Technological University Dublin

ARROW@TU Dublin

A Two-Level Identity Model To Support Interoperability of Identity Information in Electronic Health Record Systems.

Xu Chen
Technological University Dublin, xu.chen@tudublin.ie

Recommended Citation

A Two-level Identity Model

To support interoperability of identity information in electronic health record systems

Xu Chen

This thesis is submitted for the degree of

Doctor of Philosophy

Supervisor: Dr. Damon Berry

School of Electrical and Electronic Engineering

Dublin Institute of Technology

May 2016

ABSTRACT

The sharing and retrieval of health information for an electronic health record (EHR) across distributed systems involves a range of identified entities that are possible subjects of documentation (e.g., specimen, clinical analyser). Contemporary EHR specifications limit the types of entities that can be the subject of a record to health professionals and patients, thus limiting the use of two level models in healthcare information systems that contribute information to the EHR.

The literature describes several information modelling approaches for EHRs, including so called "two level models". These models differ in the amount of structure imposed on the information to be recorded, but they generally require the health documentation process for the EHR to focus exclusively on the patient as the subject of care and this definition is often a fixed one.

In this thesis, the author introduces a new identity modelling approach to create a generalised reference model for sharing archetype-constrained identity information between diverse identity domains, models and services, while permitting reuse of published standard-based archetypes. The author evaluates its use for expressing the major types of existing demographic reference models in an extensible way, and show its application for standards-compliant two-level modelling alongside heterogeneous demographics models. This thesis demonstrates how the two-level modelling approach that is used for EHRs could be adapted and reapplied to provide a highly-flexible and expressive means for representing subjects of information in allied health settings that support the healthcare process, such as the laboratory domain.

By relying on the two level modelling approach for representing identity, the proposed design facilitates cross-referencing and disambiguation of certain demographics standards and information models. The work also demonstrates how it can also be used to represent additional clinical identified entities such as specimen and order as subjects of clinical documentation.

Declaration

I certify that this thesis which I now submit for examination for the award of PhD, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work.
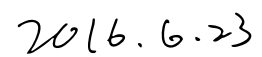
This thesis was prepared according to the regulations for postgraduate study by research of the Dublin Institute of Technology and has not been submitted in whole or in part for another award in any other third level institution.

The work reported on in this thesis conforms to the principles and requirements of the DIT's guidelines for ethics in research.

Xu Chen

Signature

2016. 6. 23

Date

I would like to dedicate this thesis to my loving parents, my wife and my son from whom I have got unconditional love and ongoing support over the years.

谨以此论文献给我深爱的父母和妻儿。

ACKNOWLEDGEMENTS

I would never have been able to complete this thesis without the guidance and persistent help of my principal supervisor Damon Berry. He was always there cheering me up and sharing his expertise and experiences in study and life.

I would like to express the appreciation to the School of Electrical and Electronic Engineering and the Graduate Research School, for their support and understanding.

I am grateful for the support of all co-authors who made contributions to my articles: William Grimson, Gaye Stephens. I would also like to thank my friends and colleagues who were willing to share their knowledge during those years in the lab: Sheng Yu, Fabrice Camous, Mark Nolan, Brian Madden, Jack Keegan, James Condron.

I would like to express my deepest gratitude to my parents, for their faith in me and providing me with unending encouragement and support.

Finally, I want to say a special thanks to my dear wife who has been doing the utmost to support me and the family as the work let us have to apart in the past several years.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF LISTINGS

ACRONYMS

ADL   Archetype Definition Language.

AHP   Allied Health Professional.

AMNCH The Adelaide & Meath Hospital Incorporating the National Children's'
           Hospital, Dublin which is located in Tallaght, South West Dublin.

ANSI   American National Standards Institute.

AOM   Archetype Object Model.

ASTM   American Society for Testing and Materials.

BFO   Basic Formal Ontology.

BRIDG Biomedical Research Integrated Domain Group.

CDA   Clinical Document Architecture.

CDISC Clinical Data Interchange Standards Consortium.

CEN   Committee European Normalisation.

CHA   Continua Health Alliance.

CIMI   Clinical Information Modelling Initiative.

EAV   Entity Attribute Value.

EAV/CR EAV With Classes and Relationships.

EIS   Entity Identification Service.

EMPI   Enterprise Master Patient Index.

ERD   Entity Relationship Diagram.

FDA   Federal Drug Administration.

FHIR	Fast Healthcare Interoperability Resources.
GEHR	Good Electronic Health Record.
GIRM	Generalised Identity Reference Model.
HCP	Health Care Professional.
HIQA	Health and Information Quality Authority of Ireland.
HIS	Hospital Information System.
HL7	Health Level Seven International.
HL7 RIM	HL7 Reference Information Model.
HL7 v2	HL7 version 2.
HL7 v3	HL7 version 3.
HSSP	Healthcare Services Specification Program.
IHE	Integrating the Healthcare Enterprise.
IHI	Individual Health Identifier.
IS	Identification Service.
ISO	International Standards Organisation.
LIMS	Laboratory Information Management System.
LIS	Laboratory Information System.
MPI	Master Person Index.
MRI	Magnetic Resonance Imaging.
MRN	Medical Record Number.
NEHTA	National E-Health Transition Authority (Australia).
NHS	National Health Service.
NPfIT	NHS National Programme for IT.
ODP	Open Distributed Processing.
OID	Object Identifier.
OMG	Object Management Group.
ORM	Object Relational Mapping.

OWL	Web Ontology Language.
PACS	Picture Archiving and Communication System.
PDS	Personal Demographics Service.
PID	Patient Identification Information.
PIDS	Person Identification Service.
PKI	Public Key Infrastructure.
PMAC	Privilege Management and Access Control.
PMI	Patient Master Index.
RBAC	Role Based Access Control.
RDF	Resource Description Framework.
REST	Representational State Transfer.
RM	Reference Model.
RM-ODP	Reference Model of Open Distributed Processing.
SDO	Standards Development Organisation.
SJH	St. James's Hospital, Dublin, Ireland which is located in Dublin, Ireland.
SOA	Service Oriented Architecture.
SOAP	Simple Object Access Protocol.
SynOD	Synapses Object Dictionary.
SynOM	Synapses Object Model.
UDI	Unique Device Identification.
UHI	Unique Health Identifier.
UML	Unified Modeling Language.
URI	Uniform Resource Identifier.
XML	Extensive Markup Language.
XSD	XML Schema Definition.
XSL	Extensible Stylesheet Language.

INTRODUCTION

1.1 Background

This work arose and was funded as part of a project called "EHR*land*", which considered issues to do with developing a regional or national electronic healthcare record (EHR) system. The author's study contributes to a key goal of the funding organisation, the Health Information and Quality Authority of Ireland (HIQA), "*has responsibility for setting standards for all aspects of health information and monitoring compliance with those standards. In addition the Authority is charged with evaluating the quality of the information available on health and social care and making recommendations in relation to improving the quality and filling in gaps where information is needed but is not currently available*" (HIQA, 2015).

More specifically, the objective of the EHR*land* project was to perform a technological assessment of the CEN/ISO 13606 standards (ISO/TC 215, 2006, 2008, 2009a,b) for exchanging partial or whole EHRs within Ireland which also has more general applicability to the international health informatics community. Taking slightly informal definition, the EHR is "an amalgamation of all the useful clinical information that has been collected and stored by different people in different locations about a patient from the 'cradle to the grave'" (ISO/TC 215, 2006).

The EHR*land* project is focused on three distinct implementation issues: architecture for managing clinical data, terminology of the EHR and identity of all entities (such as patients and healthcare professionals) in EHR workflows. The aims of this work arose out of the EHR*land* objective to investigate the two-level identity modelling approach associated with application in the clinical laboratory domain.

1.2 EHR and e-Health technologies

In modern health informatics, the EHR is widely used by professionals and non-professionals. Nevertheless, there is no single comprehensive definition for the EHR. For the purposes of this thesis, the author will use what is arguably the most authoritative definition of EHR, from ISO health informatics technical committee (ISO/TC 215, 2005) which defines an EHR (also called the "basic-generic EHR" but for the purposes of this work simply "EHR") as:

"*A repository of information regarding the health of a **subject of care** in computer processable form, stored and transmitted securely, and accessible by multiple authorised users. It has a commonly agreed logical information model which is independent of EHR systems. Its primary purpose is the support of continuing, efficient and quality integrated health care and it contains information which is retrospective, concurrent and prospective.*"

This direct quotation illustrates the standards-based consensus around the primary subject of information of an EHR - the "subject of care", which is fundamental to understanding an EHR or the use of an EHR. Clearly, an EHR is a document about a "patient". However, there is a rapidly growing literature on the secondary use of the EHR, which indicates that the secondary use of the EHR is increasingly demanded in all kind of the health domain (Botsis *et al.*, 2010). In the domains of care research, clinical research and clinical trials, research has provided ample support for the assertion that the secondary use

of the EHR can also be of benefit to patient care or to develop a better application in the domain of health (Powell and Buchan, 2005).

Despite the exclusion of secondary use from user requirements, most of the standard development organisations who are interested in the EHR (e.g., ISO, HL7, OMG, openEHR) support use of both EHR data, that is captured for the purposes of care and clinical data that is gathered for the purposes of research within their published EHR standards or specifications. The literature demonstrates that a more flexible definition of the subject of information in the health domain would be beneficial. But this more flexible approach has not been supported or recognised by the EHR standard designers. On these grounds, the author argues that the definition of EHR is severely limited in its connection to only certain prescribed physical objects such as patient, health professional, software and device.

With particular reference to Popper's "Three Worlds" (Agassi, 1974) (which will be discussed in more detail Chapter Two), the EHR is a product of the human mind. It after all is a kind of the electronic medical documentation and was recorded in paper-based form originally. The physical form of the EHR (e.g., data storage device, computer memory) and paper records are part of Popper's "World 1" of physical objects as shown in Figure 1.1. Both the paper-based documentation and the electronic documentation are used to record, store and share the health information as regards the entities within "World 1" and "World 2" such as person, sample, device and mental objects such as episode, order and process. This philosophical view of the EHR also lies at the centre of discussion on the definition of identity entity in Chapter Two.

In the health informatics domain, arguably the principal aim of interoperability between information systems is to towards a shared EHR (Garde *et al.*, 2007a,b). Shared identity information is fundamental to achieving this interoperability. At present it is difficult or even not impossible to have an EHR, EHR systems or EHR services without sharing of identities of healthcare resources or persons (such as patients, healthcare professionals, samples, pharmaceuticals or blood products) either through some forms of federation or
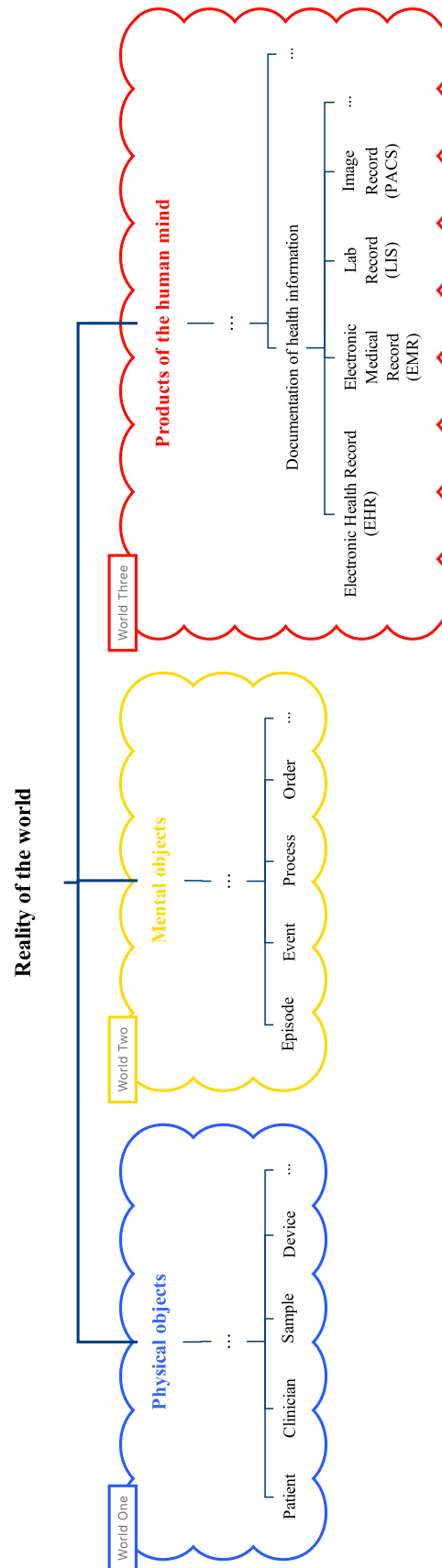
Figure. 1.1 Popper's reality of the world with a particular focus on health documentation

centralisation. Therefore a growing need for the sharing of identities has arisen and it has become the part of health informatics strategy in many countries.

1.3 Identity in healthcare

Identity[1] and its "close cousin" demographics, is the key to accurate identification of all actors and resources in healthcare. This in turn helps to improve the quality of health data, reduce costs, medical errors and inefficiencies across all clinical domains and applications (Kwak, 2005). In the context of this thesis, the management of identity as used, encompasses the concepts of identification and either identity management applications or services.

With the evolution of the discipline of health informatics, there has been a drive to leverage information technology to deliver high quality and cost effective health care, leading to increased productivity and enhanced patient safety (Yasnoff *et al.*, 2000). The effective and efficient exchange of healthcare information to support shared care depends upon rapid, usable and accurate use of identity with the EHR and this will not be implemented with efficiency unless there is a shared and secure identity management system. During the process of sharing and retrieving EHR, there are a number of types of identities in different formats in different health systems, domains, organisations, geographical locations and this heterogeneity makes the identification process difficult. It should be noted that although security and privacy are topics that are often grouped with identity, these other topics are not the primary focus of this work. However in the author's opinion, a robust enterprise wide identity management architecture can strengthen the security of an EHR system, and this study also contributes to that goal.

The exchange of health information within and between health enterprises has long been problematic. Today in many countries, the absence of a national ID for healthcare, (called a

---

[1]Identity in this thesis refers in particular to digital identity

Individual Health Identifier (IHI) in Ireland), has meant that healthcare organisations must develop their own identification systems and separate identification domains (Forslund *et al.*, 2000). Many of these systems associate the same or a similar set of trait attributes (collectively called a trait set) with each identity to help the identification process (e.g., patient name, date of birth) but the information may not be stored in identical formats or with completely identical trait sets at participating healthcare sites.

Each identity is also typically associated with one or more numerical or alpha-numerical identifiers which are called different names (e.g., Medical Record Number, Patient ID, and Subject of Care ID) in different organisations. In order to allow identities from the numerous identity domains to be matched, traits are compared and the pairs of identifiers are stored together for future use but this approach does not scale very well. In the worst case scenario, the resulting matching-and-mapping problem would need to be solved for every pair of sites, resulting in the telecommunications network's $C = \frac{n(n-1)}{2}$ topology (or connection) problem (Shaw and Mahoney, 2000) in particular, which is proportional to n^2 asymptotically. This situation is simplified if each site can refer to the above-mentioned individual health identifier in a national identifier domain. By simplifying the process of linking identities at different health sites, unique national identifiers facilitate the integration of health information. The resulting multisite access to the EHR and other health information represents an important enhancement of healthcare quality and a major step towards a regional EHR system.

1.4 Research problems and challenges

The following subsections summarise a number of problems and challenges in healthcare that are affected by the identity models and the identity management system.

1.4.1 Identity Crisis

The healthcare IT applications that rely on identity management services do not scale well because the existing identity management architectures do not support the seamless integration of semantic interoperability and compatibility for enabling EHR standards. For instance, in relation to the legacy Hospital Information System (HIS), the applications are difficult to be integrated to support enabling identity standards. On the other hand, the existing identity management architectures do not support well neither the integration of legacy system and EHR standards or semantic techniques with two-level methodology to facilitate the issues such as identification, identity matching, and identity sharing (ISO/TC 215, 2006).

Patient fatalities due to misidentification

In the United States, fatal misidentification errors are estimated to occur in 1 in 600000 to 1 in 800000 transfusions and non-fatal errors occur in 1 in 12,000 to 1 in 19,000 cases. A study points out that while misidentification is a relatively uncommon cause of death and injury, it may have a considerable impact on availability of blood for transfusion (Regan and Taylor, 2002).

Elsewhere in the literature, patient misidentification was cited in more than 100 individual root cause analyses by the United States Department of Veterans Affairs (VA) National Centre for Patient Safety from January 2000 to March 2003 (Mannos, 2003). According to an internal VA study in the USA, patient misidentification results in inappropriate diagnosis, treatment or surgery.

Patient information is missing due to errors of patients identity matching

In a study which is closer to the core topic of this work, Health Corporation RAND examined how frequently patient identity errors occur during statistical matching. It

reported that, 8 percent of the time, some key health information about a patient will be missing when an exact match is required (Hillestad *et al.*, 2008).

Massive and unnecessary duplicate identities

Initiate Systems (IBM, 2009) reports that, in an examination of six Master Patient Index (MPI) systems, there were more than 150,000 duplicate records. The duplication rate of identities was 8 percent. Larger files of 1 million or more records had a mean duplication rate of 9.4 percent, whereas those with less than 500,000 records had a mean error rate of 7.2 percent.

Given the above statistics it is reasonable to assume that clinicians are presented with a list of laboratory results for patients that could be the same. There is also a likelihood of unnecessary laboratory tests that are re-ordered for patients with a resulting waste of time and resources.

High cost identification caused by no unique identifier

Health Affairs (Hillestad *et al.*, 2005) stated that a connected EHR system in U.S.A would result in a potential efficiency saving of $77 billion per year at the 90-percent level of adoption with additional safety and health values that could double these benefits. By comparison a one-time cost of $1.5 to $11.1 billion is associated with the introduction of a unique patient identifier (UPI - the USA's equivalent of the Irish IHI), to remove the systemic errors in health.

Implementation and management of a voluntary UPI system for USA as proposed by the ASTM would (they estimate) have a cumulative cost of $25 million for the first five years for a national organisation issuing UPIs (Hillestad *et al.*, 2005). Without a national UPI, patients must be re-enrolled annually, at a cost of $1.5 billion per year; assuming that for 300 million people, it would cost $1 for the 5 minutes that it takes to manage each identity per year.

1.4.2   Lack of interoperability for sharing identities

During the process of sharing and retrieving an Electronic Health Record (EHR) across distributed systems, a range of types of identities are encountered, often in different formats, in different health systems, domains and geographical locations. This heterogeneity can make the identity management process difficult and insecure. For instance, identities are not sharable or recognisable if they are based on different existing demographic models in information systems that contribute to an EHR system.

A investigation on the status of interoperability in Germany, Spain and the UK indicates that the most important fact in determining the demand of the EHR system is to facilitate the data sharing among associated care providers (HIMMS, 2014). The Master Patient Index (MPI) is the top technical capability needed for the interoperable health system in the UK. This is likely that most people in the UK have been signed a unique identifier by the NHS but most hospitals will still issue a their own identifiers for each patient that could cause the identification error between the various health information systems (HIMMS, 2014).

1.4.3   Limitation of the secondary use of identity

The main existing demographic models support only certain types of entities that are limited to represent only some types of the identities in an EHR system. The subjects of information (or identified entities which will be described in Chapter Two) within the clinical data are not explicitly supported in EHR standards apart from the core subjects; patients and health professionals. In the author's view, the EHR should be allowed to represent other identity perspectives and types of identities. For instance, it should better support entities that are of interest for secondary use of the EHR or entities such as sample, medical device, which are used in the laboratory domain.

A study focusing on leading causes of identification errors, analysed the potential adverse consequences. While it is assumed that identification errors occur at a low frequency in clinical laboratories, misidentification of general laboratory specimens is around 1% and can produce serious harm to patients, when not promptly detected (Lippi *et al.*, 2009). What is not entirely clear from these studies is which point in the identity process causes the errors. They could be attributed to human factors, the use of technology such as bar codes and then the identity management systems, but it is likely that some portion of the errors occur within the secondary use of identities in the health domain.

In drugs and biological products, unique identifier formats such as bar code for improving the quality of medical care of patients have many potential benefits. Research prepared for the USA's Food and Drug Administration (FDA) stated that the number and complexity of medical devices is increasing. It is important that FDA quickly identifies new risks and works with companies and with medical device users to develop appropriate control of these risks. A unique device identifiers (UDI) system may be used to identify the entire life cycle of equipment as well as with the device-related information. For example, a UDI could identify which devices are compatible, such as being implanted device is working with magnetic resonance imaging (MRI) system with safety (Eastern Research Group, 2006). So the role of identity management in health is not restricted to identification of people.

1.5   Aims and objectives

On the basis of the above investigations, it seems appropriate at this point to clarify some theoretical background with examples of the relationship between reality and the representation of identity in EHRs as shown in Figure 1.2. In Figure 1.2a, the red box denotes the hierarchical relationship based upon the definition of documentation of health information as a product of human mind. The identity (refers to identity information) is

a specific entity of interest within the author's work, which corresponds to the identified entities in the "World 3". The blue curves denote the relationship between the identified entities and the physical objects or the mental objects . The enlarged portion of the "World 3" documentation of health information in Figure 1.2b delineates the boundaries of this study's subject that focuses on two-level identity modelling and integrating identities in the documentation of health information.

As a successor to the author's work in the EHR*land* project, this work will mostly rely on the various standards in the health domain (e.g., ISO/EN13606, HL7) and/or information models which were provided by Standards Development Organisations (SDOs) or third parties such as CEN, ISO, OMG. The identity management architecture that is developed to support this thesis, and the associated concepts and activities such as identity modelling, matching, identification in different contexts or across disciplines have different meanings. Camp (Camp, 2004) noted that "*the word 'identity' refers to an increasing range of possible nouns - from a set of personality-defining traits to a hashed computer password*". In the author's opinion, identity management architecture in healthcare is composed of a selection of information technologies and computerised processes to manage and securely access EHR information and resources from different organisations, domains or geographic locations. Identity management architecture should also address the problems associated with overloading of identity terminology and "*the lack of conceptual clarity*" (Camp, 2004).

A primary goal of the proposed work is to investigate the characterisation of a selection of existing demographics information models from health informatics standards with a view of modelling these features in a generalised and archetype-able identity model. The proposed work seeks to preserve the core status of the patient – health provider relationship while facilitating other types of identity. It would also seek to decompose the patient-centred record into records that focus on other subjects of information such as specimen, or individual body parts. Therefore, the image that such entities are viewed equally as
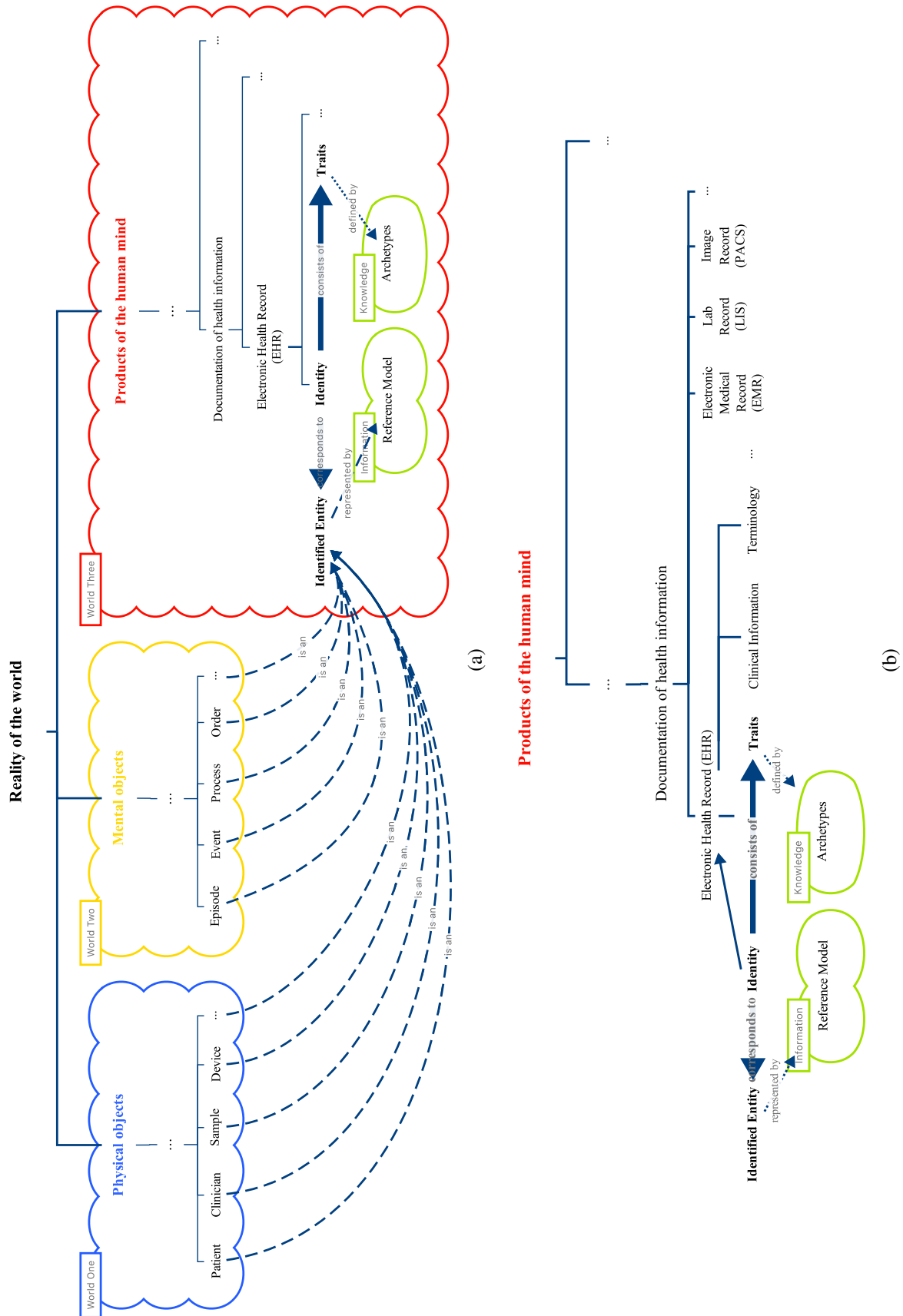
Figure. 1.2 The relationship between reality and the representation of identity in EHRs

persons in this work. These secondary identities may also be of specific interest in certain clinical studies.

