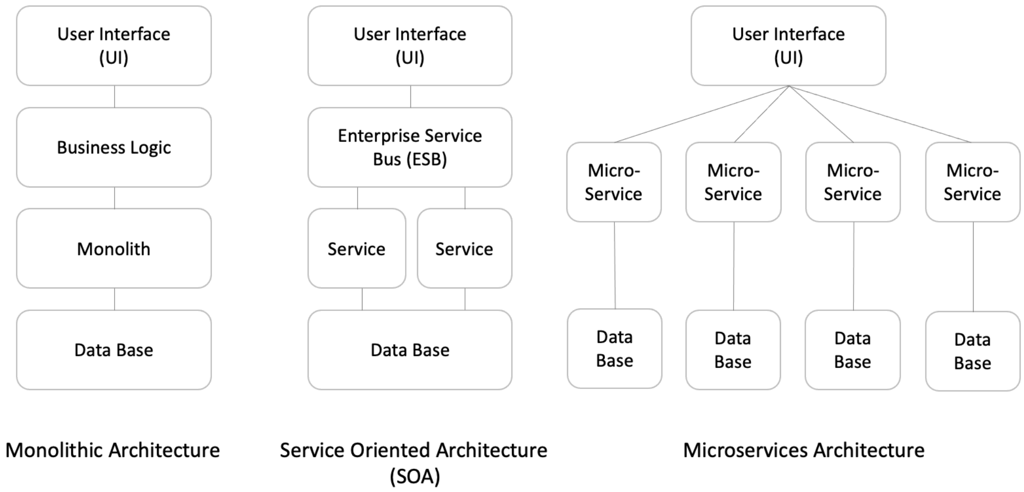


Figure 51 The evolution of IS architecture design paradigm

Monolithic applications consist of interdependent and indivisible units. To avoid the above limitations brought by monolithic architecture, Service Oriented Architecture (SOA) break the entire application into loosely coupled independent *Services.* As shown in Figure 51, *Service* in SOA refers to a reusable software functionality that can be invoked by various business logic to fulfil users' requirements (Cerny *et al.*, 2017). Each *Service* executes a specific function, for example, user authentication and top-up account. These *Services* are loosely coupled with each other, meaning each *Service* is internal cohesion and provides interfaces to external components. Ideally, the only connection between *Services* is their service contract (Brown *et al.*, 2005), a binding relationship reflecting business processes. Because of this modular architecture, development teams can develop software in parallel, and the services can be reused across business processes, so the development cycle will be significantly shortened. In Figure 51, above the *Services* is Enterprise Service Bus (ESB), a set of rules and principles that reflects business logic for integrating numerous applications over a bus-like infrastructure. The following content will discuss ESB's advantages and disadvantages in SOA in detail.

In summary, SOA encapsulates data and application functionality into a service container (Erl, 2005) to improve the system's flexibility and agility and the value stream in organisations (Hirschheim *et al.*, 2010). Compared to the monolithic architecture, SOA is characterised by: 1) modularity, which referrers the partially autonomous subsystems following module or component design principles (cohesive) (McGovern *et al.*, 2003); 2) loose coupling, which means the logical dependencies between services are as low as possible (Papazoglou and Van Den Heuvel, 2007); 3) interface standards, which is used to ensure the interoperability in heterogeneous environments and to guarantee seamless integration. (March *et al.*, 2000). This design paradigm suggests decomposing systems into Services available over a domain and integrating them across heterogenous platforms (Cerny *et al.*, 2017). The first advantage is to increase business agility. In the collection of *Services*, the *Services* can easily be assembled and reused in a manner that allows firms to respond to functional changes flexibly Second, SOA standardises resources for sharing and cooperative purposes. The self-independent *Services* can be regarded as capabilities of the whole domain and can be shared between divisions. The connections of *Services* bridge business processes and IT solutions. Third, SOA facilitates software reuse. IT engineers, rather than developing from zero, can easily reuse the ready-made *Services* to fulfil the new requirements. In turn, reuse decreases developing costs, shortens the leading time, and reduces risk. Fourth, SOA supports incremental development with lower costs. New functions can be implemented by reassembling existing services, or they can be independently developed in parallel without affecting the use of existing functions. Finally, SOA increases operational flexibility (Hirschheim *et al.*, 2010).

Another composable design paradigm is Micro-Services Architecture (MSA) (Lewis and Flower, 2014, Newman, 2015, Nadareishvili *et al.*, 2016). In Figure 51, MSA decomposes Services into smaller granularity than SOA does. The most significant difference lies in that *Microservices* in MSA are embedded with business logic; thus, *Microservices* can communicate with each other without ESB help. Figure 52 depicts the difference between SOA and MSA. ESB, a centralised control centre, facilitates the interactions between Services in SOA. This pattern of interaction is **orchestration**. In comparison, MSA does not have a centralised element for service composition; *Microservices* exchange messages by service **choreography**.