To determine the extent to which a novel (generalised and extendable) approach based on a two-level identity model facilitates improved, effective and appropriate (e.g., minimal or no logical / physical redesign, expert controlled) integration of identity models into EHR systems, the author will work towards the following project objectives[2].

O.1 investigate the two-level identity modelling approach to enable interoperability of identity information between EHR systems and other health information systems, and show the validity of using it to represent regional demographic and other identified entities in health domain.

O.2 create a generalised identity reference model for sharing archetype-constrained identity information between health information systems using diverse identities, models and services, while permitting reuse of relevant published standard-based archetypes.

O.3 apply the approach to express details of identity models to show its use for standards-based two-level modelling alongside legacy systems. Unlike existing approaches, this approach should not be specific to demographics but can also express other identity types in healthcare and in other domains.

O.4 demonstrate the benefits of using this two-level identity modelling approach the clinical models for EHRs that could be adapted and reapplied to ease identity-related integration problems between laboratory information system and EHR system.

1.6   Intended contributions of this work

The following contributions[3] are attributed to this work.

---

[2]The symbol "O" denotes "Objective".
[3]The symbol "C" denotes "Contribution".

C.1 This work will explore the state of the art in shared identity model architectures that provide a combination of processes and technologies to share identity information in the context of large scale national EHR systems and accompanying information systems.

C.2 The author will compare and analyse the features of the existing approaches to demographics and identity by employing a categorisation approach to different types of the identity models. The analysis in this work will provide evidence in support of the dynamic two-level modelling approach to improve efficiency, interoperability and expandability based on existing unstructured or single model based approaches in e-Health.

C.3 This thesis will describe the development of an approach to create a generalised identity reference model for sharing archetype-constrained identity information between diverse identities, models and services using two-level models within the health community.

C.4 The author will present an integrated approach to enhance the patient-centred record and to also enable records that focus on other subjects of medical information such as samples, or individual body parts so that these secondary identities can be used for certain clinical studies.

   (a) This approach will allow EHR systems to utilise two level models for identifying the multiple demographics entities in the different specification perspectives. This should make system integration process easier and more streamlined through fewer transformations and potentially safer too.

C.5 The work will support ease of evolution to two level models for laboratory information systems as well as EHR systems. Secondary use of more detailed feeder system data (i.e. LIS data) could benefit from the data quality enhancement that comes with

a harmonised but also multi-view approach. Secondary use could also benefit from this seamless link between record and specimen or order views.

1.7 Organisation of the chapters in this thesis

The rest of this thesis is organised as follows.

Chapter 2: *Background* gives the key terms and definitions with regard to identity in EHR. The chapter surveys the computational modelling approaches, some representative function models and interoperability standards in healthcare to account for the evolution of shared identity models and relevant information technologies.

Chapter 3: *State of the art for integrating shared identity models with EHRs* focuses solely on the integration approaches for sharing identity information with further literature review. The chapter describes the central concepts of the ontology-based identity models and federation of identity for archiving semantic interoperability of shared identity data underlying the work. It also reviews relevant research and the tools related to identity modelling in the health domain

Chapter 4: *A generalised identity reference model* introduces a generalised modelling approach (called generalised identity reference model, GIRM) for shared identity. The chapter describes the need, construction and derivation of the GIRM. In this chapter, the characteristics under which the GIRM is designed to support exchanging archetype-constrained identity information between diverse identity domains are described. It also describes the evaluation strategy of the GIRM and the

Chapter 5: *Evaluating the generalised RM approach* describes the evaluation of GIRM which demonstrate the feasibility of applying the GIRM to extending the existing demographics models. The three components of the GIRM representation will be described in this chapter, and an example of abstraction and constraint analysis of these models then illustrate the use of the GIRM. The chapter also demonstrates the evaluation process

of archetypes creation which conform to both the GIRM and the chosen demographics models.

Chapter 6: *Application of the GIRM for identifying laboratory information* present an investigative study that applies the GIRM to represent clinical identity information in the laboratory domain as well as discussion of its impact on existing modelling approach. The further evaluation of implementing the two-level identity model to integrate the shared identity and health information is described in this chapter. The last section summarises and discusses the results of these consequences as well as the limitations of the work.

Chapter 7: *Discussion and conclusions* starts by reviewing the backgrounds to the work, then describes the details of each research contribution in this thesis. The chapter summarises the whole work and discusses the research implications of this study and the possibilities for future research consideration.

Background

2.1 Introduction to this chapter

This work focuses on existing approaches to identity management in heterogeneous health-care information environments. It has already been noted that identity management is applied to associate documentation (e.g., EHR, EPR) with a subject of documentation (e.g., care, laboratory, clinical trial) and the responsible entities (e.g., patients, samples) and authors (e.g., HCPs, instruments) and also to provide secure and reliable access to documentation.

As the reader shall observe at some points in this chapter, healthcare documentation presents a particular set of challenges and leads to interesting and sometimes complex solutions. In the author's view, complexity also arises in distributed identity management in the healthcare domain. This chapter will provide background to this work by describing the evolution of many approaches and points of view that have been combined to form the generalised identity model that this work proposes and more importantly the associated way of thinking about identity.

This chapter introduces the research background about:

1. the historical viewpoints of identity in EHR and the key terms defined by the author,

2. the evolution of shared identity models and existing approach of interoperability standards and specifications,

3. the related work (main existing applications, security issues, IT technologies and tools used in the author's work) that impacts on the thesis.

2.1.1   Standards development organisations in e-Health

Before going further into the details of identity and information models, this section describes some of the main standards development organisations (SDOs) who work in the area of health informatics, SDO is a generic term that refers to a large number of standard organisations, who are engaged in developing and releasing industry-specific standards. The following SDOs, in particular, have an impact on this work and these acronyms are commonly used in the entire thesis.

The main standards development organisations working on identity management in health:

ISO/TC 215 is one of Health Informatics technical committee of ISO, which standardise the field of health information (ISO/TC 215, 1998). ISO consists of 225 technical committees (TC); the TC 215 is responsible for the field of Health Informatics of the Technical Committee (ISO/TC 215, 1998). The scope of its functions is to standardise the area of health information, information and communication technology (ICT) (Klein, 2002). The ISO/TC 215 committee is divided into six working groups (ISO/TC 215, 1998):

- Working Group 1 - Health Records and Modelling Coordination

- Working Group 2 - Messaging and Communication

- Working Group 3 - Health Concept Representation

- Working Group 4 - Security

- Working Group 5 - Health Cards

- Working Group 6 - E-pharmacy and medicines business

Furthermore, the ISO/TC 215 also supports an Advisory Group and three special groups, which are: Consumer Policies, Mobile Health, and Web Applications.

CEN/TC 251 is the main Medical Informatics technical committees of the European Commission Technical Committee of Standardisation (CEN) that includes the working groups in fields such as medical record models, terminology, codes, semantics and knowledge bases (CEN/TC 251, 2012). CEN consists of fifty five subcommittees of the European standards body, including a specialised technical committees in the medical field: TC 251 (Medical Informatics) (CEN/TC 251, 2012). The Medical Informatics CEN TC 251 technical committee includes the following working groups (Klein, 2002):

- medical record model;

- terminology, codes, semantics and knowledge base;

- communication and information;

- multimedia and imaging;

- medical equipment;

- security, privacy, quality and safety.

OMG is an international non-profit membership organisation, which produces a large number of specifications focusing on health informatics such PIDS, IS and AML. Object Management Group (OMG) (OMG, 1989) has about 300 of its international alliance, which developed the Object Management Architecture (OMA). OMA is a description of OMG definition of object-oriented applications and environments produced by the standard model. The Unified Modelling Language (UML) is one of the specifications that are developed by the OMG. An OMG Special Interest Group called the Health DTF worked on health related specifications for many years . The OMG Health DTF has published many

formal specifications on identity such as Person Identity Cross-Reference Service (PIDS) specification and HSSP (will be described in Section 2.3.5) Identity Cross-Reference Service (IXS) specification.

HL7 is a Health Information Exchange Standard institution that was accredited by American National Standards Institute (ANSI) as an SDO which mostly publishes standards for the healthcare domains such as pharmacy, medical devices, imaging (HL7, 2015). HL7 focuses on layer 7 of the ISO OSI seven layer model. Hence the name Health Level 7. The HL7 standards bring applications together from different vendors for the design of interfaces within a standard format; this family of standards allow the various medical institutions with heterogeneous systems, to data exchange. Since 1987, as a standard body, HL7 have released its standard version 1.0 and then released version v2.0, v2.1, v2.2, v2.3, v2.3.1. In 2000, the v2.4 version was released and v3 .0 version is of the official release version of the current ANSI since 2011 (Beeler, 1998b).

HSSP is a joint project with the collaborations between HL7 and OMG to support healthcare IT. The HSSP was organised in January 2005 under the meeting of HL7 EHR Technical Committee (HSSP, 2005). It provides the technical specifications based on Service Oriented Architecture (SOA) solution in healthcare, such as Identity Cross-Reference Service and Retrieve, Locate, Update Service and Decision Support Service (HSSP, 2012).

IHE is an organisation of the Integration healthcare Enterprise of US who promotes healthcare interoperability by identifying practical interoperability use cases in healthcare and developing practical profiles and guidance about implementing e-Health interoperability standards. They have developed the Patient Identifier Cross-referencing (PIX) and Patient Demographics Query (PDQ) profiles for patient identification beyond IHE IT Infrastructure Technical Framework (which will be discussed in Section 2.3.4) (IHE, 1998).

The Clinical Data Interchange Standards Consortium (CDISC) is a global non-profit standards organisation and initially founded by many multinational companies in 2000. CDISC aims to establish a series of standards to support and improve clinical data interchange (CDISC, 2000).

Continua Health Alliance (CHA) is initiative and established by a large number of different medical technology companies and many other co-sponsored manufacturers such as Intel, IBM, Philips in 2006 (CHA, 2006). The CHA actively promotes a unified connection through health technology standards to ensure that certified medical electronic devices or products able to successfully communicate and are generally compatible.

Also, these SDOs there are a number of community efforts from either the open source community e.g., openEHR (openEHR, 1999)) and CIMI (CIMI, 2011). These will be discussed in Chapter Three.

2.2 Key terms defined about identities in EHR

2.2.1 The idea of identity in information systems

The essence of Identity Management as a generalised domain-independent solution is to provide a combination of *"processes and technologies to manage and secure access to the information and resources"* of within an organisation, domain or agency while also protecting the user's profiles (Reed, 2002). However, the identity management system within healthcare is considered to be *"a low-level mechanism that enables users in a particular system or region"* to securely access information resources (Camp, 2004). Arguably the most prominent of these information resources is the electronic health record that was introduced briefly in Chapter One - which holds and shares information recorded about the health, and the care of a patient throughout their lifetime. The patient is the primary subject of information of the EHR but the patient is not the only subject in the healthcare domain. For example, in Figure 2.1, many types of identities such as

sample, specimen, device and order need to be managed in a healthcare information system environment.
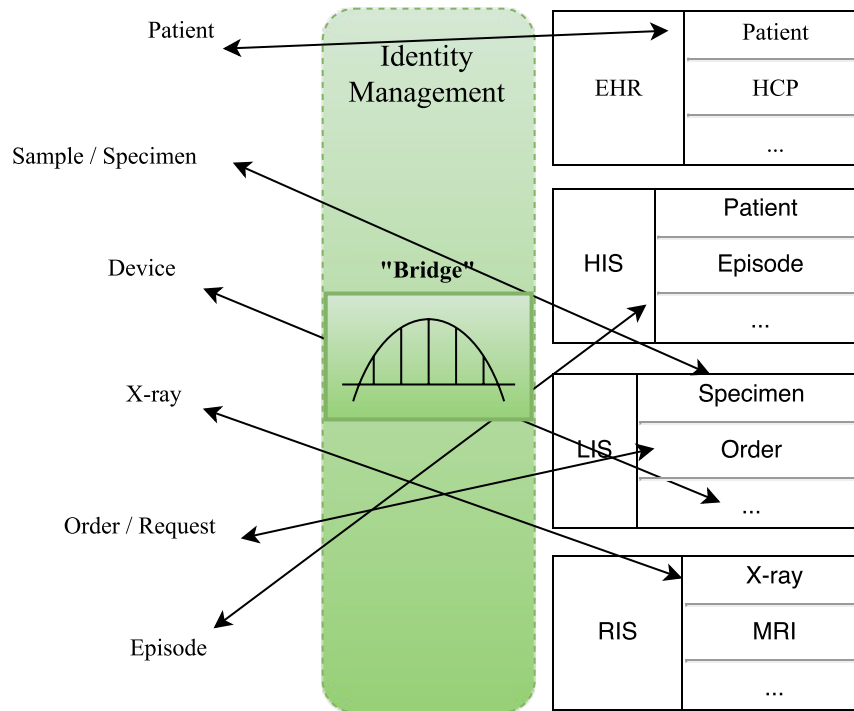
Figure. 2.1 The key to "bridge" between real world entities and information systems

Identity Management in the health domain is the "bridge" shown in Figure 2.1 between the real world and documentation. More specifically, it is the key to accurate identification of patients, health professionals and other participants in healthcare. In the healthcare domain, the identifying information of healthcare professionals, EHR users and patients, whose EHR is the accessed resource, are subject to change. A versatile model that can respond to these changes would improve the quality of health data across all clinical domains and applications. For examining these arguments, it would be desirable to interpret the key concept of an "identified entity" which refers to the existence of an identifiable thing in the health domain and to distinguish between the identity domains.

2.2.2   Identified Entity

The scope of the generic EHR is to document, archive and exchange record of care by one clinician about one patient using the computerised system. In the literature, EHR scope of ISO/TC 215, as the following quotations show (ISO/TC 215, 2006):

*"This standard is for the communication of part or all of the electronic health record (EHR) of **a single identified subject of care** between EHR systems, or between EHR systems and a centralised EHR data repository. It may also be used for EHR communication between an EHR system or repository and clinical applications or middleware components (such as decision support components) that **need to access or provide EHR data**. This standard will predominantly be used to support the direct care given to identifiable individuals, or to support population monitoring systems such as disease registries and public health surveillance."*

In the above passage, the emboldened texts illustrate two sides of an argument about the scope of the EHR. The first focuses on the patient as a subject of information. The second emboldened text concedes that EHR systems do not exist in a vacuum and often take information from other systems that do not have patients as primary subject of information.

Furthermore, with the main purpose of the EHR, the focus on either identified subject of care or identifiable individuals is limiting for interoperability and development of health information interoperability and interfaces. The concept of identified entity and the relationship should be categorised and expressed to reflect the scope of the EHR. To look at the role of identity management in the electronic health record from a different angle, let us take a higher-level view.

In 1972, Popper (Agassi, 1974) proposed his central philosophical doctrine "Three Worlds" theory which was briefly mentioned in Chapter One. In Popper's theory as Figure 2.2 shows:

1. The physical world, referred to as "*World 1*" includes physical objects and the state.

2. The spiritual world, called "*World 2*" includes psychological, state of consciousness (unconscious, subjective experience).

3. "*World 3*" refers to the product of human mental activity, including objective knowledge and objective works of art.

Reality of the world
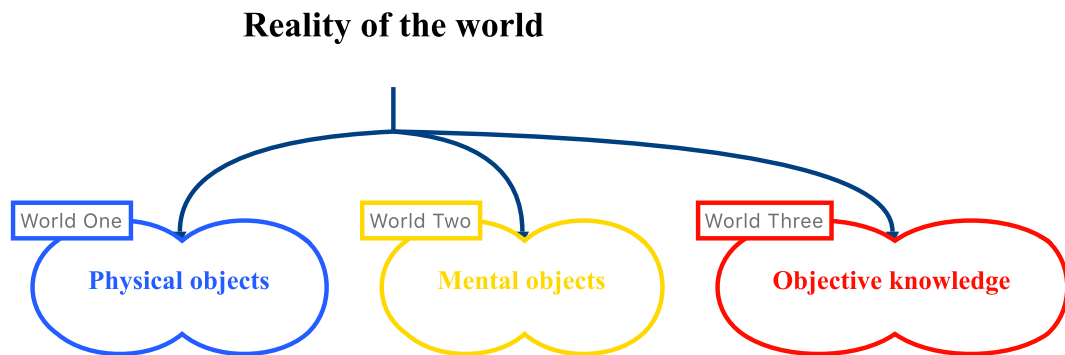
Figure. 2.2 Popper's idea of "Three worlds"

Popper believes that the three worlds he divided are "real" and "World 1" has produced "World 2" and "World 3". For example, the electronic health record (EHR) is an example object that belongs to "World 3" but the physical manifestation of that EHR also belong to the material world that is the "World 1" based on the different views. An example of this is the corresponding blips of light in an optical fibre that is communicating an EHR extract or magnetised regions on a hard disc platter that is storing it. Popper also believes that entities from "World 1" and "World 3" do not interact directly, but only through "World 2" the intermediately indirect interactions. Another example of this is the interaction of the brain ("World 1") and language ("World 3") in between. They interact through "World 2" (human consciousness), with results that have not only contributed to the evolution of the brain but also to promote the development of language. Popper believes in this critical sense, the development of scientific knowledge that is achieved through the interaction of three worlds.

An advantage of "Three Worlds" is to distinguish the existence of direct and indirect matter forms from two points of view, which are objective world and the subjective world, also known as the material world and the spiritual world (Agassi, 1974). Early 1980s, information scientists Brookes (Brookes, 1975, 1980) introduced the idea of "Three Worlds" into library science and information science research in the field, which has prolonged attention and aroused fierce controversy but Popper's idea has not been widely adopted in health information literature. Moehr (Moehr, 1994), Abbott (Abbott, 1997) and Urquhart (Urquhart, 1998) may be the earliest researchers using the "Three Worlds" on related topics. The "Three Worlds" as presented by Brookes was expanded further by Moehr (Moehr, 1994) and Bawden (Bawden, 2002) to recognise the complexity of health information. Bawden (Bawden, 2002) also provides the value angle and understanding of health information and knowledge in general.

As shown in Figure 2.3 concerning Popper's view, one role of identity management is to provide a user-accessible "bridge" between world three (documentation) and world one and two (subjects of information). This thesis will also adopt the similar mechanism inherent in interpretation of "Three Worlds" and apply it to define the identified entities in the domain of health information science apart from generalised information science (or philosophy of information).

Elsewhere in the literature, BFO (Basic Formal Ontology) or the upper ontology (Smith and Grenon, 2006) is increasingly being used in the field of information systems. The BFO approach implicitly recognises the need for an identified entity as a bridge between "reality" and its representation for information, documentation and knowledge. The BFO shown in Figure 2.4 was developed by Barry Smith and his associates in 2002. It partially adopted some idea of Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) shown in Figure 2.5 (Masolo *et al.*, 2002) and Suggested Upper Merged Ontology (SUMO) shown in Figure 2.6 (Masolo *et al.*, 2003). In the BFO, the top-level body framework consists of a series of sub-ontologies for specific domain perspectives such as open
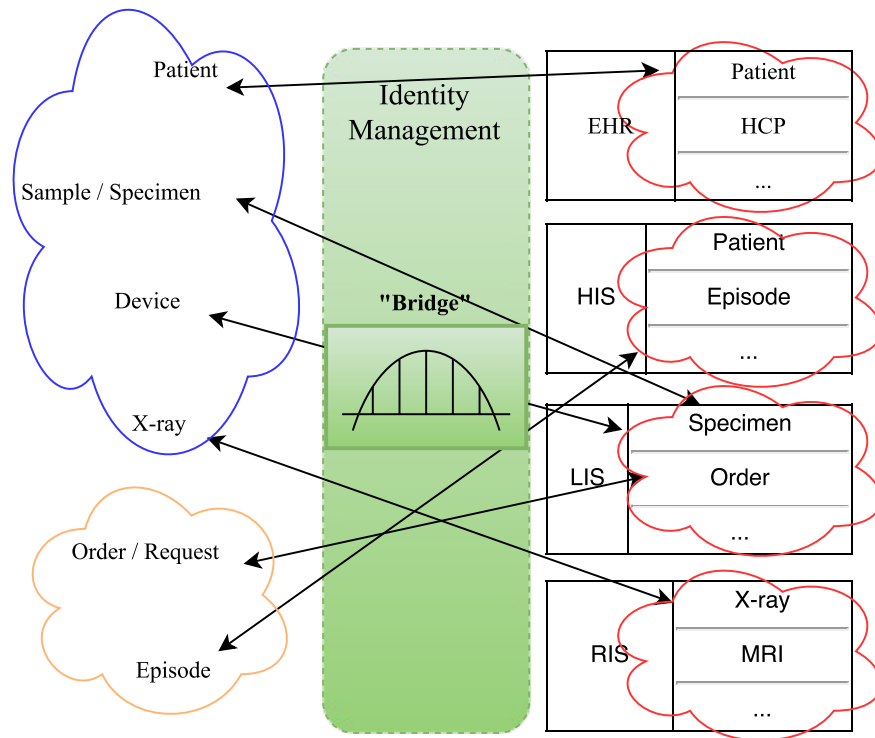
Figure. 2.3 Adoption of "Three worlds" between subjects of information and documentation

biological ontologies Foundry (OBO Foundry). The general idea of BFO is partly derived from Zemach's Four Ontologies (Zemach, 1970), which is effectively a classification between two central categories of entities, *"Continuant"* and *""Occurrent"* shown in Figure 2.4 (Grenon and Smith, 2004). According to Smith, continuant refers to the entities that cannot themselves change with the passage of time such as data, agent, and physical material. On the contrary, an occurrent entity can change itself over time such as process of injections and transplanting organs.

The BFO allows continuants not only to be physical things such as patient, health records or medical devices but also things such as qualities, processes or events (Smith, 2012). Referents, the term used by Smith (Ceusters and Smith, 2005, 2006) also refers to all identified entities including continuants and occurrents such as patients, diseases and diagnoses. In agreement with the discussion earlier in this chapter, a case is made by Smith et al. for tracking more referents than just patients, health care professionals,
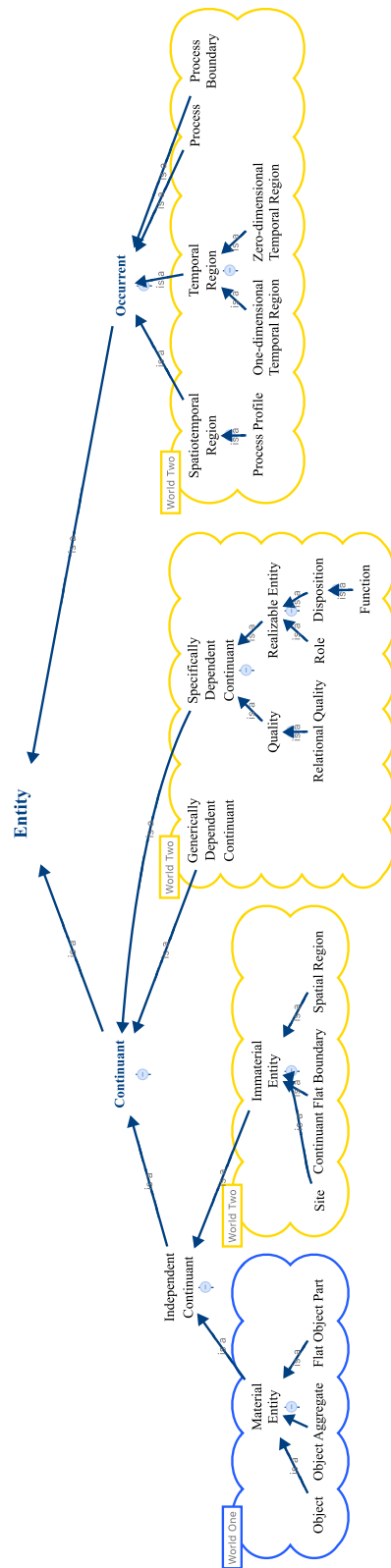
Figure. 2.4 Taxonomy of basic categories in BFO hierarchy
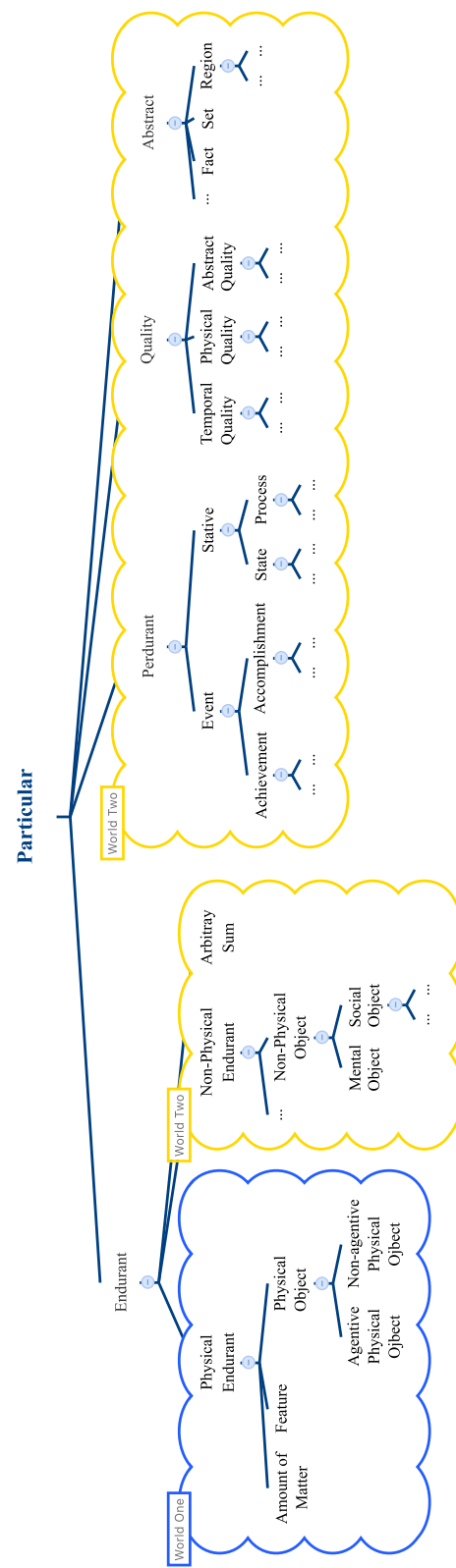
Figure. 2.5 Taxonomy of basic categories in DOLCE
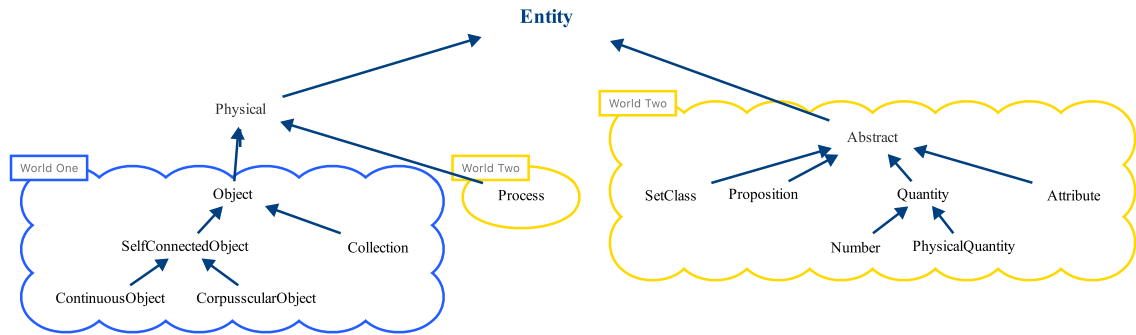
Figure. 2.6 Taxonomy of basic categories in SUMO

and organisations. Ceusters and Smith, have suggested in some publications that it would be beneficial to incorporate other entity types, such as disorders, body parts, treatments, and to make these entities identifiable in the EHR (Ceusters and Smith, 2006). A case is being made for tracking referents, a term which can be applied to participants or actors in an EHR. It will be shown later on that, this point of view is compatible with the central argument of this work.

As Smith and Ceusters argue, the ability to identify entities relevant to the health of the patient helps to keep track of their evolution, their disappearance, and also the appearance of new entities (such as polyps, fractures, transplanted organs or bone components). This precise documentation of parts of the patient is ever more relevant in the context of distributed care where patients will receive care from different health professionals at various locations and times. Requirements for referent tracking have already been published (Ceusters and Smith, 2006) and a system has been implemented (Manzorr *et al.*, 2007). The focus is not only on the allocation of universally unique identifiers (UUID but also on actually keeping track of the entities and their relationships in time. The referent tracking system (RTS) can be seen as an identification service for all entities that are referenced in the EHR. As it is a fore-runner of this work, such entities can also be identified in other health information systems such as laboratory information system (LIS), or picture archiving and communication system (PACS).

Recorded Information

World Three

care
information

admin
information

history

option

instruction

observation   action

assessment

proposal

intervention
request

investigation
request

diagnosis   risk   prognosis   recommendation   scenario   goal
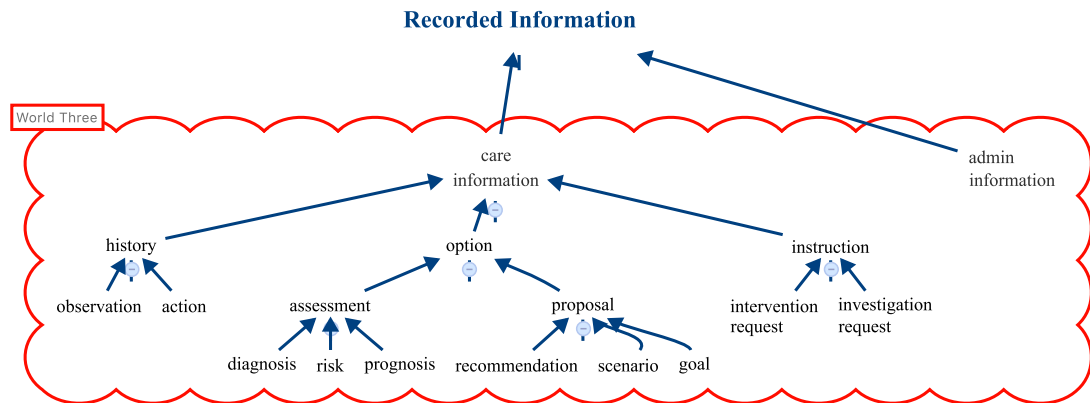
Figure. 2.7 OpenEHR Ontology of Recorded Information

Recent representation models of the EHR or EHR extract communication, like openEHR (which will be discussed in Section 2.3.3) ontology of recorded information (shown in Figure 2.7) look differently at the idea of recording information that describes the structure of reality in healthcare (Beale *et al.*, 2006). Systems designed using these specifications model documentation rather than modelling the real world entities. The openEHR approach allows the domain experts to configure and build the shape and content of the EHR, not to provide precision in the definition of the ontology. They can use a language to assemble and constrain generic EHR blocks. A particular configuration of generic EHR blocks constitutes what is called an archetype which will be described in more detail later in this chapter, and contains instructions on, for example, the data capture of a blood pressure measurement.

Despite this flexibility in describing health information, the main EHR communications standard (first introduced in Chapter One and described in more detail in Chapter Three) CEN/ISO EN13606 does not allow this level of flexibility when it comes to entering data about identified entities. In line with the focus on the patient that was referred to in Section 2.4.2, the entities can only be of three main types: organisations, persons, or software/device. In order to allow health professionals to determine the data they need to identify all the entities they wish to refer to in the EHR, generic identity trait "blocks"

29

(similarly to EHR "blocks": folders, sections, entries) need to be available to be added where required into identification archetypes which describe the data required for the identification of a particular entity.

To provide another example of the wide range of identifiable subjects of information from the healthcare domain, the following modified list within Biomedical Research Integrated Domain Group (BRIDG) Model (CDISC, 2015) of CDISC (which will be discussed in Section 2.3.4) shows examples of common identified entities in the electronic health record based on author's definition that are commonly associated with identifiers of one type or another.

1. Patient (including genotype)

2. Patient Anatomical location

3. Gene

4. Geographical Region

5. Health Care Provider

6. Healthcare Provider Organisation

7. Sample / Specimen

8. Device

9. Order / Request

10. Pharmaceutical Product

This is just one specification that takes a view of identity that is not coherent with the EHR specification focus on patient and health professional. This topic will be revisited later.

2.2.3 Unique Identifier and traits

At time of writing, the Irish health service is doing initial trials on a national health identifier for patients called the individual health identifier (IHI) (Martin, W, 2015). Clearly, a unique national identifier will serve many purposes in e-health. In particular, it is expected that a national health identifier will enhance the provision of quality health care services by facilitating the accurate and rapid identification and compilation of an individual's health records (HIQA, 2009). The adoption of a single standard identifier will also lead to more efficient processes and improved patient safety (National E-Health Transition Authority, 2006). If a unique identifier is independently introduced and is not just an extension of an existing number, it may avoid recognised problems from earlier identifier systems. For these reasons, it is important to consider the scope of intended use of the identifier carefully. Listing 2.1 presents 24 archetypes that include at least one mention of *identifier*, as determined by a regular expression search in the openEHR archetype repository. Further, the results indicate some potential candidate identified entities that could be associated with the identifiers that are mentioned within existing openEHR archetypes. Further, the results state that the main identified entities associated with the identifiers by existing openEHR archetypes. An independently assigned identifier would require the creation of a new system for assigning and maintaining the numbers as well as separate technology infrastructure and administrative structures so the development and implementation would require a huge investment. Nevertheless, the positive attributes of having a unique national identifier for patients are still leading those of negative side by many countries' experiences (Hannan, 1991).

Listing 2.1 A summary of use of identifiers in the openEHR archetype repository

```
1  openEHR-DEMOGRAPHIC-ADDRESS.address\-provider.v1
2  openEHR-DEMOGRAPHIC-ADDRESS.address.v1
3  openEHR-DEMOGRAPHIC-CLUSTER.biometric\_identifier_iso.v1
4  openEHR-DEMOGRAPHIC-CLUSTER.provider\_identifier.v1
```

```
 5  openEHR-DEMOGRAPHIC-ITEM_TREE.person\_details.v1
 6  openEHR-DEMOGRAPHIC-PARTY\_IDENTITY.organisation\_name.v1
 7  openEHR-EHR-CLUSTER.device.v1
 8  openEHR-EHR-CLUSTER.individual\_professional.v1
 9  openEHR-EHR-CLUSTER.organisation.v1
10  openEHR-EHR-CLUSTER.specimen.v1
11  openEHR-EHR-OBSERVATION.imaging.v1
12  openEHR-EHR-OBSERVATION.lab\_test-blood\_glucose.v1
13  openEHR-EHR-OBSERVATION.lab\_test-blood\_match.v1
14  openEHR-EHR-OBSERVATION.lab\_test-esr.v1
15  openEHR-EHR-OBSERVATION.lab\_test-full\_blood_count.v1
16  openEHR-EHR-OBSERVATION.lab\_test-hba1c.v1
17  openEHR-EHR-OBSERVATION.lab\_test-histopathology.v1
18  openEHR-EHR-OBSERVATION.lab\_test-immunology-ANA.v1
19  openEHR-EHR-OBSERVATION.lab\_test-immunology.v1
20  openEHR-EHR-OBSERVATION.lab\_test-lipids.v1
21  openEHR-EHR-OBSERVATION.lab\_test-microalbumin.v1
22  openEHR-EHR-OBSERVATION.lab\_test-thyroid.v1
23  openEHR-EHR-OBSERVATION.lab\_test-urea\_and\_electrolytes.v1
24  openEHR-EHR-OBSERVATION.lab\_test-urine\_protein.v1
```

At the very least, this list suggests that in addition to the already-included individual healthcare professional and subject of care the identified entities, specimen, medical device, healthcare provider, and blood product are also suggested.

The identifying traits or demographics details are critical components of identity. The traits of an identity (also called identity credentials) can be represented by certain attributes of a documented object, specifically, the attributes that aid identification of the associated real world entity. For example, a patient identifying trait has three parts that are like traitName "First Name", traitType "String" and traitValue "Paul". Many EHR standards have used the different representing traits of demographics (which will be described later in this chapter).

2.2.4 Identity Domain

An identity domain is similar to the concept of domain that occurs in many places in the computerised world. An identity domain, whether it is for patients, health professionals, or other types of healthcare resources can be seen as a set of identifiers and traits for each identified entity that have been assigned to in the different healthcare organisations with certain administrative policies. For example, in a hospital, a hospital information system (HIS) is closely aligned with identity domains for patients (e.g., Enterprise Master Patient Index or EMPI) and users and each identifier within these identity domains is unique. Identity services operate on the Identity Domain to create cross-references identity sharing, linking and merging. A typical medium size hospital would have identity domains that incorporate hundreds of thousands of identities.

2.2.5 Feeder systems for the EHR

It has been noted that electronic health records have been designed based on detailed requirements documents from the users who are "frontline" carers such as doctors and nursing staff and are responsible for signing information into the record. However, the EHR is not the only health documentation system that is involved in the care process. It is also worth considering the variety of allied health professionals who support the frontline healthcare staff in primary care. These include biomedical scientists, pharmacists, medical physicists and radiologists, clinical engineers, phlebotomists, dieticians, physiotherapists and speech language therapists. There are a number of traditional information systems that are used extensively by allied health professionals as follows.

Laboratory information System (LIS), also referred to as a Laboratory Information Management System (LIMS) in modern hospital laboratory, is used to manage the process of preparation, analysis and reporting of clinical investigations on biological samples of different types. All of these biological specimens can be related back to the patient, but the

primary subject of workflows is the specimen rather than the patient, and this is reflected in the LIS design, as shall be demonstrated in Chapter Four. The laboratory test archetypes which use the two-level modelling approach (will be discussed in Section 2.3.3) have been recently developed and used for the LIS, which support ISO EHR standards (Kopanitsa and Taranik, 2015). This study illustrates the use of archetypes beyond the conventional scope of the EHR but according to the previous discussion, more subjects of information are required to be identified in the laboratory feeder systems.

Order Communications system (OCS) generally developed separately from a LIS, allows the collection and analysis of blood and other specimen types to be ordered by clinicians and taken by suitably trained health professionals such as phlebotomists. The order comms can also be extended to other areas. The main focus here is on the early stages of ordering process and documenting orders. openEHR provides a mechanism for modelling orders and associated actions by the *Instruction* class and its subclass of *Action* (Beale *et al.*, 2006). Any archetypes derived from that classes correspond to orders and so would need to interact with an order comms system. If the OCS is documenting orders, shouldn't an order (a world two entity) be an identified entity in the scope of the OCS?

Pharmacy Information System (PIS) focuses on prescriptions and encapsulates a "drug database" and is linked to local stocks of medicines where appropriate. Also, the literature shows the development of applying archetypes to PIS since 2011 (Farfán Sedano *et al.*, 2011). Again an argument could be made for making pharmaceutical products, or prescriptions a key identified entity within the scope of the PIS.

Picture Archival and Communication System(PACS) / Radiology Imaging System(RIS), Medical physics typically produces large volumes of 2-D and 3-D image data. These digital images are typically managed by a series of information systems that come under the broad heading of PACS/RIS. Again, in this case although images are linked by patient identifier to the patient, the focus is also very much on the image (yet another identified entity) until such time as a report is created which is associated with the patient.

The archetypes have been developed to RIS based on the openEHR approach (which will be discussed to support representing the radiology information in Section 2.3.3) (Araujo *et al.*, 2013).

On the basis of the surveys above, there is a lot of demand for these systems that can use and share information that may or may not be signed into the EHR. This is particularly true for information which is from feeder systems and so not wholly from the health professional's creation. This, however, requires the shared EHR to be applied to the subject of the feeder system.

2.3 The evolution of shared identity models

The identity models (which will be discussed later in this this chapter) are used to abstract either "physical" concepts like patients, samples or orders that serve as the blueprints for constructing physical objects, or "mental objects" such as processes or orders that are less tangible, but according to Popper as also from the real world (Agassi, 1974). Shared identity models, particularly those that describe and identify human subjects, (called demographic models) have been used through information system in healthcare by many standards development organisations such as CEN and HL7. Later in this work, the author will employ a categorisation of different types of shared identity models. This approach is derived from a similar analysis that was used by Bird (Bird *et al.*, 2003) to classify the different types of EHR system model.

2.3.1 Unstructured and semi-structured models

In the unstructured modelling approach, an identity is composed of a simple set of textual entries with no clear classification. This approach if used in an EHR allows it to contain the different forms of identity information, as well as the convenient disposal of the changes in the medical domain. However, this approach causes problems in the long run, because

the resulting infinite number of possible variations in a detailed data structure cannot be successfully managed and queried and cannot easily be extended and made compatible with other systems so that it can impact on the interoperability of EHR. An example of this approach for EHR systems is by using an extensive makeup language (XML) documents to represent the clinical information such HL7 CDA (Clinical Document Architecture) level 1 documents (Dolin *et al.*, 2006) without reference to the associated HL7 reference information model (RIM).

The Personal Information Model (PIMO) (Sauermann *et al.*, 2007) is an example of the semi-structured model that represents the real-world *things* of personal information or knowledge such as contacts, emails and roles and is a lightweight model according to Abiteboul's description (Abiteboul, 1997). The scope of PIMO is to provide a simple lexical classification for individuals by allowing the users to create and personalise its classes, properties and instances but the PIMO lacks the fixed technological structure or schema of data.

2.3.2 Static / single-level models

In this approach, the identity model is represented (in the case of a relational database) by having one or more columns in a table for many of the traits. Figure 2.8 illustrates an instance of the static demographics model that is source from the data elements and interrelated components of ISO/TS 22220 (ISO/TC 215, 2011b). In the view of diagram, the static model approach is designed to support different types of demographics data to be contained in generally fixed set of data structures. For example, the demographics data for the role of a patient may contain telecom, address and name and other demographic traits.

In literature, many EHR-related specifications have the analogous approaches for modelling demographics information. For example, Figure 2.9 shows the equivalence of the demographic concepts between the ISO 13606 and openEHR (which will be discussed in Section 2.3.3). Furthermore, the ISO 13606, openEHR and HL7 RIM allow only a

limited number of identified entities without explicitly resorting to any external terminology (Ceusters and Smith, 2005; Smith and Ceusters, 2006). They are designed in a fixed way to be efficient and to produce faster queries. However, both a static model based approach and unstructured approach may lead to inconsistencies and so lower quality of useable data, and simultaneously it is hard to maintain these types of system over the long-term.
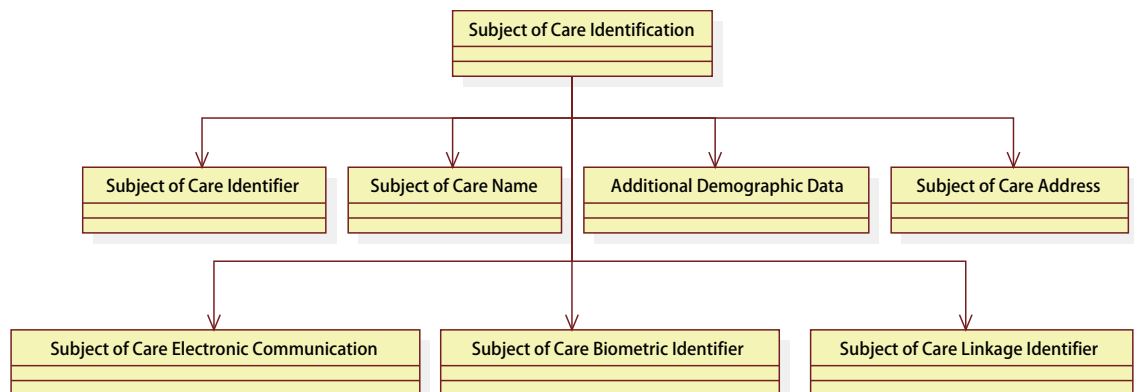
Figure. 2.8 ISO/TS 22220 - A "fixed" demographic model for Subject of Care Identification

Another example of the static model, object dictionary (OD) was previously used for modelling in relational databases and the Human Genome project (Slezak *et al.*, 1995). The idea of an OD model is to provide the structural definitions for clinical objects that could be seen as a data dictionary for storing object definitions. The Synapses project (which will be discussed in Section 2.3.3) adopted the OD approach to providing a standardised set of definitions of healthcare objects, which were mapped onto the local data objects and the Synapses SynOD is a precursor of archetypes. Experience has shown that it is difficult to establish the high level of flexibility and change complexity to low level, need much discussion, and occupied a large number of project work. Level the boundary between problems have not disappeared, it is obvious that many more discussions are required around two level models.

Elsewhere in Literature, the Lightweight Directory Access Protocol (LDAP) is somewhat analogous to the static modelling approach that allows multiple clients (e.g., desktop software often in Microsoft Windows environment) within the enterprise networks to use a

Figure. 2.9 The equivalent demographic concepts between CEN 13606 and openEHR

uniform protocol to access data which tends to be hierarchical, tree, according to the access with the different privileges. The Lightweight Directory Access Protocol (LDAP) schema can be defined by devising a LDAP Data Interchange Format (LDIF) file (Schultz *et al.*, 2011; Sermersheim, 2006). LDIF files can be created for different entities, so it could be argued that there is some flexibility there, but the reality is that the community tends to only adopt standard LDIF formats. In addition, this process is very much in the domain of a technologist rather than a clinical domain specialist. Also, communication of identity would rely on members of a community employing the same technologist-specified LDAP schema. That makes LDAP more or less equivalent to a fixed model that has many of the same features as a common database. For this reason, LDAP with its static schemas and the lack of explicit support for federation by itself, is limiting to manage these problems. However, the LDAP could be a useful component of an identity management infrastructure. It should be noted that LDAPs speciality is authentication and authorisation.

In domains characterised by complexity, large numbers of concepts, and/or a high rate of definitional change, systems based on models such as single level, object and data models to describe entities and attributes are expensive to maintain and usually have to be replaced after a few years. For organising highly heterogeneous health data, Entity Attribute Value (EAV) style model has been used in early medical informatics applications (such as clinical patient records) and other domains such as artificial intelligence (LISP) and operating systems (Microsoft Windows registry) (Nadkarni *et al.*, 1999).

According to Nadkarni (Nadkarni *et al.*, 1999), flexibility is one of the advantages of the EAV approach that allows a user to extend the number of properties associated with an object without requiring additional database fields. However, in many cases, this complexity is worth the effort. For instance in the case of sparse database schemas that need to be constantly changed, an EAV based database design has remarkable feature of flexibility that also supports data growth without the need for schema redesign.

Nadkarni extended the EAV approach to model heterogeneous data by using EAV/CR (EAV with classes and relationships) (Nadkarni *et al.*, 1999). Nadkarni's revision of EAV also incorporated the idea of an object dictionary that was mentioned in the previous section, a concept that is not unlike archetypes. EAV/CR model is similar to the object-oriented database but it gets more clear impression of object and logic to be regarded as a superset of the object-oriented concept. The comparatively lightweight EAV/CR modelling approach, enables rapid addition of attributes associated with an object without requiring additional fields/tags/attributes but it also sacrifices some efficiency for data recall. This also restricts the use of the EAV style approach to use on highly sparse, heterogeneous attributes for typical clinical data.

2.3.3 Two-level models

Two noteworthy EHR projects were funded under the EU 3rd Framework and Advanced Informatics in Medicine (AIM) programme. The Good European Health Record (GEHR) project (Laires *et al.*, 1995) project involved 21 participating organisations across 8 European countries from 1992 to 1995, and sought to create among other things a high quality "fixed" reference model for clinical information. Commencing slightly after GEHR in 1995-1998 and involving many of the same partners, the Synapses project as previously mentioned, developed the object dictionary concept, introducing a reference model (SynOM) and perhaps influenced by the work of Slezak and others (Slezak *et al.*, 1995), developed the idea of clinical objects definitions derived from the SynOM building blocks, that could be edited by users. This set of definitions of pieces of clinical information was called the Synapses Object Dictionary (SynOD) and was intended to be authored by non-technical users (Grimson *et al.*, 1998).