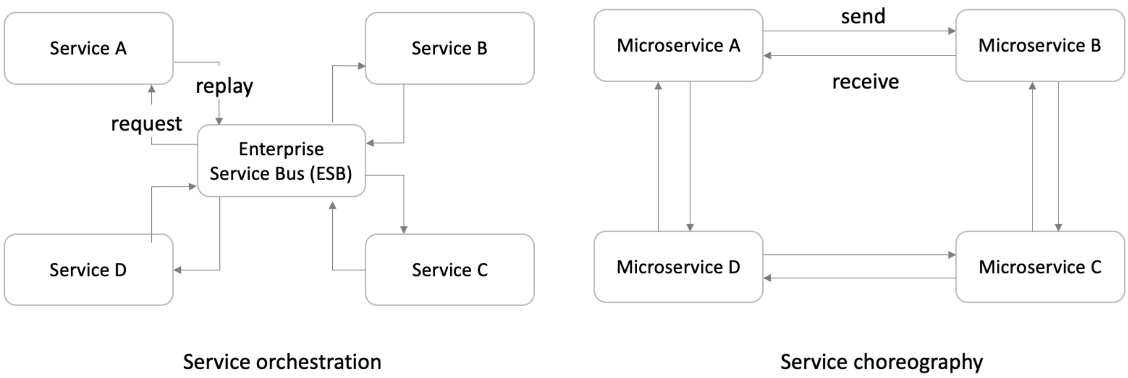


Figure 52 Services interaction in SOA and MSA

Service orchestration requires a centralised business process coordinating activities over different user requests, which could be a bottleneck in dealing with communication overhead or distributed transactions when the whole system needs to scale up (Xiao *et al.*, 2016). In SOA, ESB takes on services orchestration function and integrates business processes inside, which is the backbone for SOA. Through ESB, business logic can be flexibly reconfigured. Therefore, the integration platform of SOA is smart also complex. That is why SOA is referred to as a 'simple service and smart pipe' (Kratzke and Quint, 2017). In contrast, *Microservices* can work independently but coordinate with each other through a predefined set of cues or events. Each *Microservice* is an independent and autonomous program that can fulfil only one task. Programs can work together and have a universal interface for user invoking (Wolff, 2016). In Figure 51, each *Microservice* connects its own database, which is different from the shared database adopted by SOA. Hence, each *Microservice* can be treated as a container that encapsulates data storage technology, dependencies, and programming platform. Furthermore, a shared database in SOA means the data share the same context, while in MSA, the data employed by *Microservices* have their own context.

Compared with SOA, the vital distinctive characteristics of MSA are 1) *Microservice* is at a smaller granularity of service than the *Service* defined in SOA; 2) MSA introduces the new concept of bounded context to define the scope of where a service can be meaningfully adopted; 3) each *Microservice* is operationally independent (Dragoni et al., 2017), which means each *Microservice* can be customised, scaled, and deployed independently.

In summary, MSA is more flexible to system evolution, which is the foundation of agile development. Moreover, MSA provides a method to build distributed and independent services that interact through a shared fabric. That is why Cloud-native systems embrace MSA for constructing modern applications. Therefore, the industry is in the shift toward MSA. However, MSA is not considered the next-generation system architecture following SOA in academia. Regardless, Adrian Cockcroft at Netflix describes the MSA approach as 'fine-grained SOA' pioneering based on the web at scale; MSA is not born to address the limitations of SOA (Lewis and Flower, 2014). Both of them have their limitations. Cerny *et al.* (2017) summarise the differences and limitations in the following table.

Table 19 MSA and SOA comparation (Cerny *et al.*, 2017)

concern	MSA	SOA
Deploy	Individual service deploys	Monolithic deploy, all at once
Teams	Microservices managed by individual teams	Services, integration, and user interface man- aged by individual teams
User interface	Part of Microservice	Portal for all the services
Architecture scope	One project	The whole company/enterprise
Flexibility	Fast independent service deploy	Business process adjustment on top of services
Integration mechanism	Simple and primitive integration	Smart and complex integration mechanism
Integration technology	Heterogenous if any	Homogeneous/Single vendor
Cloud-native	Yes	No
Management/ governance	Distributed	Centralised
Data storage	Per unit	Shared
Scalability	Horizontally better scalable. Elastic	Limited compared to MSA. Bottleneck in the integration unit or a message parsing over- head. Limited elasticity.
Unit	Autonomous, un-coupled, own container, independently scalable	Shared Database, units linked to serve business processes. Loosely coupled.
Mainstream communication	Choreography	Orchestration
Fit	Medium-sized infrastructure	Large infrastructure
Service size	Fine-grained, small	Fine or coarse-grained
Versioning	Should be part of architecture, more open to changes	Maintaining multiple same services of different version
Administration level	Anarchy	Centralised
Business rules location	Particular service	Integration component