In 1998, GEHR participants agreed to change the name into Good Electronic Health Record (Laires *et al.*, 1995). The Australian GEHR project that followed GEHR refined the SynOM and SynOD ideas from Synapses and "two-level modelling" as we know it

today was born. Also following the GEHR project, an EU open source project, openEHR was started in late 1999 and it adopted the two-level modelling approach to pursue its aims of development, implementation and evaluation methodology for EHR systems (openEHR, 1999). Recent representation models of the EHR or for EHR extract communication, such as openEHR and the EN13606 standard for EHR communication (Bird *et al.*, 2003), allow domain experts to configure and build the shape and content of the EHR. They enable health informaticians to use graphical tools to assemble and constrain generic EHR blocks. These specifications utilise a so-called two-level model. After examining the current literature, it seems reasonable to conclude a roadmap shown in Figure 2.10 for clarification of past outstanding contributions and influences regarding the development of two-level modelling in the health domain.
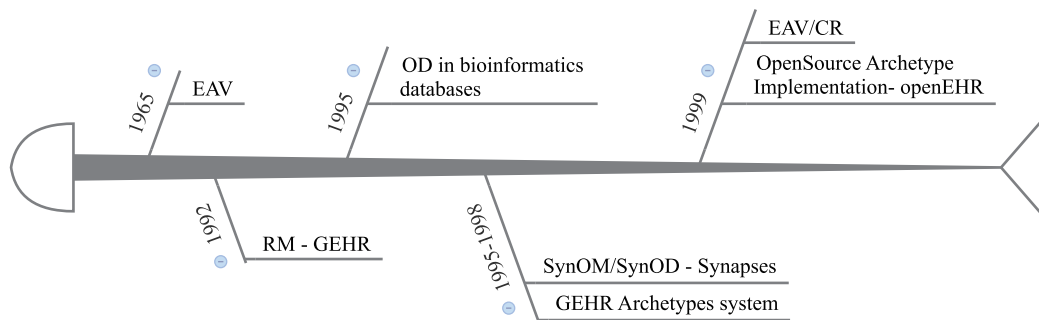
Figure. 2.10 A timeline for the development of two-level modelling

Reference model and archetypes

Figure 2.11 shows the relationship between parts of two-level models and instances of each. In the first level on the left-hand side of the two-level modelling approach, a general model called a *"reference model"* focuses on the generic frame for healthcare domain (Garde *et al.*, 2007a). This approach also introduced another important feature called an *"archetype"* which is used to specify the second level on the right-hand side and specify meaningful content for the EHR. Archetypes are a formal method for representing the

agreed structure of the record components that can be communicated or shared between EHR systems. They can constrain the organisational structure of an EHR that is built using constructs from the reference model at bottom left of Figure 2.11. Archetypes allow the general and abstract concepts from the reference model to be specialised and constrained into useable and agreed information constructs that are recognised by health professionals at bottom right of Figure 2.11. Another way of considering an archetype is as a type of software template that restricts the ways in which the concepts from the reference model can be combined to form sensible (medical and legal) agreed EHR constructs. Groups of domain experts can construct archetypes in the same way that national organisations currently specify messages for exchange of health information.
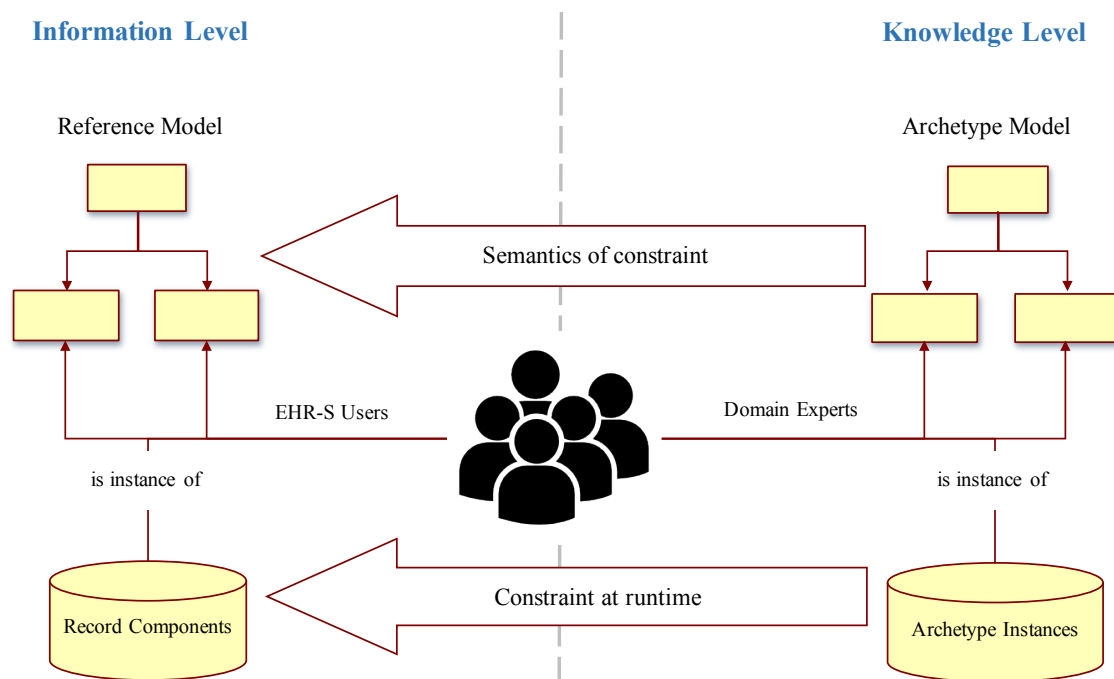
Figure. 2.11 Two-level modelling approach

Reference Model (RM) comprises a set of reusable EHR building blocks. A particular configuration of these generic EHR blocks constitutes the second level model that is composed of archetypes. The two-level modelling approach, as refined in the openEHR project and adopted in the EN13606 standard, also provides an archetype object model

(AOM) and a scripting language called the Archetype Definition Language (ADL) for representing archetypes in tools and as exchangeable documents (OpenEHR, 2005, 2006). The intention is that domain modellers can use convenient graphical tools based on the AOM and ADL to define, manage and share archetypes.

Archetypes allow domain experts to define detailed high quality and very specific models to represent information that is commonly found in the electronic health record. The domain experts construct archetypes with building blocks that mirror the concepts of the reference model. One archetype could, for example, contain instructions on the required characteristics for data capture of a blood pressure measurement. Archetypes are intended to incorporate many variations that occur for each archetyped topic of clinical information. So for instance, the blood pressure archetype should be capable of representing different types of blood pressure measurement archetypes are therefore referred to as providing a *maximal* model of healthcare documentation. In the openEHR series of specifications,

Templates are used by domain experts and developers to make choices to limit the archetypes to particular types by adding additional constraints that suit the particular implementation site. This combination of community-shared archetypes and site-specific templates allows the model to be both flexible and well-constrained.

2.3.4   Function models for sharing identity in healthcare

Previous studies have shown general modelling approaches for hierarchical categorisation within healthcare. However, the literature shows no consensus on categorisation for function models, which indicates that the function models could employ multiple-level and/or dynamic modelling approaches for the identity management. The following identity modelling approaches focus more specifically on the function models for identity that are commonly served by information system such as patient identification process, demographics service and master patient index.

IHE PIX/PDQ

Integrating the Health Enterprise (IHE) PIX/PDQ Profiles was aimed at stimulating integration of healthcare information resources and the IHE technical framework. IHE have defined several profiles for interdepartmental communication (IHE, 2008). The Patient Identifier Cross-referencing (PIX) Profile provides cross-referencing of patient identifiers from multiple Patient Identifier Domains by supporting the transmission between an identity source to the PIX manager and correlating information about a single patient from sources that know the patient by different identifiers (Noumeir, 2008) and a implementation of a master patient index and how to link to identity domains. The Patient Demographics Query Profile (PDQ) (Noumeir, 2008) allows distributed applications to query a central patient information server and retrieve the patients' demographics information (such as when and how to search or visit the information).

MPI/EMPI

Master Patient Index (Enterprise Master Patient Index, EMPI) is a type of index directory for the patient's basic information retrieval (Lenson, 1998). The primary use of EMPI is to provide a unique patient identification through multiple health information systems. To achieve interoperability between the various systems, to ensure the integrity and accuracy of the patient identification, the distributed information of person will be collected in MPI systems. It has been used widely to establish the MPI within the hospital to achieve large-scale systems integration, sharing of resources and EHRs.

CDISC - BRIDG Model

HL7 does not provide comprehensive support for clinical research (Smith and Ceusters, 2006). Perhaps noticing a "gap" in the market, CDISC began developing a "static" Domain Analysis ModeI (DAM) in 2003, which aimed to make standards of clinical research

data that can be exchanged with HL7, and it is Biomedical Research Integrated Domain Group (BRIDG) model. Since the BRIDG model has been accepted by Regulated Clinical Research Information Management Technical Committee (RCRIM TC) and supported by FDA, it has gradually been developed into an independent standard for information exchange between healthcare and clinical research.

The BRIDG model provides a shared view to enable the semantic interoperability between documentation for clinical research and other health and life sciences, which is possibly the best for secondary identities. As the list in Section 2.2 shows, just like the HL7 RIM, the BRIDG model recognises the primary use of multiple identified entities apart from patient and health professional as a part of health information communication for the purposes of clinical research. If such health information was originally collected in an electronic health record for the purposes of healthcare, then this is called secondary use.

IS (Hl7 function model)

The Identification Service (IS, formerly called Entity Identification Service or Identity Cross-Reference Service (EIS or IXS)) (HL7, 2014; OMG, 2009a,b) was developed through numerous iterations under a joint agreement between HL7 and OMG Health Domain Taskforce, in the Healthcare Services Specification Project (HSSP) (HSSP, 2005). The IS development aimed to develop standard specifications to support the improved provision of health care. The HSSP IS was one of the constituent service specifications of the HSSP, which provided a set of services/projects (e.g., Decision Support Service, Common Terminology Services 2, Privacy, Access and Security Services) to support healthcare IT.

The IS is a general specification that is intended to be used for identifying different types of entities, not just persons as in the case of PIDS. IXS entities could be patients, health care professionals, medical devices, information systems possibly even pharmaceutical products within the disparate systems, a single enterprise and/or across a set of collaborating

health organisations. The IXS specification could be seen as a super-set of PIDS, and in the author's view, it is a move towards simpler but more powerful and flexible use of identification of abstracted entities rather than focusing exclusively on a single type of entities such as patients as in the case of PIDS.

Where an earlier IS meta-data model used an approach that was analogous to the EAV model that aims to manage the matching of identities across identity domains. It extensively uses an approach that is similar to EAV/CR in order to create sets of identifying information with a single class with a set of attributes. This construct is named *semantic signifier*. The *semantic signifier* approach proposed in EIS (former IS) is considered to be equivalent to the localised reference model (OMG, 2009a) that is used to identify the structure and semantics of information for EHR communication. Such a mechanism has been extended to represent the entity types in the current IS specification (HL7, 2014). The IS version 1.0 has been adopted as a normative specification and it was formally released as a part of HL7 v3 standard on the basis for the collaboration with OMG in 2014.

2.3.5 Demographics Service - a special case of identity sharing

It has been noted above that communication between health care information systems is the key to securing closer co-operation in a shared care setting, improving handling of patients regarding quality and continuity of care and patient safety. To ensure that health care professionals have access to information about an individual patient with the corresponding privilege, several standard specifications for identity services have been developed over the last few years. A brief summary of some of the main innovations that have led towards a shared identity service is as follows.

PIDS

The Person Identification Service (PIDS) is a service specification that was adopted by the Health Domain Taskforce of the Object Management Group (OMG) (OMG, 2000) or

managing identities of persons within a particular domain. It was originally intended for use within the health domain, but its use was extended to other domains. The PIDS standard includes an interface that supports the ability to connect multiple PIDS components/servers together in a federated manner. These PIDS components were designed and validated for interoperability with a variety of pre-existing person demographics models and record-format standards though healthcare. This design of PIDS standard was to ensure that the specification could permit most pre-existing person identifier management systems and interfaces to participate as members of a complex integration environment.

PDS

The United Kingdom's National Health Service (NHS) Personal Demographics Service (PDS) is part of the NHS Care Record Service, which supports access control and identity management in the United Kingdom (NHS, 2011). The demographic information will form part of each person's electronic NHS Care Record. The PDS is the national electronic demographic service for the United Kingdom and it allows a patient to be identified by NHS staff. The PDS is expected to provide secure, efficient and convenient access to demographic information for 50 million patients in the UK (NHS, 2008) within the NHS Connecting for Health Initiative, which in turn is part of the National Programme for IT (Crompton, 2007).

In many cases, patients' demographic and identity information is stored in local databases from where it can only be accessed from the same organisation or geographical area, which can result in delays in identifying a patient, accessing their correct medical information or even in providing treatment. Becker has noted that the specification and development of the NHS SPINE and the Personal Demographics Service (Becker, 2007), is quite open, leaving room for differing and therefore possibly conflicting interpretations.

2.4   Interoperability specifications in healthcare

From the previous discussion, it is clear that identity management plays a vital role in integration and development of national EHR systems. Interoperability of health information is subject to a number of important and detailed requirements specifications that have been developed by different SDOs such as ISO/TS 18308 (Requirements for an EHR Architecture) (Lloyd and Kalra, 2003), HL7 v3 RIM and openEHR. However, as previously noted, within those specifications the focus in on patients and healthcare professionals, Meanwhile, in the era of connected health, identity management systems must be allowed to interoperate with other systems such as LIS, HIS, PACS, RIS. The following subsections of this chapter survey each of these specifications along with some of the relevant technologies impact on the work.

2.4.1   Requirements for the EHR

ISO/TS 18308 - Requirements for an EHR Architecture is *"to assemble and collate a set of clinical and technical requirements for an electronic health record architecture (EHRA) that supports using, sharing, and exchanging electronic health records across different health sectors, different countries, and different models of healthcare delivery"* (Lloyd and Kalra, 2003). This scope of ISO/TS 18308 depicts a vision for a sequence of activities and general uses of the EHR to support sharing and exchanging health information. However, these requirements are limited to the primary use of the EHR in CEN/ISO 13606 which has been discussed in Chapter One.

Nevertheless, the openEHR architecture recognises the requirements of the EHR which states *"although one of the requirements for the openEHR EHR was a "patient-centric, longitudinal, shared care EHR", it is not limited to this, and can be used in purely episodic, specialist situations, e.g., as a radiology department record system"* (Beale and Heard, 2007). This statement appears to be close to the author's view. However, the existing

openEHR base architecture do not allow the flexibility of the secondary use of its EHR in other domains of health, especially in the feeder systems that do not deal with person - the primary subject of care of the EHR (which will be discussed in detail in Chapter Four).

2.4.2   EHR Communication

CEN/ISO ENV12265, ENV 13606 1-4 and EN13606 1-5

The primary use of identity in this work is to support the development of shared electronic health records. A series of standards developed by CEN TC251 have evolved towards the current EHR communication standard. The first of these standards, ENV 12265 was a foundation pre-standard, which defined the principles of the electronic healthcare records in 1995 (Rossi Mori and Consorti, 1998).

The successor to ENV12265, ENV13606 (Eichelberg *et al.*, 2005) was published by CEN/TC251 in 1999 and the primary aim of ENV13606 was to enable the electronic health-care record communication exchange between heterogeneous systems under European standards. ENV13606 had four parts:

- Part 1: Extended architecture

- Part 2: Domain termlist

- Part 3: Distribution rules

- Part 4: Messages for the exchange of information

The third standard and the centrepiece of recent European health informatics standard-isation were developed by the "EHRcom" project team of CEN Technical Committee 251. This standard has sought to fulfil the EHR requirements that are expressed in ISO TS-18308. This standard is called CEN/ISO EN13606 - "EHRcom". It is a five-part standard that defines and describes various critical aspects of the exchange of electronic healthcare records. EHRcom consist of five parts:

- Part 1: The reference model

- Part 2: Archetype interchange specification

- Part 3: Reference archetypes and term lists

- Part 4: Security requirement and distribution rules

- Part 5: Exchange model

The ISO EN13606 EHR standard is intended to support sharing of health records on a regional or national scale ultimately leading to a shared national EHR system. EHRcom supports the two-level modelling approach that is intended to make health information systems more adaptable and more under the control of domain experts through the use of archetypes.

The five parts of EHRcom are at various stages of the CEN/ISO standardisation process at the time of writing (the last part was published by ISO in 2010). Part two of the EN13606 standard describes, formally, the information model for archetypes and also provides details of information model that can be used to describe archetypes, to exchange them and to persistently store them. EHRcom also incorporates a representation of EHR access policies that are dealt with in part 4 of the standard. Part 4 also describes how an EHRcom compliant system should deal with various security related scenarios (ISO/TC 215, 2009b).

2.4.3 Health Level Seven (HL7) v2.x and v3 Models

As mentioned in Section 2.1.1, HL7v2.x (and v3) models use internal codes in many places to allow a concept or property to be used in different contexts, but in order to cope with diverse use cases, it is sometimes necessary to allow to some level of flexibility / ambiguity at the field level in order to allow a model to be generally applicable. Of course, for a well-designed and safe electronic health record, ambiguity should be minimised. The HL7

organisation recognised the limitations of the v2.x implementations some time ago, and began work on HL7v3. The HL7v3 model is quite general and can be specialised, not with archetypes as in the case of EN13606 but with artefacts called RMIMS, DMIMS and CDA templates. However, HL7 RIM is intended to be multipurpose; it is not specifically designed for the EHR.

2.4.4 Clinical Document Architecture (CDA)

The Clinical Document Architecture (CDA) (Dolin *et al.*, 2006) is an EHR standard of HL7 v3.0. The HL7 has released it for the exchange of health information in medical industry; CDA documents are encoded in XML as a basis for describing the structure of medical records though the development of schema of by acting as a data definition standard. The semantics of the HL7 is by means of constrained profiles of the HL7 Reference Information Model (RIM) and Coded Vocabularies (Beeler, 1998a). In the CDA standard, according to the classification of the structure of medical record information, information is in turn divided into three stages: Level One (header of message), Level Two (basic structure content) and Level Three (detailed content) (Dolin *et al.*, 2006). The CDA level 3 and CDA templates brings a "bridge" between the EHR communication and the HL7 approach.

2.4.5 Identity-related specifications for person

ISO/TS 22220: Identification of subjects of health care

ISO/TS 22220 (ISO/TC 215, 2011b) defines a detailed trait set for identification of subjects of care (not including nonliving subjects) which consists of demographics and other identifying traits (such as subject of care identifier, subject of care biometric identifier, subject of care linkage identifier). The ISO/TS 22220 specification also provides guidelines on applications of the model and makes recommendations about the nature and form of health care identifiers.

ISO/TS 27527: Provider identification

ISO/TS 27527 (ISO/TC 215, 2010) provides a technical specification for uniquely identifying Health Service Providers in the context of shared EHR implementations. Its focus is similar to that of the specification for identification of subjects of health care. ISO/TS 27527 also describes a number of traits for identifying individual health providers and a separate set of traits for health provider organisations.

ISO/IEEE 11073: Personal Health Data

ISO/IEEE 11073 (ISO/IEEE, 2008) provides a family of interoperability specifications to standardise the data exchange protocols between personal health devices such as blood pressure monitors, thermometers, weighing scales and glucose meters. The wide application of ISO/IEEE 11073 focuses on the use of the health devices for person that acts as an umbrella for medical device. In the ISO/IEEE 11073, the part 90101 of health informatics - Point-of-care medical device communication details interfaces to a wide selection of Analytical instruments.This specification was consulted for contemporary ideas on modelling medical device in this work. It is worth mentioning that besides the ISO/IEEE 11073:90101, a specification of ASTM that fulfils the similar functions in particular the data exchange protocols between clinical instruments and computer system is another reference source to represent identity information within the clinical laboratory information system in this work.

OID / II: Object Identifier

An Object Identifier (OID) is a globally unique string identifier consisting of numbers and dots, just like an IP address (ehr_system.root.oid = 1.2.33.22) but these numbers can be the infinite number (Berry *et al.*, 2010b). Each section in OID represents a domain such (1.2.345.xxx is registered by HIQA or HCP). The rest of the number (xxx.xxx.xx.124.32.23)

can represent other identifiers like local system id or other ids of demographics. The OID idea is also supported by CEN, which is defined in CEN/TS 14796 that are used as attribute data types (IIs) in the EHR_EXTRACT and DEMOGRAPHICS Packages (CEN/TS, 2004). Within the CEN EN 13606 specification, OIDs are used as the ROOT of an Instance Identifier (II) which is a complex data type for different type of entities. To keep the design consistent with the EHR communication standard (EN13606), the author has adopted OIDs and IIs for managing identity domains and complex data types that provide binding to the reference model. The data types used for implementing the proposed model are taken from ISO/CEN EN 13606 standard for EHR communication.

2.5 Technical specifications for security and access control

ISO/TS 17090: Public key infrastructure (PKI)

The Public key infrastructure (PKI) for health informatics (ISO/TS, 2002) is an ISO authentication specification The application of PKI in healthcare is fairly similar to its use in the computer network security business and communications as when a person signs a letter and puts it in a sealed envelope and then sends it. A "signature" guarantees the authenticity of identity of the sender and a sealed envelope can ensure the confidentiality of the letter.

The PKI standard (ISO/TS, 2002) consists of two parts:

- *Part 1: Framework and overview* defines the information structure (PKI) of the basic concepts and provides a collaborative work schedule required the establishment of PKI to ensure that health information communications. It also defines the participants in the communication of health information, as well as major security facilities.

- *Part 2: Certificate profile health information - Public element of information structures*. This part introduces certificate profiling and management in relation to PKI.

The PKI and related access control methods in literature can secure the confidentiality through the use of a method for information related to key encryption, but they are not designed for the purpose of allowing other types of entity. If anyone wants to use PKI for health informatics, only the public key encryption system itself is not enough. Confidentiality relies on the policies defined by the organisation or authority of health care and this is major part of the set of document. Authenticity is also a big problem and it requires an existing and working solution of identification system/service and it relates to the work in privilege management (will be discussed in Chapter Three).

SAML, SSO, OAuth and OpenID

Technologies such as SAML (Anderson and Lockhart, 2004), OAuth (Hardt, 2012), SSO and OpenID (Bellamy-McIntyre *et al.*, 2011) that are developed for specific purposes all facilitate exchanging authentication and authorisation of identities. They do not operate on the high level of abstraction required to identify federated healthcare demographics information or the other identity information in the domain of health. These integration profiles for access control help to access the EHRs in various secure ways. A summary of the security and access control modelling approach will be described in Section 3.5.2.

2.6   Summary

This chapter began with a brief introduction of the work and a series of definitions of the key terms and concepts in identity models and service in the health domain, that included a section on evolution of shared identity model. During this chapter, the author has surveyed the state of the art of the main existing identity relevant standards and specifications in the context of the availability (or lack thereof) of shared identity models for the subject of

information of an integrated care EHR system. However, the single modelling approach or the function models for the specific domains are limited for integrating identity information and EHR in a more realistic situation within the feeder systems. This leads to literature review of the more specific integrated approaches for various identities in EHR system which will be introduced in Chapter Three.

LITERATURE REVIEW: INTEGRATED USE OF SHARED IDEN-

TITY MODELS WITH EHRS

3.1 Integrated use of shared identity models

The approach of combining a reference model and sets of archetypes for documenting

clinical information was introduced in the previous chapter. The approach is based on using

a fixed and abstract information model (level 1) that is refined by a large online community

of clinical domain specialists. These domain specialists (clinicians) use graphical tools

to produce specific and flexible definitions for different parts of an EPR such as blood

pressure, Barthel index, eye exam. These definitions are called Archetypes and represent

the second level of the model that allow the information system to be both complex and

accessible to data modellers. As archetypes do not need to be fixed, they can be adaptive

as healthcare processes and technology evolves. The process of developing models of this

type is called Clinical Information Modelling (CIM). A literature review by Moner in 2013

found there were more than three hundred research papers in the broad topic of clinical

information modelling, of which over 50 dealt with two-level models (Moreno-Conde

et al., 2015). The ISO, CEN and HL7 standards development communities have adopted

different styles of the multi-level modelling approach.

One of the consequences of the approach is that the resulting clinical information models are highly detailed. As we shall see later in this work, the combination of reference model and archetypes could correspond to a heavyweight model according to Abiteboul's description (Abiteboul, 1997), with associated disadvantages that are somewhat mitigated by the split into two levels. Another consequence is that the application of the maximal model (which has been mentioned in Section 2.3.3) needs archetypes for multiple purposes, which in turn, requires community-wide consensus on the archetypes to be used. However, while the development of a draft archetype to suit the purposes of a single site can be achieved quite quickly, arriving at a consensus on archetypes that are to be shared at a national or international level requires substantial time and effort from domain experts.

The two-level models support communities of health informatics like EN13606 and openEHR have no vigorous support of a large community of domain experts to lead or take part in the consensus-based development of archetypes so that the work has proceeded slowly (Berry *et al.*, 2010a). This situation above has gradually improved since more countries follow the example of Sweden (Lundell, 2010) and Minas Gerais region in Brazil (B-RES, 2011; Lincoln, 2010) in implementing two-level based EHR. These comments are made about two level modelling for the purposes of documentation of healthcare. They can also be applied to identity modelling. So the combination of reference model and archetypes in a two level model should be comparatively heavyweight, and rather than being domain independent as in the case of an EAV or EAV/CR model, classes in a two level reference model should contain domain specific properties. As in the case of two-level models for documentation, archetypes for demographics and more generally for identity modelling should adopt the maximal set approach, in order to avoid developing multiple similar archetypes. Finally, if a shared two-level identity model infrastructure was to be developed a community of identity experts would need to be engaged in the development of the proposed identity archetypes.

3.1.1 EN13606 Revision – introducing two level models for demographics

Work on more generalised approach to demographics (roughly corresponding person based identity management) including two level models has already begun. The group that is responsible for the extension of the EN13606 is at time of writing in December 2015, appears to have adopted the use of archetypes to allow the EN13606 demographics model to be extended by archetypes. The current EN13606 Demographics model (which has been described in Section 2.4.2) only focuses on the subject of care and does not allow this level of flexibility and precision when it comes to entering data about demographic entities or other entities or resources. The identified entities of the EN13606 Demographics model can only be of three main types: organisation, person, or software/device.

The EN13606 group has recognised these issues and the changes from the demographics model of EN13606 revision aim to integrate the demographics model with the EHR extract while still allowing independent demographic modelling. The changes make the revised demographics model closer to a function model by adding Functional_Role, Link, Demographics_Role_Relationship, which is not constrained by the hierarchical structure of the fixed demographics RM. More specifically, the EN13606 Demographics model uses the demographics archetype to define attributes such as roles, relationship and facility in favour of allowing other kinds of actors and entities rather than the subject of care person or health professionals within the current version of the demographic record.

However, as previously noted, the scope of this forthcoming demographics model is still limited when modelling the demographics entities. In order to allow other types of the identified entity, more flexibility with respect to demographics models, generic identity trait blocks (with an equivalent level of generality to EHR building blocks for clinical data such as folders, sections, and entries) could be constructed to be configured into demographics archetypes which describe the data needed for the identity management of a

particular entity. The changes made from the current EN13606 model could also require a mapping mechanism for the existing archetypes to be used.

3.1.2 Clinical Information Modeling Initiative (CIMI)

The CIMI initiative started in 2011 with the collaboration of its members from main existing SDOs and vendors in healthcare (e.g., EN13606, HL7, NHS, openEHR) (CIMI, 2014). To improve interoperability of HL7 standards on representing clinical information, the CIMI specification adopted the two-level modeling methodology after reviewing many existing modelling approaches. The structure of the CIMI models appear to be a "lightweight" version of openEHR Archetype Object Models (AOM) with a minimal hierarchy of *Cluster*, *Entry*, *Section* and *Composition*.

Significantly from the point of view of this work, the CIMI reference model follows the HL7 RIM approach to use the *Entity*, *Actor* and *Role* class to model the subject of information, which allows modellers to represent the real-world entities including Group, Agent (Device), Person and Organisation. The approach of the Entity-Role-Actor allows the CIMI to create a direct mapping of CIMI RM and HL7 RIM.

The CIMI employs openEHR ADL 2 (which will be described in Section 3.1.4) as its modelling language but it is committed to support a number of formats (e.g., UML, JSON). In comparison to openEHR archetype repository (CKM), more archetypes would be expected to be created for the CIMI archetypes (CIMI, 2014).

3.1.3 Fast Health Interoperable Resources (FHIR)

Resulting from heightened activities in health informatics in the past couple of years, FHIR, an implementation-oriented EHR specification was created by HL7 that extracts the experiences and lessons in the modelling design of HL7's v2 , HL7 v3 and CDA (FHIR, 2014). The scope and process of the FHIR development conform to the patterns of an open

source project that make the FHIR model much more faster to implement for the health IT developer community.

More specifically, the FHIR revises the HL7 messaging mechanism by modelling the clinical concepts as a set of resources and profiles for healthcare. An example of the FHIR resource can be encoded within a simple XML file with the structure information under the following headings; *resource identity*, *summary*, *extension with URL to definition* and *administrative or clinical data* (FHIR, 2014). This FHIR profiling is analogous to the archetype modelling but it does not allow specialisation or template profiling of the FHIR Composition Resource. With regard to defining the subject of information, FHIR uses Administrative Resources including *Person*, *Animal*, *Organization*, *Agent* and *Patient*, which is also equivalent to the Entity-Role-Actor architecture within the fixed HL7 RIM .

The literature shows (Boone, 2013; Grieve, 2013), the FHIR is intended to gradually replace CDA (which has been discussed in Section 2.4.4) as an unstructured model. The author believes that one of the key contributions achieved by the FHIR is to use and support the RESTful APIs to exchange data within health information system, which in the author's experience accelerates the implementation process to lower the development cost.

3.1.4   openEHR AOM/ADL 2

The openEHR AOM defines the structure of the archetype which can mostly be represented using XML-based instances. It represents the archetypes in object-oriented fashion as a treelike object model as shown in Figure 3.1. For example, an openEHR patient archetype consists of six nodes which conforms to the openEHR demographics model. It was noted in Chapter Two that openEHR archetypes are typically expressed in their own machine-readable modelling language Archetype Definition Language (ADL) that describes the semantic constraints of instance of the openEHR reference model. Listing 3.1 illustrates a part of the constraints by example of the same patient archetype definition. In this case, the person is defined by the building blocks as follows; *details* archetype (containing

person_details as shown in line 6 other than name or address), identities (*person_name* as shown in line 12) archetype, contacts (*address* as shown in line 20) archetype and relationships archetypes (defining a patient's relationships such as family ties).

Identities — Name

Contacts — Addresses

Relationships — Relationships / Third party payer / Healthcare provider/Health professional

Patient — Details — Demographic details

Description — Concept description / Purpose / Use / Keywords / References

Attribution — Archetype ID / Original Author / Translators / Licencing
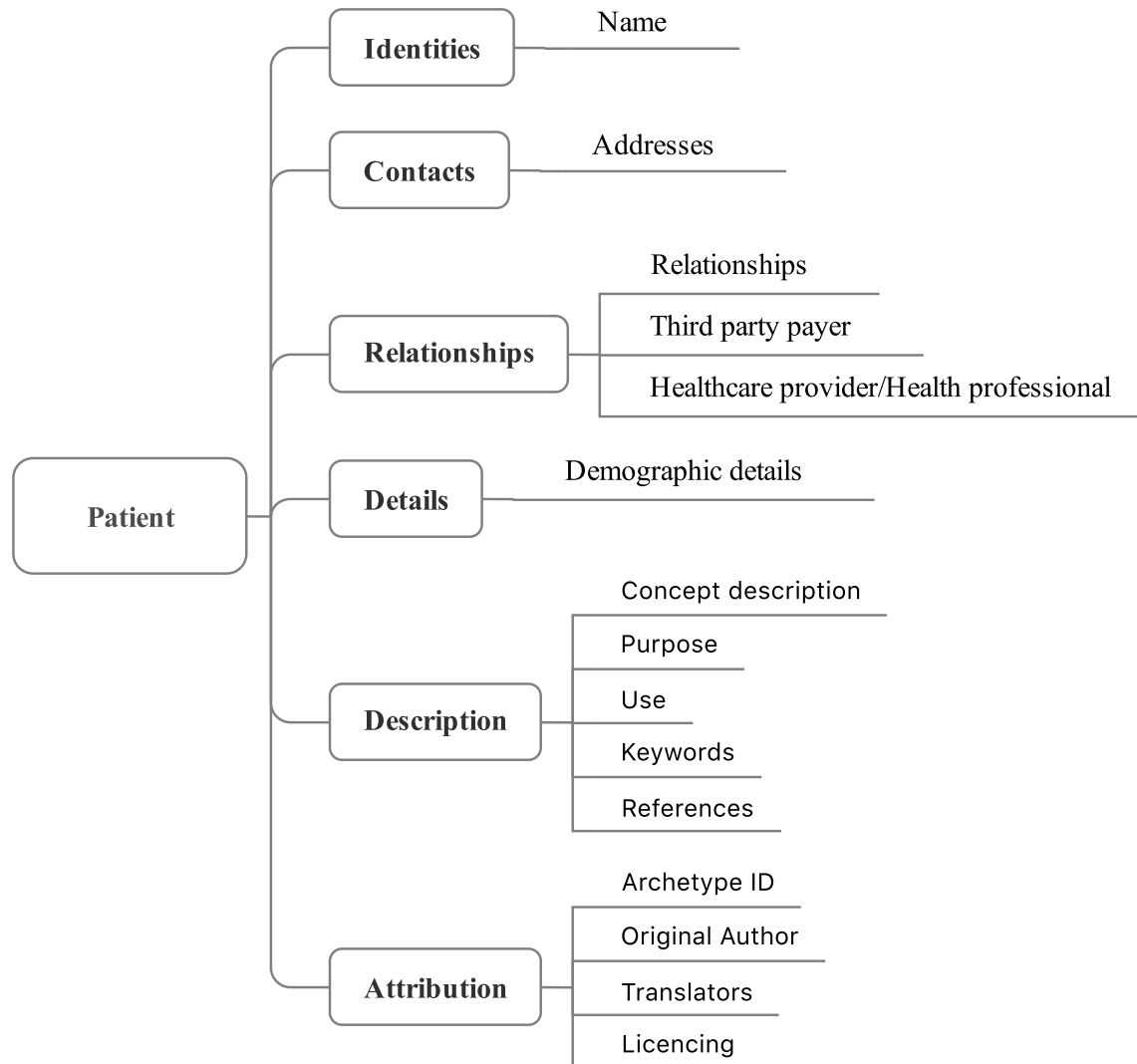
Figure. 3.1 A mindmap of the openEHR Patient archetype

The openEHR community formally released the AOM/ADL version 2 specification in 2014 after taking a number of views and comments from the CIMI members (openEHR, 2014). As an extension of major work of openEHR, an archetype editor tool (ADL Workbench) and an open-source archetype repository (Clinical Knowledge Manager, CKM) were developed (openEHR, 2014). Apart from the technical updates of the openEHR ADL

Listing 3.1 A raw ADL file showing a portion of the openEHR Patient archetype

```
1  definition
2      PERSON[at0000.1] matches {  -- patient demographic data
3          details matches {
4              allow_archetype ITEM_TREE[at0001] occurrences matches {1..1} matches {
5                  include
6                      archetype_id/value matches {/(person_details)[a-zA-Z0-9_-]*\.v1/}
7              }
8          }
9          identities matches {
10             allow_archetype PARTY_IDENTITY[at0002.1] occurrences matches {1..1} matches {
11                 include
12                     archetype_id/value matches {/(person_name)[a-zA-Z0-9_-]*\.v1/}
13             }
14         }
15         contacts matches {
16             CONTACT[at0003.1] occurrences matches {1..1} matches {  -- Contacts
17                 addresses matches {
18                     allow_archetype ADDRESS[at0030] occurrences matches {1..1} matches {
19                         include
20                             archetype_id/value matches {/(address)([a-zA-Z0-9_]+)*\.v1/}
21                             archetype_id/value matches {/(electronic_communication)[a-zA-Z0-9_-]*\.v1/}
22                     }
23                 }
24             }
25         }
26         relationships matches {
27  ...
```

tools, a breaking change of the openEHR ADL 2 is to allow the archetypes to be more terminologically capable which leads to a new internal coding system and terminology binding. Another change to Archetype Object Model (AOM) is to use the meta-data model approach so that archetypes can conform to the standards such as ISO/TS 13972 (Detailed clinical models, characteristics and processes) rather than the RM only (Goossen *et al.*, 2010).

3.2 Data integration of shared identity data

3.2.1 Granularity versus abstraction

In the author's experience, granularity of model and data query performance are often mutually contradictory. A high level of granularity ensures the flexibility of the data queries, but reduces the efficiency of query by means of increased abstraction. In modelling, it is an important problem of balancing the granularity and abstraction. A second problem is that

granularity of document can impact accuracy of recording information but it is especially important in the domain of health as mentioned in Chapter One. The solutions, similar to the data warehouse (Inmon, 1996), are to add additional dimension of abstraction to the atomic data level so it can combine the flexibility and efficiency. This is also the principle of two-level modelling approach that places atomic data on the level of clinical information modelling. Whether and to what extent that the approach is pragmatic, though, it will take a quite a while to solve in this thesis.

3.2.2 Mediated schema

There is a rapidly growing literature on data integration, which indicates that data sharing can make better use of existing data resources. The modern data integration aims to establish an integrated approach with general and implementable significance for the distributed heterogeneous data sources, by providing a unified query interface to multiple data sources. The mediated schema (also called mediator) (Levy *et al.*, 1996) is commonly used approach for data integration by providing a unified logical view of data to hide the details of the underlying data, allowing users to look upon the data sources as a whole. Wiederhold defines a mediator as "*a software module that exploits encoded knowledge about certain set or subset of data to create information for a higher layer of applications*" (Wiederhold, 1992).

The EHR domain, in particular, has characteristics that are problematic for mapping such as changing attributes of continually generated clinical data and higher frequency updating of reference information model as in the case of HL7. And the data is still stored within each local data source while each data source wrapper must convert the data to conform to the "middleman" model. The literature shows that the mediator schema approach was utilised with the EHR in the SynEx (Xu *et al.*, 1999) project to transform data between different syntactic and semantic formats.

An adoption of the mediated schema approach on the two-level models architecture fills the gap of managing these domain-specific data. The data of the numerous complex attributes are non-sparse and can be managed by standard-based archetypes. During data integration of identities, a mediated schema also helps mapping the attributes and definitions back to their origins which is derived from the same reference model.

The archetype-based approach to a mediated schema that will be introduced in the next chapter (described in Section 4.3) for use in this work contains two distinct building blocks:

- A generalised, data-type aware but domain/problem specific reference model (in the lightweight EAV style) derived from ontological identity

- Standards-based archetypes such as EN13606 and openEHR in the health domain that are compatible with that model. The combination of reference model and archetypes can be considered to be heavyweight from the point of view of Abiteboul's definition (Abiteboul, 1997).

The mediation schema applies a two level model to separate generalised information and domain knowledge from the heterogeneous EHR data. In this thesis, the author implements the mediated schemas that employ the identity model based on existing biomedical standards. Generalised identity information is highly sparse data that can be stored in the database structure based on EAV design. According to Nadkarni, flexibility is one of the great advantages of the EAV approach that allows a user to extend the number of properties associated with an object without requiring additional database fields (Nadkarni *et al.*, 1999). In the simplest case, a patient table A is constructed by an EAV patient table B so it allows to add more traits without modifying the database structure as shown in Figure 3.2.

It has been mentioned that EAV introduces additional complexity. Taking the example above where EAV is used with a relational database, this means that a simple SQL statement

Patient A			
patient_id	first_name	last_name	Date_of_birth
101	William	Chen	21/02/1976
102	David	Wang	02/04/1965
103	Emma	Fang	23/11/1981

Patient B		
patient_id	trait	value
101	first_name	William
101	last_name	Chen
101	date_of_birth	21/02/1976
102	first_name	David
102	last_name	Wang
102	date_of_birth	02/04/1965
103	first_name	Emma
103	last_name	Fang
103	date_of_birth	23/11/1981

Figure. 3.2 An example of two different "Patient" data tables

for querying the patient with the birthday less than "*01/01/1980*" in Listing 3.2 will be transformed into more complex SQL statement in Listing 3.3 to produce the same result.

Listing 3.2 A SQL statement for querying the table "Patient A"

```sql
SELECT first_name, last_name
FROM PatientA
WHERE date_of_birth < '01/01/1980';
```

Listing 3.3 A SQL statement for querying the table "Patient B"

```sql
SELECT MAX( CASE trait WHEN 'first_name' THEN value END ) AS first_name,
    MAX( CASE trait WHEN 'last_name' THEN value END ) AS last_name
FROM PatientB
WHERE patient_id IN (
  SELECT patient_id FROM PatientB
  WHERE trait = 'date_of_birth' AND CAST ( value AS DATETIME ) < '01/01/1980' )
AND trait IN ( 'first_name', 'last_name')
GROUP BY patient_id;
```

But in many cases this complexity of EAV database structure (table: Patient B) is worth the effort. For instance in the case of sparse database schemes that need to be constantly changed, if based on EAV-style database design they will have remarkable feature of flexibility that also supports data growing without schema redesign (Nadkarni *et al.*, 1999).

The numerous complex attributes that correspond to non-sparse data can be managed by standard-based archetypes that exceed the EAV's advantages.

On the other aspects of data integration, a mediated schema helps to map the attributes and definitions back to their origins when the mediated schema of two EHR standards is derived from the same GIRM. The author also realises the drawbacks of mediated schema; it is less efficient as well as technically more difficult while mapping the data types from two different standards. By considering the advantages and disadvantages of the implementation, it is worth using mediated schema approach to facilitate flexibility, reusability, semantic interoperability even though it requires more cost of efficiency on basic database operation.

3.2.3   Object Relational Mapping

Object Relational Mapping (ORM) is a well-known approach to persist objects into a relational database by implementing mapping relationships between objects in an object oriented model and underlying relational database. ORM as an enabling technology for middleware solutions, essentially converts one form of data into another, which also suggests the additional overhead cost of function. For example, there are four statements which can correspond to the multiple queries to the database as follows.

1. A person is born in 1980.

2. A patient is older than 80 years.

3. A sample is collected in 2 hours.

4. The red blood cell count is less than 4.2 million/mcL.

Each of these statements can be related to an object that is mapped to the relational database. The ORM model simplifies this operation process so the user can access the data without understanding the database structure. However, this requires a large amount of

Table 3.1 Mapping objects in BLOBs using EAV

patient_id	trait
101	BLOB for the first patient object
102	BLOB for the second patient object
103	BLOB for the third patient object

work on mapping to enable the practical use in the development environment. Another problem of mapping the object-oriented concepts such as Aggregation, inheritance and associations to the relation database is noticed by the author who proposes to integrate all related object's attributes into single table by the EAV-style database structure (Keller, 1997). Table 3.1 demonstrates a solution of mapping the table "Patient B" (in Figure 3.2) to a simple EAV-style table. In this case, the BLOB consists each pair of attributes and values of the object, which can reduce the complexity of the SQL query as shown in Listing 3.3.

There are numerous ORM-based open source persistence frameworks which facilitate the convenience and flexibility of object mapping such as Hibernate, JPA and iBATIS which is used in both the EHR*land* project and the implementation in this thesis. Within those ORM frameworks, the mapping are usually created by the XML document which normally represents the semi-structured data. In addition, the semi-structured data can also be stored in databases, which allows developers to persist data from the archetypes into the relational database (Florescu, 1999; Reinwald *et al.*, 1996).

While EN13606 focused on EHR communication, and so does not place special emphasis on persistence, openEHR specifically focuses on storage of information according to two level models. The openEHR community suggests a node+path approach for archetype persistence by using ORM as shown in Figure 3.3. The node in an openEHR archeytpe has unique identifier of its path, for example by path codes as shown in Listing 3.4.

Listing 3.4 The raw ADL codes of path and node structure in archetype

```
1  use_node  ITEM_TREE  /data[at0001]/events[at0001]/state[at0001]
```

There are certainly many ways of storing the archetype objects using ORM approach. In the approach as shown in Figure 3.3, the clinical information objects as represented by ADL path codes will be serialised into blobs for data persistence by each single node. A significant benefit of using ORM to persist the archetype objects is that it integrate the SQL query language and ADL for querying archetypes in relational database by using openEHR Archetype Query Language (AQL).
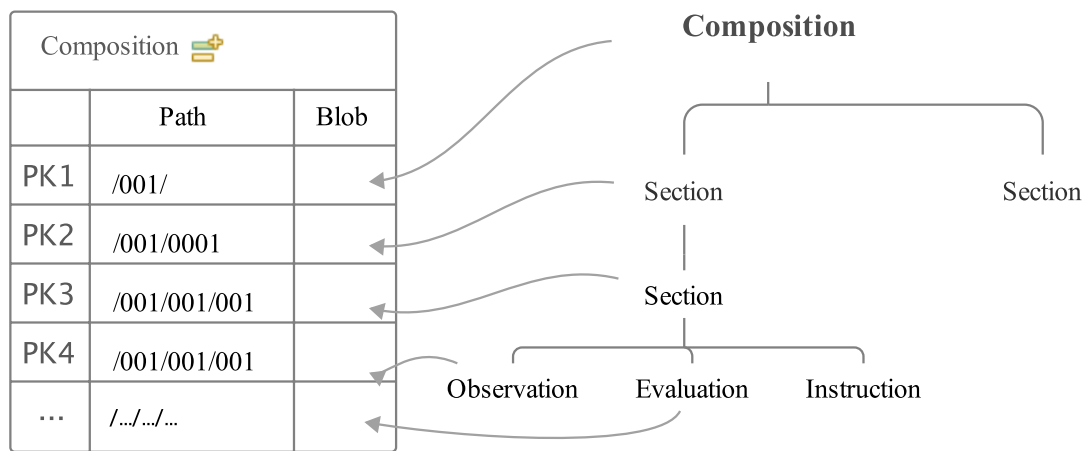
Figure. 3.3 An ORM for openEHR archetype serialisation (openEHR, 2008)

3.3 Ontology-based modelling in e-Health

According to Shirky's definition Ontology (Shirky, 2005),

"*the main thread of ontology in the philosophical sense is the study of entities and their relations. The question ontology asks is: What kinds of things exist or can exist in the world, and what manner of relations can those things have to each other? Ontology is less concerned with what is than with what is possible.*"

The application of ontology in e-Health, involves positioning a vocabulary of basic terms and the relationships within the subject or the domain of health, as well as denotative meaning of combining these terms and their relationships. In an ontology, the fundamental link between the entities is the *is_a* relation (an example of one type of relationship). Next,

a arguably more irregular and concept specific network of *is_part_of* relationships or additional types of relations like *has_a* could be added but the base of ontology is the *is_a* tree.

In addition to the link between individual entities, the ontological approach can also be used to help describe the identifying characteristics (traits) of each identified entity. According to the American National Institute of Standards and Technology (NIST), an ontology of identity credentials (referred to as identity traits in this work), "*is an explicit specification of a conceptualisation of identity credentials, including the actors, actions, and objects that establish the relationships of their production, use, and destruction*" (Gregor and Dutcher, 2006). More specifically, an ontology of identity credentials in the domain of health when used within an ontology of healthcare identities, provides rich meanings and functions for representation of individual identified entities in many health information systems. By comparison the identity model that is developed for this work would be used to describe and provide an index of instances of a small subset of the entities in such a healthcare ontology. Only those entities which were useful subjects of documentation would be considered to be modelled by the identity model and archetypes proposed in this work. The creation process of an ontology according to Bermejo can be summarised as follows (Bermejo, 2007).

1. Define the domain of discourse for the identities (which is the domain of health in this work).

2. Enumerate a range of key terms with regard to the identified entities in this domain such as patient, healthcare professional, sample, medical device.

3. Define the taxonomy model of identity to reveal identity semantics and relationships in e-health with correspondents and characteristics.

To facilitate the efficiency and reusability of the identity reference model, the author surveyed a number of relevant ontologies of identity which distinguish between concepts

of an identity or that are associated with identity information, and other generic attributes and relationships. Another example of an ontology of identity is, *"the set of characteristics by which a thing or person is recognised or known and that constitutes it's individuality or set inclusion"*, which is defined by National Cancer Institute Thesaurus (NCIT) (NCIT, 2015) as shown in Figure 3.4. It should be noted that there are over 300 biomedical ontologies with associated definitions are stored in the BioPortal ontology repository where this generalised identity ontology comes from (BioPortal, 2015).

Identity ⟵ Characteristic ⟵ Property or Attribute

Figure. 3.4 A visual summary of the identity ontology defined by NCIT

Also, there are many tools for creating ontologies. Protégé a Java-based open-source visual ontology editor which is developed by Stanford University is employed in this thesis (Protégé, 2011) to demonstrate typical features of ontology as shown in Figure 3.5. Before continuing, it is worth mentioning that there are two main machine-readable languages for representing ontology surveyed in this thesis.

- The **Resource Description Framework** (RDF) is a W3C data exchange model for expressing the semantic web resources (Klyne and Carroll, 2004). RDF Schema (RDFS) provides a set of semantic description of RDF resources by defining classes and attributes (Brickley and Guha, 2004).

- **The Web Ontology Language** (OWL), is similar to the RDF, but with more vocabulary and syntax features for expressing classes and properties. Therefore, OWL has more options for expression of ontology than RDF, RDFS so it has a greater ability to understand content and context compared with other languages (McGuinness and van Harmelen, 2004; van Harmelen and Frank, 2009).

Protégé provides many reasoning engines to examine consistency of the instance of ontology in support of OWL and RDF specifications (Noy *et al.*, 2003). Figure 3.6
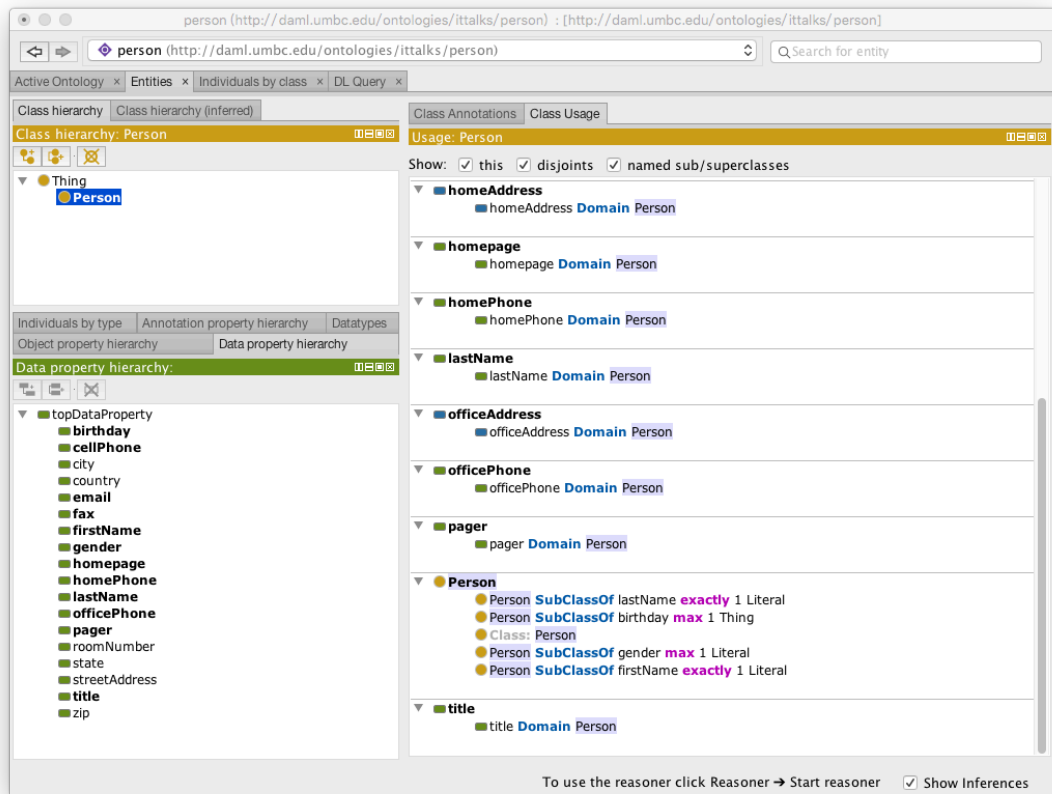
Figure. 3.5 A GUI snapshot of the visual ontology editor Protege showing the Person entity from the vCard ontology and its traits (UMBC, 2015)

illustrates another example of a "patient trait ontology", a single patient entry in a patient ontology which consists a large number of "trait" attributes and there should be more as we can expect (Webprotege, 2015). There appears to lead to a distinction between an ontology of identity and an identity model in the health domain.

- An ontology often consists of a hierarchy of identities representing "real world entities" from extremely abstract concepts at the root of the ontology and becoming increasingly concrete and specific with each individual branch, with semantics of the identity domain and relationships in a categorisation hierarchy.

- In the author's view, an identity model is intended for modelling a small number of identified entities in that hierarchy that are of interest to users and need to be

identified and documented, such as patient, specimen, medical device, order. In particular, the instances of these entities with the identifiable traits and values allows the "real world entity" to be identified so they can be traced back to the relevant documents in the information system.

In other words, an ontology is like the "truth" in representing reality, but in clinical modelling these ontological "truths" are incompatible with the highly changeable clinical information (Blobel *et al.*, 2014). In the author's view, abstraction of an ontology and instances of the "real world entities" are on the different level of modelling in real life situations.

Nevertheless, a suitable ontology does help to identify candidate identified entities and furthermore to tease out the semantics of domain and structure of an identity model. In this work, the generalised concept of identity should consist of identifiers, traits and corresponds to participants/actor and domains. In any case, the ontology approach is helpful for learning about reality. The semantics model of this generalised identity that is derived from the ontological distinction will be described in Chapter Four.

3.4 Federation of identity

Another tool that is used in the management of identity at an enterprise level is federation. According to Sheth and Larsson (Sheth and Larson, 1990), a federated database system refers to a heterogeneous distributed database system with autonomy of component database systems. A federated demographics or identity system by comparison then is a heterogeneous distributed demographics or identity system which facilitates autonomy of component demographics or identity systems. Smith (Forslund *et al.*, 2000) has pointed out that experts were predicting rapid growth of the adoption of federation of demographics information in 2005. However, despite this early promise by April 2008 he made the
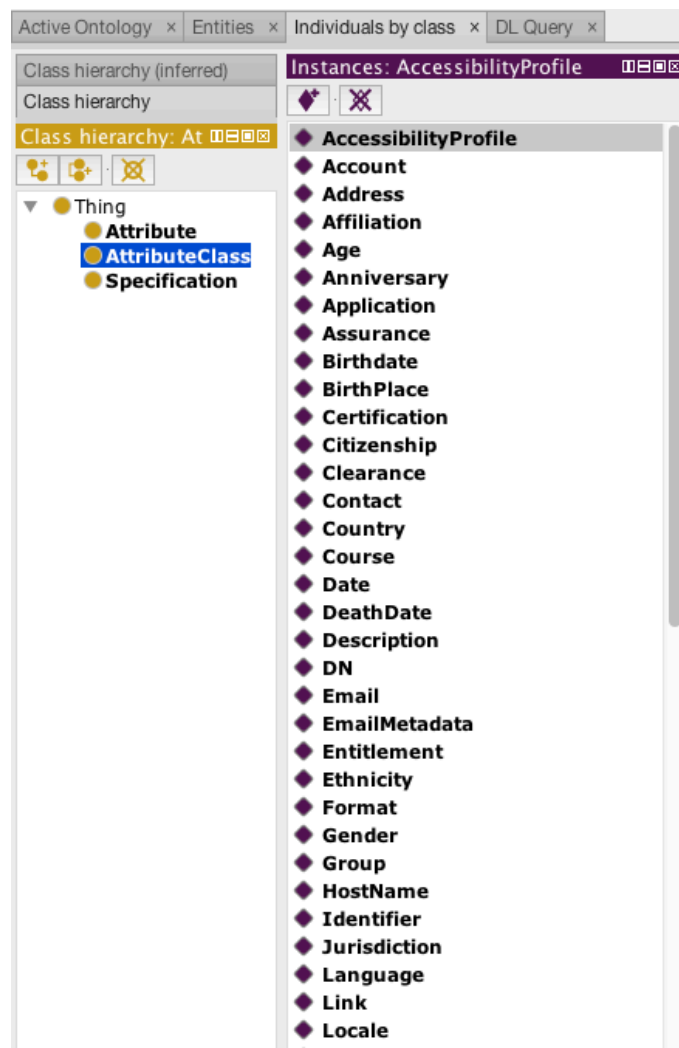
Figure. 3.6 An example of the attributes within a patient trait ontology

point that the predicted growth had not happened Smith postulated that this was due to the following factors.

- The federation technology obfuscates the problem and the concept is cumbersome and inaccessible to be described by language.

- It is difficult to gain and easy to lose trust that is the primary concept underlying federation.

- There are many challenges of extending internal identity to external partners or service providers for organisation.

73

- Federation identity management might not be easily achievable even with an implementation of cross-domain single sign-on that will require identifying a new application for existing application.

- Liability of federation could be an issue that lead to data loss and security breaches due to unchecking on the potential liability of federated partners for companies.

In any case, centralised identity management is the preferred approach for management of health identities in Ireland (HIQA, 2014). The preceding sections have presented a number of approaches that support identity management, including mediation and transformation, linking to ontologies, centralised and federation of identity management systems. The complexities of the identity domain present difficulties for any single approach when taken alone and this affects performance characteristics such as efficiency, feasibility, compatibility and scalability. Although this work effectively mixes these techniques, the author has sought to combine them while developing the two-level models based on data integration approach that requires the combination both of centralisation and federation-based migration of identities.

The author compared and analysed the centralisation and federation for identity management, in particular, focused on the complexity involved in a national service. The author believes that it is possible to extend this approach to other types of identity domains in healthcare such as identification of pharmaceutical products and their supply chains, identification of diagnostic devices, or of health organisations and health organisation units.

3.5 Related topics

3.5.1 Object-role Models

The notion of roles in information modelling was first proposed by Bachman and Daya's role data model (Bachman, 1980; Bachman and Daya, 1977)in 1970 based on the observation that "*most conventional records and relational n-tuples are role oriented. These typically deal with employees, customers, patients, or students, all of which are role types. This role orientation is in contrast with integrated database theory which has taught that each record should represent all aspects of some one entity in the real world.*" (Bachman and Daya, 1977; Steimann, 2000a). In the scenarios within the health domain, a person can play multiple roles of the entity. For example, the openEHR demographics model uses *Relationships* to deal with multiple roles as shown in Figure 3.1. Indeed "entities" such as patient, healthcare professional and healthcare provider can also be considered as roles. This approach would in principle reduce the cost of coding every single type of the entity in the database schema. The OR modelling can dramatically reduce redundant codes (Bachman, 1980) by defining the fields of roles instead of the entity types in database.

A difference between the Object-Role (OR) modelling and the Entity Relationship (ER) modelling or ER-based UML is that there is no concept of attributes on OR modelling which allows more flexibility to model the complex information such as businesses process, clinical activity (Halpin, 1998). A key benefit of the OR models can be represented by the verbal expression in natural language by using Natural Language Information Analysis Method (NIAM) which allows the domain experts to verify the correctness of model contents (Halpin, 1998, 2010).

However, the OR model has the problem of semantic conflict and semantic redundancy due to the dichotomy of types and role. There is an ontological distinction between "roles" and "nature types" according to Guarino (Guarino, 1992) in that nature types do not change

with the time (the person is always the person) but roles can change (the student become the teacher or professional) (Steimann, 2000b). An example of this kind of problem according to Steimann the "misconception" of using a single abstract union, "Party" to as a common supertype for all nature types and roles such as person, organisation and agent (Steimann, 2000b), an approach which is adopted by many SDOs (e.g., HL7, CIMI). Each nature type or role can be the subtype of the *Party* but the instance of them should not be from same type (Steimann, 2000b). That is to say, the instance of person does not share constraint (or property in common) with the instance of organisation from *Party*, and hence the roles should be supertypes of the nature types, not subtypes of the*Party*. To solve this complex type hierarchies, (Elmasri *et al.*, 1985) proposes a category concept which allows the nature types and the roles not to be the subset of *Party* but to be categorised. However, the category approach may result in finding and resolving duplicate dependencies in one other by replacing the role model of all relationships. To some extent, this approach does not look for a single model to abstract the relationship of the nature type and roles and therefore it takes account of the vertical modelling like the EAV.

In this thesis, the author looked at similar ideas of the category that weaken the relationships between identified entities which will be described in Chapter Four. Despite the function of representing semantic equivalence, the object-role model also has the advantage in access control which allows the health records to be securely and appropriately accessed based on the user and group permissions for different roles. This can be regarded as a function which should be accounted for security from the development environment in case it is requested and deemed necessarily.

3.5.2 Security and Access Control

It has been noted that one of the main uses of identity is in access control. Typically, a user is being allowed to access a resource such as the patient record, or information relating to

a specimen, prescription, order or something similar. So the access control must interact with at least two identities.

- The user

- The identified entity (patient, specimen, prescription, order,..) whose record the user is being allowed to access

Three principal variations on the access control model have been widely used commercially: the **Discretionary Access Control** (DAC) in which access is determined by the system rather than the owner and is the basis for access control on UNIX and Linux, **mandatory access control** (MAC) which is often employed within database management systems, and **Role Based Access Control** (RBAC) (Blobel, 2004a). RBAC is probably the most popular access control scheme in use today and controls collections of permissions relating to everything from complex operations such as an e-commerce transaction, to simple as read and write operations.

RBAC separates the user from specific authorisation. In the design of RBAC, the user must have the authority to adopt specific roles to be set, so different abstract descriptions of the licensing authority can be made to easily specify a different role in the collection for each user and give users different levels of authority with respect to access to the documentation associated with an identified entity. Moreover, it reduces the amount of administrative work needed to add or delete users.

Despite being the most popular access control scheme RBAC alone is probably not the complete answer for all needs of access control, especially for providing a comprehensive and satisfactory access control solution for shared electronic health record system. For example, as Becker points out, the security and confidentiality requirements as described in the NPfIT output-based specification OBS and subsequent documentation *"are highly challenging and beyond the capabilities of current access control technologies including role-based access control (RBAC)"* (Becker, 2007). The three-part of ISO 22600 PMAC

(Privilege Management and Access Control) (Blobel *et al.*, 2006) standard is intended to provide a more inclusive and generalised view of access control management to health information resources as well as the services required for managing privileges and access control across the different information systems of healthcare. A number of cooperating access control viewpoints have been described by Blobel (Blobel, 2004a) from different perspectives for the health domain. Figure 3.7 shows these access control viewpoints and they are summarised below.
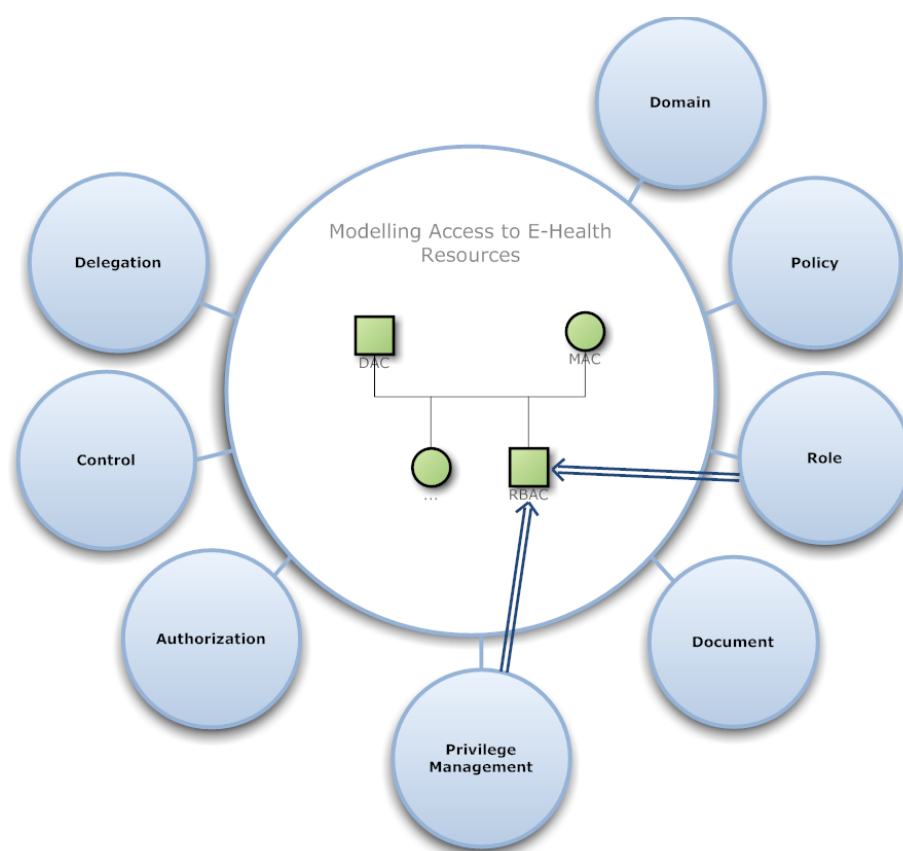
Figure. 3.7 A summary of different model viewpoints on access control

The **domain viewpoint** allows information resources to be grouped into communication domains that share an agreed security policy. Domains can be aggregated into super-domains or broken into sub-domains. These domains could be described according to OMG's definition, as having three components, common policies (policy domains), common environments (environment domains) and common technology (technology domains).

The **policy viewpoint** facilitates a range of different policy types. For instance authorisation policies contain sets of permitted actions; event-based obligation policies define actions which must be performed when certain conditions are met; refrain policies declare actions the subjects must not perform; delegation policies define which authorisations can be delegated and to whom. A basic security policy model is usually characterised by a list of the attributes. For example, policy identifier, policy name, policy authority, policy domain and operation code.

The **role viewpoint** allows privileges to be indirectly assigned to users as individuals are given roles and roles are associated with a set of privileges. This separation allows privileges associated with a role to be updated without needing to modify the role membership. A generic role model can be distinguished in two types which are functional roles and structural roles. It will also corporate with two class types of entity and act.

The **document viewpoint** is that processes and entity roles must be documented and signed for expressing the particular relations between entities and processes. The combination of processes and relations leads to multiple signatures (e.g., in the case of delegation). A simple document model could have several attributes such as document process, document entity role, document user activity.

The **privilege management viewpoint** is used by ISO PMAC specification and it allows the system authority to assign the privilege to individual actors or to groups of individual actors which can be a human user, a system or application, and playing the closed role to role viewpoint. The general privilege management model consists of three classes, object, privilege asserter and privilege verifier.

The **authorisation viewpoint** is quite similar as privilege management viewpoint and used by OMG RAD specification and authorisation logic is encapsulated within an authorisation facility that is external to the application. In order to perform an application-level access control to clinical object, an application requests an authorisation decision

from such a facility and enforces that decision (OMG, 1989). The general authorisation process has two parts, granting privilege and assigning permission by the security system.

The **control viewpoint**, illustrates how control is exerted over access to a sensitive object operation (Klein *et al.*, 2003). Access control is the process which determines whether a claimant's privileges permit him/her/it to access a service provided by a target component. In this context, access is broader than acquiring some data. It might refer to any service offered by a target component (e.g., deletion, computation, transfer). There are four components in the model: the claimant, the verifier, the target and the control policy.

The **delegation viewpoint**, a source authority can delegate to certain delegation administrators the privileges to create and manage the identity management for an authorisation entity (Klein *et al.*, 2003) There are three components of the delegation model: the verifier, the source of authority, and the claimant. The verifier grants an entity known as the source of authority with global privilege within the context a delegation occurs. The claimant asserts its delegated privilege by demonstrating its identity. The source of authority may also process a request from an entity to delegate its privilege by issuing an attribute certificate to another entity.

Other access control schemes have been described in the health domain. **Identity Based Access Control** (IBAC) (Magnussen and Stavik, 2006) also means that, regardless of where or when an individual appears on the network, policy appropriate for that individual can be enforced. Policy based on the individual means that non-trusted users can be prevented from accessing the network even though they connect through a seemingly legitimate connection point. Identity-based access control makes it possible for mobile users to roam throughout the network, and yet continue to have the appropriate access to the resources based on business need. **Process Based Access Control** (PBAC) (Blobel, 2004b) is an authorisation system where each (web) service publishes a list of operations that it can perform and PBAC determines which operations can be called by each user in different contexts. In this case, the process can be considered as a "World 2" concept.

In order to satisfy access control requirements then, it seems that an identity model needs to be able to provide identity information on the user and role who is seeking access to information, and the identity of the entity or resource that they are seeking information about. Principally, (in the case of the EHR) this entity is the subject of care or patient. In other information systems in healthcare the subject of information could be many other "World 1" or "World 2" entities as described in Section 2.2.2.

The *information distance model* applies increased access control restrictions depending on the "information distance" between the information about an entity and an actor who seeks to interact with it. So for instance in the electronic health record, the originator (the subject of care) is "closest" followed by the producer (author/interpreter) of the information and next comes the administrator (user) of information. Lovis *et al.* (Lovis *et al.*, 2007) described zones of medical responsibility and physical location to indicate medical responsibility and therapeutic relationships which can supplement more general role assignment so that the EHR of subjects of care who enter the care flow of a particular heath unit can be accessed by healthcare professionals with appropriate roles within that health care unit.

Distribution rules define the behaviour of an access control component which is adopted by the EN13606 part 4 (ISO/TC 215, 2009b), and the attributes of an access control policy can be categorised into who, when, where, why and how. Sucurovic (Sucurovic, 2007) indicated that for calculating access control decisions, the attributes governing access across these categories can be processed using logical AND operations while attributes within a single category can be processed using logical OR operations.

The progress towards the electronic health record has led to a significant "fading of boundaries" between health information systems (Hammond, 2003). The reliable identification of notable instances of health phenomena is increasingly important in the context of distributed multidisciplinary care where patients will receive care from different health professionals at different locations and times. Among the basic functionality

81

requited to support this trend, provision of integrated identity management and access control facilities is worth considering. Another element is the need for generalised identity of the type promoted in this thesis.

3.5.3   Clinical Terminology

Clinical information can be classified in many ways according to its semantic diversity and its rich meaning under different clinical contexts. Many of these contexts are central to the core EHR. Others relate mainly to clinical research and clinical trials, health management or population health. To harvest all the possible but rigorous intended meaning when communication happens, meta data which may describe the intended clinical meaning are to be placed to aid comprehension. Terminology is a mechanism that operates at the meta data level to improve semantic interoperability.

The clinical terminology approach provides a standardised description of medical terms and concepts to reduce ambiguity, with effects similar to the clinical ontologies that further provides a standardised and consistent relationship between reality and medical concepts. Indeed some clinical terminologies also function as ontologies. It is very beneficial to eliminate ambiguity of medical vocabulary in EHR communication. The most common of international medical code sets and tools include the International classification of diseases (ICD), Current Procedural Terminology (CPT), Systematised Nomenclature of Medicine - Clinical Terms (SNOMED - CT) and Logical Observation Identifiers Names and Codes (LOINC) (AMA, 1970; Cornet and de Keizer, 2008; WHO, 2010). Clinical terminology coding systems have also been introduced to the two-level modelling approach such as openEHR, EN13606 from the beginning. The codes (terminology identifiers) with "at00xx" in Listing 3.1 can be used for linking archetype terms to the terminology codes. The integrated use of clinical terminology in openEHR is not about representing the clinical entities of reality but provides bindings by the archetype codes to the terminology codes in ADL.

While archetypes act as constraint specifications on record structure, the terminology that is embedded in archetypes or instances of archetypes (i.e the data that a particular medical record is holding) serves as the canonical vocabulary and the ontology of medicine for inference. As the development of the mechanism is progressed by international organisations and open communities, great benefit can be realised by allowing these mechanisms to cooperate rather than compete. When a combined approach is adopted, useful and reusable user generated knowledge can be analysed and studied to improve the technologies. This is as true for identity-based archetypes as it is for those intended for clinical documentation, even if the number of archetypes required for identity of certain types of subject of information is much less than for clinical documentation.

3.5.4   Health Natural Language Processing

The integrated use of shared EHR discussed above has improved significantly over years, but many challenges remain to effectively capture, process and analyse the mass medical data from unstructured, complex and heterogeneous sources (Friedman, 2009). Such scenarios fit well the scope of Natural Language Processing (NLP) in the domain of health. Especially while dealing with free-text EHR data, the NLP is a important solution for a large number of ambiguity and unstructured data. And the process is complicated, including often combing through various technologies such as information retrieval, speech recognition and machine translation (Demner-Fushman *et al.*, 2009). The NLP is out of the scope of thesis - an identity system that needed to rely on NLP would be troublesome to say the least, but it is worth mentioning that it has been the important research aspect of the EHR.

3.5.5   Clinical modelling tools and projects

Regarding more recent discoveries (Frade *et al.*, 2013), there are four publicly available archetype editors, which are the Linkoping LiU Archetype Editor (LiU, 2006), the LinkEHR-Ed Archetype Editor (Maldonado *et al.*, 2009), the Ocean Informatics Archetypes Editor (Ocean Informatics, 2013) and the openEHR ADL Workbench (openEHR, 2014). The clinical modeling community have produced more than just archetype editors, such as ADL designer, template designer and ADL compiler but lack of maturity of the tools (openEHR, 2015). An archetype form creator is also developed within the EHR*land* project that only focuses on EN13606 archetypes (ISO/TC 215, 2008). At the time of writing, there is a small amount of work towards developing demographics archetypes, but there has not been an attempt to harmonise the main demographics information models and approaches.

In this work, the author will demonstrate the use of the LinkEHR tool to create archetypes. The LinkEHR is a Spanish project which focuses on semantic web technologies for management of the EHR. The LinkEHR editor GUI supports two schema formats of the reference model representation (.XSD, .RM) but this work employs the XML schema for representing identity models. XML Schema (XSD) is an XML-based mark-up language which describes the structure of an XML document. A feature of XML schema is to provide the validation of data elements, attributes and types by declaring simple type or complex type definitions. Furthermore, XSD is also extended to be widely used in a number of related applications such as XML Cursor, XML Beans and XQuery (Chamberlin, 2002; Gao *et al.*, 2009).

The LinkEHR editor requires a set of mandatory settings for creating a new reference model such as Organisation name, Reference model name. In the meantime, the types of the represent complex elements need to be selected and added from a XSD file that is including the specified XML representation of reference model by LinkEHR editor.

An example of the EN13606 demographics archetype that are being edited by using the LinkEHR editor is shown in Figure 3.8. The EN13606 demographics archetype is encoded as an ADL document initially and then it can be mapped to XML schema for creating XML instance which is used for retrieval and query of the demographic entities.
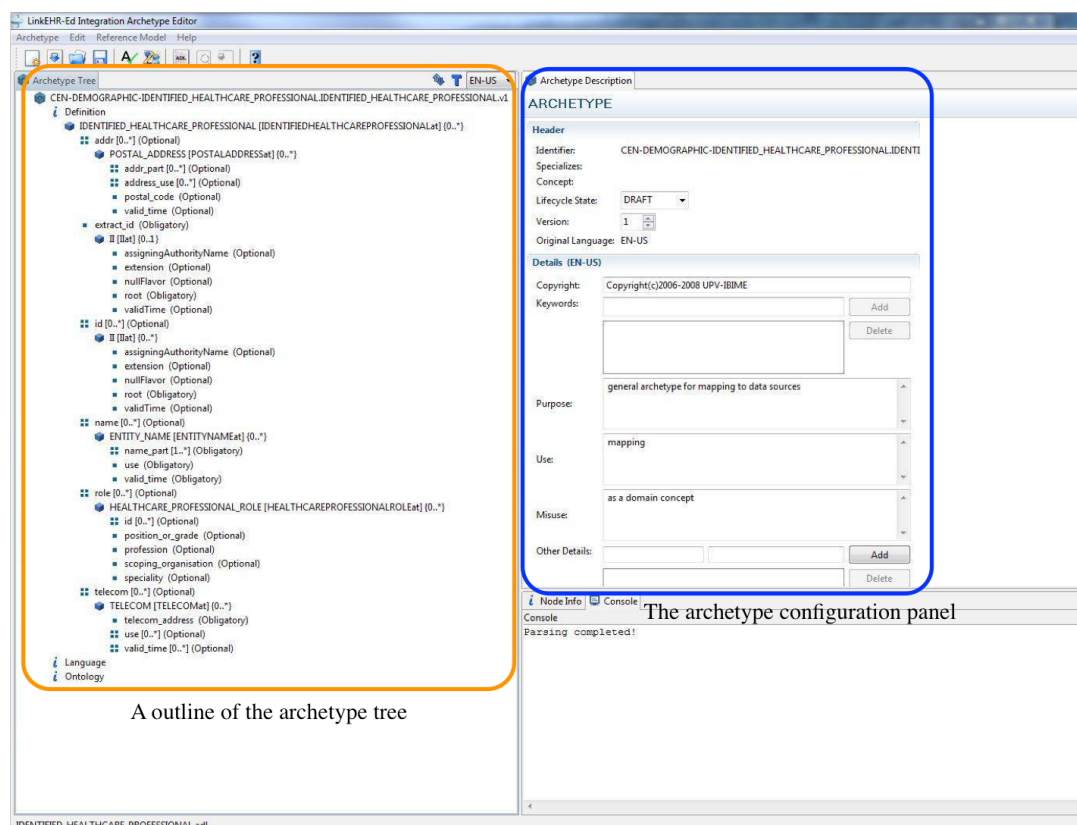
Figure. 3.8 A snapshot of editing archetypes by the LinkEHR editor

In this work, XSD is also used as a tool to represent the model that can be instantiated to XML-based messages for exchanging or to be serialised into object relational database for the persistence.

3.6   Summary

In this chapter, the author briefly surveys the motivation for integrated use of shared identity models, while looking at the main existing solutions for shared EHR and outlines the

need for development of a general two-level approach for the subject of information in the domain of health. After introducing the essential techniques of integrating the two-level modelling approach, the author surveys various situations and related work towards the interoperability of shared identity within EHR systems. The principal objective of the literature review in this chapter is to investigate the features of a feasible two-level modelling approach for shared identity.

DESGIN AND EVALUATION STRATEGY FOR A GENERALISED

IDENTITY MODEL

4.1 Overview

Previous chapters have charted the state of the art in the area of identity management and identity information models. It has been seen that specifications that incorporate two level models either use those models solely for the task of documentation – to hold clinical information – or as in the example of the current openEHR specification or the proposed next version of EN13606, they allow archetyping of the patient and the healthcare professional in a demographics model. Those health information specifications that model entities other than patient and healthcare professional do not employ two level models. This leads to the main thrust of this work.

In this chapter, the author describes the approach of a generalised two-level model for identities called the General Identity Reference Model (GIRM). The GIRM is designed using the two-level modelling approach which is an evolution of the different approaches to demographics model, identity model and associated technologies in previous chapters, two and three and associated technologies. Clearly these approaches are already widely used for representing clinical information in EHRs. As mentioned in previous chapters,

the subject of information in healthcare documentation is not the same as subject of care. The question arises of to what extent they could be used to represent regional demographic RMs and other general identity concepts within the "World 3" documentation which has been described in Section 2.2.2. To illustrate the methodology regarding the semantic interoperability of an EHR, the GIRM is also used as a case study to show how to represent the subject of documentation that is not in particular for demographics perspective but also is allowed to express other identity types in healthcare and other identity domains.

The chapter is structured as follows:

1. Section 4.2 explores the challenge of fixed identity models by using comparative case analysis of the demographics models, in particular, which in the author's view are not well suited for sharing heterogeneous identity information in healthcare.

2. Section 4.3 will demonstrate how to integrate the EHR and various identities within the different domains by using the GIRM as a mediator in this chapter. The process of examining the features from existing standards and specifications in Section 4.3.3, will lead to the derivation of the GIRM that involves reasoning of the modelling approaches.

3. Section 4.4 will discuss the idea that this thinking about generalised identity should be fully incorporated into health information models for electronic health records and their feeder systems in order to support better integration of information between feeder systems and EHR systems.

4.2 Challenge of the sharing heterogeneous identity information in healthcare

4.2.1 Investigation of the main existing demographics models

In this section the author investigates five working models for subjects of care from the specifications (which has been introduced in Section 2.4) ISO/TS 22220, HL7 RIM, EN13606 and openEHR. This study mainly focuses on constituent demographic elements and corresponding details (patient and healthcare professional) to be supported by these models.

EN13606 demographics model (ISO/TC 215, 2006) provides a demographic package related to entities that comprise the demographic aspects of the reference model for this standard. The package consists of two primary entities, subject of care (*person*) and *health-care agent* of subject of care as well as three related party entities (*healthcare professional*, *organisation* and *software_or_device*). In the package, EN13606 defines for each of these entities certain demographic traits, including *telecom*, the usage group of *entity_name*, *postal_address* and the detail classes of *entity_name_part* and *postal_address_part*. This demographic model is more specific than the other three models but it is not extendable.

HL7 RIM v2.14 demographics model (HL7, 2014) brings in a number of classes that not only support demographics in particular but also cover numerous external but related functional-components. The HL7 demographics model classifies the multiple entities by the categories and roles. For example: a person entity can plays the roles of patient or employee and the entities can be assigned as *livingSubject*, *notPersonLivingSubject* and *material*. The full extent of this interesting but nevertheless, heavy and complex model is not relevant to this work at this stage, and the study only takes the person entity with the patient role which is defined as a semantic profile of HL7 v2.14 Patient. HL7 v2.14 also has the valid features of usage group and data value for *name_part* and *address_part*.

The minimum identified traits used for patient as a basis contain patient identifier, patient name, and date of birth, administrative sex and patient address.

ISO/TS 22220 demographics model (ISO/TC 215, 2011b) only defines a fixed detailed structure for the subject of care identification (not including non-living subject) which consists of seven data elements: *Subject of Care Identifier*, *Subject of Care Name*, *Additional Demographic Data*, *Subject of Care Address*, *Subject of Care Electronic Communication*, *Subject of Care Biometric Identifier* and *Subject of Care Linkage Identifier*. The model can be seen as an extended model for subject of care person identification of CEN 13606 demographics.

ISO/TS 27527 demographics model (ISO/TC 215, 2010) gives more specific and structured fixed model for the usage of provider identification which can be seen as an extension of the EN13606 demographic model for provider identification, it also includes the complex data elements from HL7 such as field of practice, registration details, and organisation start date. In this comparison study, the data elements listed above are considered equivalent to the trait type.

openEHR demographics model (Beale *et al.*, 2006) is specified for either a standalone or a "wrapper" service for an existing system such as a patient master index (PMI), and the concept of party used in openEHR is roughly equivalent to the entity. However, the demographic details of this model are generalised to the complex data types. This is similar to the approach taken by HL7 v2.14. For example, the subclass of party address contains an attribute details which is part of the reference model that contains the details of address in one of three different forms.

The author analysed what is covered in the normative attribute sets of each model and identifies what has to be added to mapping of the EN13606 as it was mentioned that these specifications are limited, and do not model enough identifiable entities. Table 4.1 gives some perspective of similarity between EN13606 and the other models which is high at this stage. There are 14 mismatch concepts that indicate ISO/TS 22220 and

ISO/TS 27527 partially match EN13606 as they fit their specialised scopes for patient and healthcare professional. This similarity can be advantageous for data integration because it makes the mediated schema more simple to create for EN13606. However, it is far more complicated that an accurate mapping of individual model can be very time consuming among the heterogeneous EHR systems. Taking this process a step further, the similarity of the concepts among these existing demographics models lead to the integrated use of shared identity model.

Table 4.1 A comparison of the concepts of subject of care exsited in selected models

EN13606	HL7 v2.14	OpenEHR r1.0.1	ISO/TS 22220	ISO/TS 27527
IDENTIFIED_ENTITY	ENTITY	PARTY		
TELCOM	ENTITY.telecom	ADDRESS.type: electronic	SUBJECT OF CARE Electronic Communication	Provider Electronic Communication
PERSON	PERSON	PERSON		Individual Provider
IDENTIFIED_ HEALTHCARE_ PROFESSIONAL	EMPLOYEE	PERSON		
ORGANISATION	ORGANISATION	ORGANISATION		Organisational Provider
SOFTWARE_ OR_DEVICE	DEVICE	AGENT		
HEALTHCARE_ PROFESSIONAL_ ROLE	ROLE	ROLE		
Link	RELATIONSHIP_ LINK	PARTY_ RELATIONSHIP		
SUBJECT_OF_ CARE_PERSON_ IDENTIFICATION	PERSON	LIVINGSUBJECT	SUBJECT OF CARE IDENTIFICATION	
ENTITY_NAME	Entity.name	PARTY_IDENTITY	Name Usage Group	Name Usage
ENTITY_NAME_ PART	Role.name	PARTY_IDENTITY	Name	Individual Provider Name
POSTAL_ADDRESS	PLACE	ADDRESS	SUBJECT OF CARE Address	Provider Address Details
POSTAL_ADDRESS_ PART	PLACE.addr	ADDRESS		Address line

Along with the drawbacks of two-level modelling (which has been discussed in Section 3.1), a barrier still exists while adopting the two models that require domain experts to understand how to use the RM concepts. Along similar lines, the common use of the fixed models on demographics between distributed systems, which can be difficult to extend

as changes to the fixed information model can cost more of changing software (Berry *et al.*, 2010a). Such fixed models do not tend to support consensus-based development of concepts by the domain community. However, the literature also implies that introducing this flexibility into demographics model would comes at a cost of design-time complexity in the systems that use this dual model approach (Garde *et al.*, 2007a). This is undoubtedly true in the author's opinion. It is up to the HI community to decide whether the the benefits of the flexible approach of the type presented here is worth the "trouble" associated with that extra complexity.

4.2.2 Access control of different subject of information

The essence of managing identities in information system is to provide a technological process to allow the right users within an identity domain to access the right information resources within an identity domain. Identity in healthcare is more complex. The healthcare professional who is typically the "user", and the patient record ("the resource") both need to be identified, indexed, queried in a constantly changing non-uniform federated environment so there are two clear identity domains.

There is no doubt that the provision of secure widely-shared patient records which can nevertheless only be appropriately accessed by the right health professionals at the right time is a complex goal which requires complex solutions. Multiple viewpoints in particular could be considered as a composite of identity management and access control to support different use of access control to access different subject of information. For instance, the health professional's views are quite different from the clinician's view in the laboratory. The existence of a flexible identity model that allows additional subjects of information and additional healthcare actors to be identified, would allow access control policies and domains that are central to PMAC to expand beyond just the EHR to include other healthcare information systems. A Biomedical scientist, or other allied health professional, in addition to the information that is available in their own departmental system (e.g., LIS,

PIS, PACS), may only need access to a portion of a patients EHR to allow them to do their jobs. PMAC-type functionality in order to produce policies and domains for interoperable environments, therefore depends upon more liberal interpretation of user and subject of information.

In addition to the problems associated with the integration of legacy systems including the harmonisation of access control approaches and linking of identity domains to support interoperability between clinical information systems there are numerous access control issues which general solutions have yet to be found including identification of medical devices and pharmaceutical products, health organisations or units.

4.2.3   Ownership of the ideal EHRs

As mentioned in Chapter One, the exact scope of the EHR and the type of data that should be permitted into it has been a subject of dispute for many years. This is associated with clinician "ownership" and responsibility for the record and leads to the idea of "attested" EHR data.

According to the definitions of the EN13606 / openEHR records, all EHR data should be "signed in" to the record by the healthcare professional. This idea arises partly because the EHR is supposed to be a legal document. Clearly then, there must be an "outer" corpus of un-attested information from which the signed record is produced with the appropriate authorisation.

While, in the electronic health record, the primary actors in the healthcare process are the patient (subject of care) and the clinician (healthcare professional) some specifications additionally refer to other entities such as healthcare organisation (e.g., ISO/TS 27527, openEHR) medical devices (e.g., EN13606). It is fair to say however, that these other entities are not given central status and in the author's view this emphasis on the two key entities is related to "clinician ownership" of the EHR (Beale *et al.*, 2006).

The recent evolution of patient-centric care and medical documentation, to some extent nudges record ownership in the direction of the patient and the growing importance of this outer corpus un-attested information can also be applied to documentation that is associated with other health viewpoints. Thus, multiple viewpoints often means focus on multiple identities for the subject of information. In Figure 4.1, the author depicts its scenarios about whether patient-centric documents can be reused by the wide domains in healthcare.
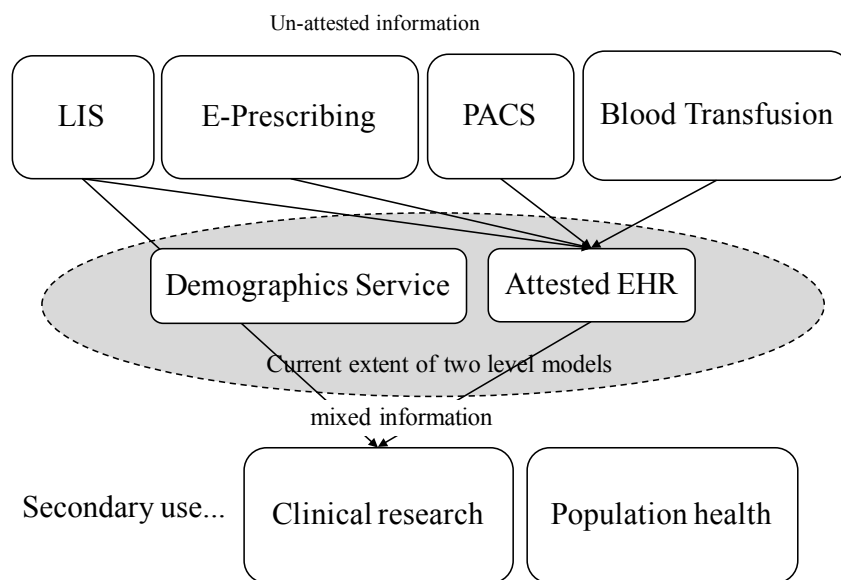
Figure. 4.1 Cross-border interoperability of the EHR

Whatever about the separation of signed and unsigned data, the author believes that it is natural to wish to attempt to use a similar approach to represent both attested and unattested EHR data. However, there are many examples of unattested data that require identified entities other than patient and healthcare professional. These include as shown in Figure 4.1, LIS data, omics information, and information for blood transfusion, organ transfer, and personal health records. These are all examples of information that may or may not ever be signed by a clinician as all this information is created elsewhere. In a similar fashion, information that has been signed into the EHR can be put to a number

of secondary uses; for biomedical research or clinical research; to support clinical trials which require additional identified entities.

A question then arises, why shouldn't these additional identified entities (e.g., specimen, researcher, organ, limb, medical device, control group) be represented using two level models? If such identities are allowed to be represented even outside the record using two-level models, a general identity model such as the one presented in the next section is required (Späth and Grimson, 2011).

4.3 A generalised identity model

In this section, the author proposes a generalised identity RM that can be used to represent demographics models and concepts based on the architecture of two-level models.

4.3.1 Reasons for the proposed approach

As discussed in the previous section, there are various reasons why a general model is needed, which can be summarised as follows:

1. The existing demographics model are complex but lack extensibility.

2. The ambiguity is not eliminated by the use of the federated data models according to Hu and Peyton's report: "*often , there is no common identifier and the identifying information is inconsistent and often ambiguous* (Hu and Peyton, 2009)").

3. The organisations are obliged to rely heavily on automated entity resolution approaches. Banton *et al.* (Banton and Filer, 2014) noted that the "*quality and integrity of MPI record data is highly compromised*")

4. A more inclusive use of identity (Ceusters and Smith's tracking referents) can support multiple identity and indexing contexts for EHR reuse (Ceusters and Smith, 2005).

5. The lack of standards associated with data collection for identity (Arellano and Weber, 1998).

6. A generalised identity model has a capacity to represent common features of existing models even though the subject of information is different.

7. Identification of those entities helps to keep track of their evolution, their disappearance, and also the appearance of new entities (e.g., polyps, fractures, bone components transplants).

8. As mentioned previously, this approach it is hoped would facilitate interoperability between other "feeder" information systems and the EHR in that they would be better enabled to "talk" the same language of archetypes when contributing information that should be signed into the EHR.

9. The approach facilitates multiple access control viewpoints on healthcare information by allowing users to access information according to their particular expertise.

4.3.2 The features of a generalised identity model

Whether identities are intended to represent demographic concepts (in the case of demographic services) or other entities, they are typically associated with one or more numerical or alpha-numerical identifiers. These identifiers are traits which can be used individually and conveniently by a software application to select a single unique identity or particular entity within an identity domain (OMG, 2000). In this thesis, the author adopt and extended a view of digital identity in healthcare that is derived from the ISO/IEC 24760-1's definition (ISO/JTC, 2011) as shown in Figure 4.2.

Identity could be considered as equivalence between a set of traits that allow humans to uniquely identify an entity and a set of identifiers that allow the corresponding entity to be identified automatically within their particular identity domains. Identifiers for the
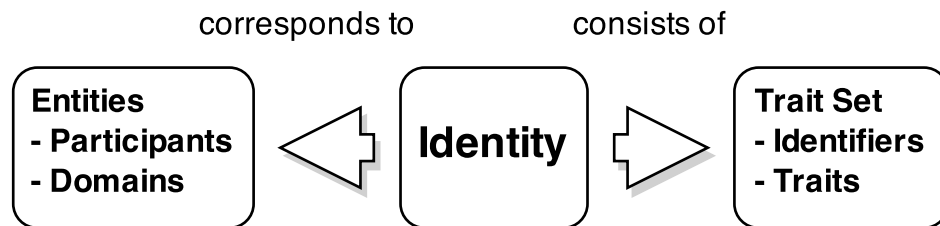
Figure. 4.2 Semantics model of identity in healthcare.

same entity can be called different names (e.g., MRN, Patient ID, and Subject of Care ID) in the identity domains of different organisations and they can adopt different formats.

As the author mentioned before in Chapter Two, the ontology-based referent tracking approach allows identity models to be extended to represent identities other than patients, health care professionals. Ceusters and Smith (Ceusters and Smith, 2006) suggest that it would be beneficial to incorporate entity types, such as disorders and body parts into the EHR. This more inclusive and unified use of identity would make these entities more identifiable and usable in health documentation.

Notwithstanding the overheads introduced by a two-level approach to identity, a more flexible approach to identity in the EHR will allow information in the EHR to be accessed and indexed (with appropriate safeguards) for secondary purposes. Indexing the record according to these additional identity contexts is likely to lead to efficient reuse of health information. A number of candidate identities for the electronic health record to support multiple identity and indexing contexts for EHR reuse. The ability to identify those entities relevant to a patient's health helps to manage the life-cycle of a healthcare related entity. Reliable identification of notable instances of health phenomena is increasingly important in the context of distributed multidisciplinary care where patients will receive care from different health professionals at different locations and times.

To guarantee successful identification of an entity such as a patient's fracture a set of suitably discriminated traits need to be captured within the EHR or directly in some external RTS. For example, for a fracture, one may need a reference to a bone ID, an X-ray

ID, and a time stamp. The essential identification traits need to be agreed by the health care professionals at each location where the entity is used. Trait sets may evolve as new technologies change the identification process or our knowledge about the entities evolve. New previously unknown entities will appear which require particular identification traits. In other words, an identity RM should have a capacity to express common features of existing demographic models but should be extensible with compatibility of Popper's "Three Worlds" and also recognise the referent tracking approach.

It is the author's aim to work towards a generalised RM that is consistent with referent tracking, that allows the identities to be defined by domain experts and also allows the referent approach to be optionally incorporated into EHR systems. While the referent tracking approach could assume the role of the clinical RMs for EHR systems, the approach that will be investigated in this work is to employ the flexibility of clinical RMs and define a simple identity model for entities in the EHR to which clinical data can be attached.

It is interesting at this point to speculate about some characteristics of a two level identity model. It is first of all used for the purposes of structuring information that is used in identification rather than structuring information that is used in documentation – the original use case of two level models. Secondly, only some properties of an identified entity could be considered as traits that would be used for the purposes of identifying that entity. Other properties, particularly those that change over time, are naturally part of the documentation process and would tell something about the current state of the entity rather than its identity. So similarly to the way in which data elements in an EHR are either for identifying a patient, or form part of the record that documents that patients care. This distinction between traits and other properties needs to be made as part of the identity modelling process and this will be discussed later in Chapter Five. Next, depending on the type of entity to be identified, the number of identifying traits will vary. In particular for "World 2" entities, the number of identifying traits will be low, due to the fact that the entity is not a real world entity and so possibly only has computer recognisable identifiers

and only a very small number of human recognisable traits. Therefore there is a possible argument for treating "World 2" identified entities in a different way - perhaps using a more lightweight approach of just naming the entity and allowing it to have identifiers.

Finally, there is the issue of healthcare which was mentioned earlier in Section 4.2.3 of data always needing to be referred to a subject of care. So even if information is originally recorded about, such a specimen as subject of information, by the time that the information is attested into the EHR, it is considered to be information about the subject of care. In some sense, in order to be consistent with EHR specifications, the primary identified entity in the original context is suppressed in favour of the patient as subject. Subsequently, in some secondary care scenarios, the patient may be merged into some anonymised demographic grouping which then becomes the primary subject of information. The generalised identity model in allowing the subject of information to change, while allowing the subject of care to be remembered at all times, enables information to be shared with or shared by an EHR.

4.3.3   Derivation of the GIRM from features of existing models

According to the investigation in Section 4.2.1, three of the existing standard demographics models (HL7 v2, ISO/TS 22220 Demographics and ISO/TS 27527 Demographics) could be described as static one-level models: they are expected to cover the entire conceptual space of their individual domains. For example, ISO/TS 22220 and ISO/TS 27527 are static information models describing the demographic information. Although these three examples are not particularly large or complex, the size and complexity of some models of this type can make them difficult to learn. They often include optional features to describe the domain of interest. They lead to distributed systems which can be difficult and costly to extend. They can produce multiple descriptions of the same concept. For this work, the data elements of both ISO/TS 22220 and ISO/TS 27527 are considered to be equivalent to types of *Trait*. While constituent parts of each data element are examples of Trait part. The use of *trait parts* will be described later in this chapter.

The general structure of the EAV/CR approach is compliant with the needs of a generalised RM but certain identity-specific features can be added so that the generalised identity model reflects its intended purpose. This basic form can then be constrained and extended by archetypes. This is somewhat analogous to adding "dynamic" EAV tables to a "static" relational database.

In addition to the constraint issues of models, the author compared and analysed the chosen models to select their common features for the proposed model as follows.

- The original IS model includes the concept of HL7 v3 *trait* (HL7, 2014).

- The EN13606 model includes the concept of *Name Part* and *Address Part*. HL7 data types use a similar approach, so the author adopt the idea of a *trait part*.

- EN13606 uses the concept of *Identified_Entity* which is very close to Entity as used in IS (ISO/TC 215, 2006; OMG, 2009a).

- EN13606 also includes the concept of the *validity time* of a trait, so it has been added to identity components.

- There is an increasing emphasis on strong data typing in health information models. This is supported by making trait parts contain a EN13606 or ISO 21090 data type.

- Both OpenEHR and HL7 incorporate a similar concept for role, link for their entity class.

- Following the style of EN13606 in order to make certain identity components archetype-able, it is useful to add *name* and *archetype id*.

Based on a study of selected demographics and identity models, Table 4.2 shows the essential aspects of GIRM. The GIRM is designed to allow exchange of identity data using XML based messages that are compliant with different standards (e.g., systems based on ISO/TS 25757 and ISO/TS 22220). Figure 4.3 shows a class diagram for the GIRM

prototype with CEN/TS 14796 data types (CEN/TS, 2004), that are referenced by the ISO EN13606 (ISO/TC 215, 2006) specification. Properties or elements can be added based on a generic example. For instance, a sequence number may be an optional property of the *TraitPart* class. Certain complex data types are required to be incorporated into the GIRM according to implementation requirements, for example, simple text, coded text, Uniform Resource Identifier (URI), Object Identifier (OID) and Instance Identifier (II) (CEN/TS, 2004).

Table 4.2 Observation of the characterisation of the GIRM

Characteristics of identity model	How handled in GIRM	Standards using feature
Identity	RM class	EN13606, OpenEHR
Identity component	RM properties / data type	EN13606, OpenEHR
Archetype id	RM properties / data type	EN13606
Trait	EAV with meta data model (HL7, 2014)	IS, EN13606, ISO/TS 22220, ISO/DTS 27527
Trait part	RM class	EN13606, ISO/TS 22220, ISO/DTS 27527
Validity time	RM property / data type	EN13606, OpenEHR, IS, HL7
Replacing version	RM property / data type	IS, OpenEHR, HL7
Unique identifier	RM property / data type	EN13606
Identity link	RM class / ontology-based	OpenEHR, HL7, IS
Data value	RM property / data type	EN13606, OpenEHR, HL7, IS

Consequently, the main features of GIRM according to Figure 4.3 are defined as follows.

Main classes in the GIRM schema:

1. *Identity* is instance of an identity that corresponds to the identified entity of EN13606. An identity may contain zero Traits when it is initiated.

2. *Identity component* - is an abstract and super class of Identity and Trait.

3. *Trait* is a characteristic of an identity such as name, address and assigned after within certain domain. Trait is either a complex trait or simple trait (with the Data value). For preserving information integrity of a complex trait, a trait part therefore can not be a simple trait in the design of the GIRM schema.
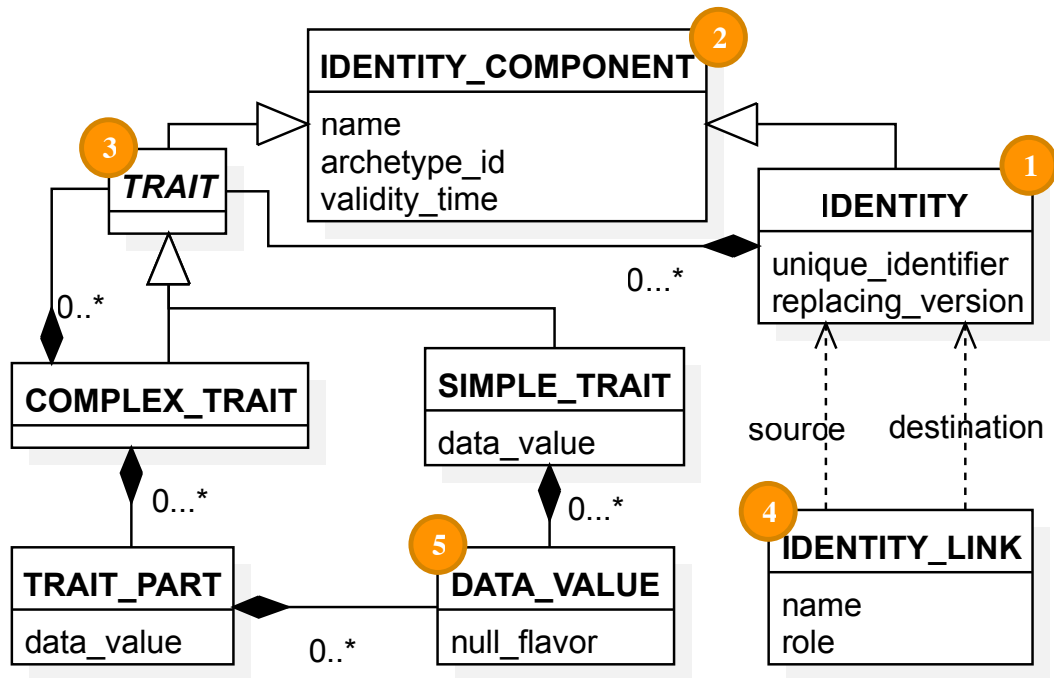
Figure. 4.3 The class diagram of the GIRM prototype

4. *Identity link* provides a functional role for the linkage between identities.

5. *Data value* is the container for an actual value in the RM and it may contain a single value that is defined by the complex data type or null attribute.

Main properties in the GIRM schema:

- *Name* is used to define the responsibilities undertaken by a class including the archetype-able class based on its competency.

- *Archetype id* refers to the unique identifier of the archetype.

- *Trait part* represents the details of a trait. It can be used to contain any numbers or texts if a trait has additional descriptive attributes.

- *Validity time* is the valid time interval for the definition of trait or trait set.

- *Replacing version* means that identities are always versioned when created.

- *Unique identifier* is required for uniquely identifying the identity.

The design of a generalised two-level identity RM facilitates the identity management in multiple identity contexts as well as the cross-referencing of heterogeneous identity domains. The design has flexibility and expressive power that can be extended and reused by two level modelling. Data types are a prerequisite for the application of reference models. There may even be a need for certain systems to support multiple sets of data types. (For example, one set of data types for persistence, another for communication). A joint ISO/CEN initiative to develop harmonised datatypes resulted in early 2011 in the successful ballot of a set of harmonised data types (ISO/TC 215, 2011a). These harmonised data types are intended to bridge differences between data types that are consistent with the two level model approaches of openEHR and EN13606 and data types that support HL7 v3. Unfortunately however, this harmonised set of data types is not as compact as the original CEN datatypes that were used satisfactorily in this work.

4.4 The evaluation strategy of the GIRM

The previous sections have described the design for the identity modelling and structure of the GIRM. In this section the author will discuss how the benefit and expected cost of the GIRM will be measured, and how to factor these results into an overall evaluation plan. Then the evaluation of the GIRM approach will proceed in the following chapters.

For validating the GIRM approach with the features mentioned in Section 4.3.2, the following questions[1] should be answered by the evaluations.

Q.1 How does the complexity of two-level models compared to the complexity of fixed models?

Q.2 How should the GIRM be used alongside a two-level modelling tool to replicate the existing demographics information models with archetypes?

---

[1]The symbol "Q" denotes "Question".

Q.3 Can the GIRM be used to extend the existing demographics models by for instance adding traits into the existing identified entities?

Q.4 How can the GIRM be implemented to achieve identity interoperability by an example of exchanging instance messages between two different standard-based demographics models (EN13606 and HL7)?

In order to answer questions 1 to 4, the initial evaluation is setup on the basis of feasibility, effectiveness and consistency of the GIRM creation process in Section 5.2.1 of Chapter Five. The further evaluation leads to question 5 which will be encountered in Section 4.4.3 and focuses on the suitability of exchanging GIRM-based identity information in Section 5.3 of Chapter Five.

4.4.1   Initial evaluation of the GIRM

Figure 4.4 illustrates the initial evaluation process of using the GIRM to construct existing demographics RMs and extend use of identifiers with trait sets to categorise and identify the other entities in healthcare including parts of the human body, samples, diagnostic devices and pharmaceutical products. The details of the initial evaluation plan are described in relation to the questions and the following investigations.

Investigation 1: The author will evaluate the GIRM by expressing and extending the demographics reference models at the meta data level, while demonstrating how archetypes can be constructed to produce structures which are similar to the selected reference models. It is expected that the GIRM has a high level of abstraction and is simple to instantiate to create GIRM-derived archetypes that are quite complex in modelling clinical information.

Investigation 2: The data entities and data elements of existing demographics models (e.g., EN13606, ISO/TS 22220, ISO/TS 27527) will be analysed and expressed using a generalised identity reference model. In this work, the GIRM-derived reference instances
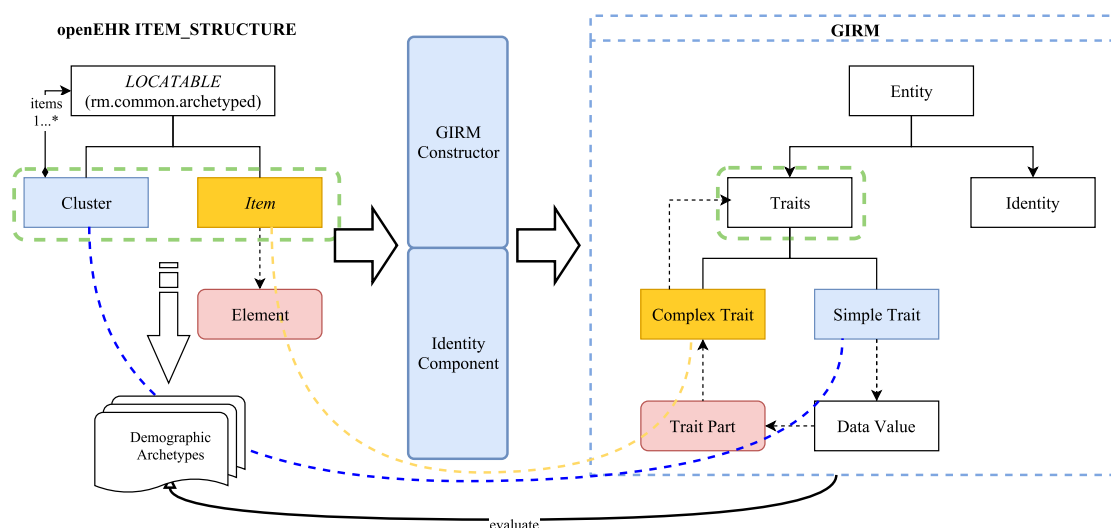
Figure. 4.4 GIRM construction using features of the existing demographics models

were represented by XML schema. This assesses the potential and effectiveness for extension of the experimental models using archetypes.

Investigation 3: A set of archetypes will be created that fully cover all the entities of the CEN demographics model, involving e.g., organisation, person identification, software or device to validate use of the extensibility of the GIRM. The evaluation will show that the GIRM-derived archetypes decompose the demographics information into *complex_trait*, *trait_part* and also the recursive aggregation feature to enable more flexibility when constructing the EN13606 demographics entities.

Investigation 4: A set of demographics reference instances and archetypes based on the chosen standards (ISO/TS 22220 and ISO/TS 27257) will be created using the GIRM to validate suitability of its use in modelling demographics information from other "fixed" models.

After that, the results of the initial evaluation comprising investigations 1 to 4 will be described and explained in relation to the performance of the GIRM creation process.

4.4.2 Evaluating the instance of the GIRM for exchanging identity information

Application 1: Figure 4.5 shows the further evaluation of generalised identity reference model at the meta data level as a step towards exchanging archetype-constrained identity information between diverse identity domains. This evaluation process will be run through the selected demographics models which have been investigated in Section 4.2.1. The message exchanging prototype of the demographics instances between HL7v2 PID and EN13606 Person Identification was developed. This prototype should demonstrate good query capabilities. The whole process is represented as a workflow diagram with step by step descriptions.

4.4.3 Application of the GIRM for identifying laboratory information

Following the evaluation of the GIRM, the author believes that adopting this generalised identity modelling approach to identify clinical laboratory information is particularly useful according to the author's experience. This investigation is therefore considered to be an important part of the overall evaluation. The author will utilise the GIRM approach to help to extend identity management in the domain of clinical laboratory medicine.

Some questions to be answered within the evaluation are also summaries as follows.

Q.5 How should an information model be arranged to allow laboratory information that is destined for the electronic health record to be recorded about the patient, but also about the specimen and the order?

Q.6 How can the GIRM be used alongside archetypes and two-level modelling tools to construct the identity model to support documentation about subject other than patient and healthcare professional?

```
        ┌──────────┐         ┌──────────┐
        │  Legacy  │         │   GIRM   │
        │    RM    │         │          │
        └──────────┘         └──────────┘
             │                     ┊
      expressed using           create
             ▼                     ▼
        ┌──────────┐         ┌──────────────┐
        │ RM schema│         │ GIRM mediated│
        │          │         │    schema    │
        └──────────┘         └──────────────┘
              \                 /
               \  normalise    /
                ▼             ▼
            ┌──────────────────┐
            │ Reconstructed RM │
            │     schema       │
            └──────────────────┘
                     │
                 validation
                     ▼
            ┌──────────────────┐
            │ Archetype-based  │
            │     schema       │
            └──────────────────┘
                     │
               to instantiate
                     ▼
            ┌──────────────────┐
            │    Extended      │
            │    archetype     │
            └──────────────────┘
```
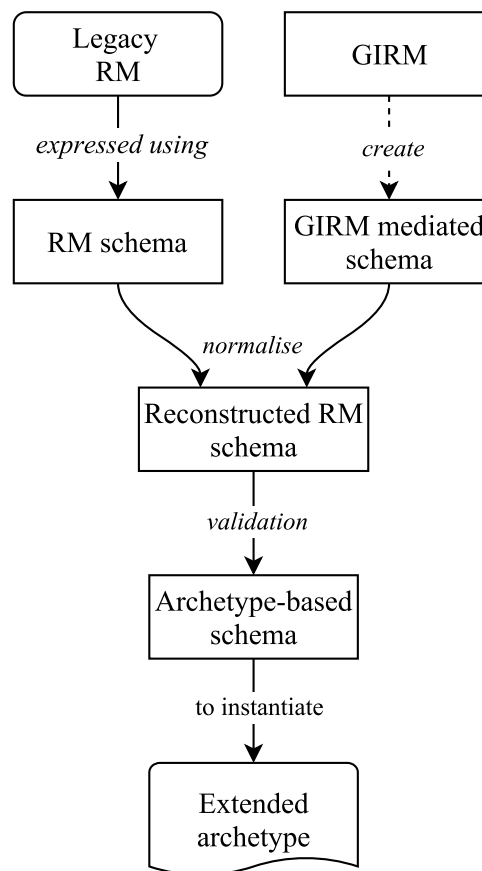
Figure. 4.5 Flowchart of using GIRM mediated schema to extend demographics RM

Q.7 How to decide which properties are *identity properties* and which are *record properties*?

This stage of the evaluation is required to test on utilisation of the GIRM approach to demonstrate how a generalised two level model that has been previously developed and validated by the author, can be applied to specimen, order, medical device or other candidate identified entities. This requires an investigation of the relationship between the differing perspectives of clinicians, lab professionals, and patients. For example, one idea of the application is to allow medical device as a "patient" for instrument vendors. The GIRM approach can be utilised for instance, to provide an "instrument record" to keep the medical device "well" so it can "do its job" as a medical device. The example application will

demonstrate how to apply two level approaches by applying a generalised identity model to a LIMS while using existing defined archetypes. The approach will be shown to involve the following investigations.

Investigation 5: Analysis, of the data elements and data entities of interest for laboratory based information systems and for the messages that pass between them. The application will show difference and similarity between the existing laboratory information models (90101 vs ASTM).

Application 2: Mapping/overlapping between the fields of specimen/ instrument using 11073 and ASTM. In this example application, provision of a set of archetypes were provided that fully cover all the fields of certain specimen, order/request and instruments in these scenarios using generalised two level model within LIMS. It will also demonstrate the principal distinctions between identity information and clinical information that is derived from the idea of the specimen as the "patient".

Application 3: An example application of the extent of use of the GIRM that were not based on patient and healthcare professional between the information systems (legacy database) and archetypes relating to specimen/instrument using 11073/openEHR will be presented. The example application will demonstrate that meta models of surveyed open source laboratory information systems are extracted to create the GIRM-based archetypes for identifying the clinical identity information.

4.5   Summary

As a first step towards achieving semantic interoperability between two-level model-based EHR systems, this chapter has described the investigation and derivation procedures for a generalised identity reference model based on the two-level modelling approach. The author has applied two-level modelling approach to improve the integration of identity reference models into EHR systems. This modelling approach was created to assist in the

"baseline" task of identity modelling in the presence of differing underlying demographic models or schemas where one schema employs two-level models and archetypes. The evaluation of the GIRM in the next chapter will show the validity and the extensibility of two-level identity reference models.

EVALUATION OF THE TWO-LEVEL IDENTITY APPROACH

5.1 Overview

Now that the rationale for the GIRM has been described along with the characteristics of the GIRM itself, it is instructive to investigate how the GIRM itself, or identity models with similar features might be used and attempt to demonstrate the added utility of this approach to identity management.

5.1.1 Initial evaluation

The author anticipates that the GIRM will provide a more general approach to linking documentation to subject of information, which can give system developers a generalised identity model on which to identify the demographics identities or indeed other entities, to add domain specific details by domain experts in more concrete archetype models and make it possible for these identities to be the subject of information of shared information in health information systems other than the EHR particularly those that contribute information to, or accept information from the EHR. This chapter will describe evaluation of these aspects of the GIRM as follows.

Firstly, the author will evaluate the feasibility and consistency of the GIRM which has been described in Chapter Four. The initial evaluation of the GIRM using mediated schema approach for exchanging demographics information is presented in Section 5.2.1, which describes the generalised model to be implemented in EHR system by the measurement of complexity metrics, abstraction metrics and details at archetype level. The evaluation will show its use for standards-based two-level modelling alongside legacy systems by expressing details of EN13606 demographics model in an extensible way. A set of archetypes to describe two relevant ISO "fixed" model-based specifications (ISO/TS 27527, ISO/TS 22220) are created using the LinkEHR archetype tool (Maldonado *et al.*, 2009) showing that the GIRM can be used for the exchange of archetype-based EHR identities.

5.1.2 Further evaluation

Secondly, the author will evaluate the practicability of the GIRM by applying the GIRM to LIMS in Chapter Six. The example applications will show that information systems other than the EHR require more flexibility with respect to "identified entities". To do this, the author will demonstrate that this approach can provide a highly flexible and expressive means for representing subjects of information by the sample sets of data elements and entities from working information systems in the laboratory domain. The applications will also demonstrate reuse of relevant published standard-based archetypes.

There are also the secondary uses of healthcare data and the wider health domain that includes clinical research, population health, management and other activities, that may involve secondary use of healthcare information. This wider scope includes additional identified entities - researchers, population groups / regions. It has been noted that the perspective of allied health professionals such as the laboratory professional is different from that of the clinician. While clinicians focus on diagnosing and treating patients, laboratory professionals are concerned with processing laboratory specimens to produce

laboratory results and reports. It has been argued that the primary "subject of care" in clinical laboratories is the specimen. To maximise data quality it is important to recognise the independence of the specimen from a laboratory perspective. This is particularly true in cases where patient identity is withheld for security, privacy, or technical reasons. In this case, the specimen becomes at least for some ICT stakeholders that is the primary identified entity within laboratories.

Actually, as the reader shall see later, another very important concept in the laboratory world is the order, an entity that is not even from the physical world but is part of Popper's "World Two". A similar case can be made for many of the list of identified entities that were identified current EHR information models in following the user requirements of their primary users; the clinicians do not take account of this more general perspective on identity.

Unlike existing approaches, the GIRM is not specific to demographics but can also express other identity types in healthcare (which has been discussed in Chapter Four). A piece of evidence of documentation in the healthcare domain that does not focus on the patient, but yet end up contributing information to the electronic health record. Examples of "identified entities" that accumulate documentation in the laboratory domain include "specimen / sample", "order", and possibly "medical device".

5.2 Evaluating the generalised RM approach

Of course arguably the key requirement of the EHR also needs to be kept in mind. That information recorded in the EHR is about the *patient*. Even if in its original setting this information was recorded about a specimen, medical image or pharmaceutical product, and after being incorporated in the EHR, in secondary use, the same health information is to be associated with a group of patients or with a research site who are overseeing the

group. The fact is that modern shared care requires the patient to be the central subject of health information in the EHR but not the only subject of information.

5.2.1 Methods of the initial evaluation

Today in many countries, healthcare organisations must develop their own identification systems and separate identification domains with the continued introduction of shared care, and where many jurisdictions, the issue of managing and matching multiple identities for patients and health professionals is still an important one. And with thousands of legacy systems and the complexity of demographics across regions, there is little chance of a single convenient and global "fixed" demographics model. So flexibility in demographic representation can be useful in these environments.

A formal two-level model of generalised identity which has been proposed in the previous chapter conforms to the EAV-style "design pattern" for flexible addition of properties and relationships which are nevertheless constrained by the rules imposed by the GIRM reference model itself. This evaluation will show that GIRM permits the reconstruction of existing demographics reference models that protect the unique relationship between patients and health professionals but allows other perspectives and identities to be represented. The evaluation process is designed to assess investigations 1 to 4 (which has been described in Section 4.4) for purposes of alignment with the GIRM construction process in Figure 4.4 of Chapter Four.

5.2.2 *Investigation 1*: Measuring the level of abstraction and constraint of GIRM

The demographics models of five specifications which have been investigated in Section 4.2.1 does raise a number of issues related to managing heterogeneous data however, and so this investigation will continue by compare them using elements of Graph theory. Graph

theory has been widely used for assessing modelling of complex information systems (Christensen and Albert, 2006).

For this assessment, the author employs a model driven development (MDD) approach to measuring the level of abstraction and detail of the different types of identity reference models, which can be used for evaluating the efficiency of semantic interoperability of the model represented by UML class diagrams (Nugroho and Lange, 2008; Pilgrim, 2007). The seven selected EHR reference models each exist at different levels of abstraction and so in the author's view provide a representative sample of existing demographics metadata. They are CEN EN13606 Demographics (ISO/TC 215, 2006), OpenEHR Demographics (Beale *et al.*, 2006), HL7 RIM v2.4 (HL7, 2015), ISO/TS 22220 (ISO/TC 215, 2011b) and ISO/DTS 27527 (ISO/TC 215, 2010) and two other identity models GIRM and IXS v2 which can also be used for representing demographics information at this stage.

The measurement is taken by using Graph theory where graph $G = (V, E)$ with number of vertices V and number of edges E. The definition in Equation 5.1 of semantic size (*ss*) is the number of all attributes (*A*) of the underlying model. This parameter is representative of the "level of abstraction" of the model (Pilgrim, 2007).

$$ss(M) = |A|^1 \tag{5.1}$$

The size of model reflects the "level of constraint" is defined in Equation 5.2 by the sum of

- the number of vertices V of the model, the number of classes in the UML class diagram,

- the number of edges E, the number of pairs of these vertices,

- and the graph itself (count one) to avoid a division by zero error.

---

[1]"|A|" denotes the number of attributes in A.

$$size(M) = |V| + |E| + 1^2 \tag{5.2}$$

In the measurement approach of (Pilgrim, 2007), the ratio of abstraction and constraint (rac) of a model (M) is defined in Equation 5.3 as the semantic size (*ss*) of a model (M), divided by its size (*size*).

$$rac(M) = \frac{ss(M)}{size(M)} \tag{5.3}$$

Table 5.1 provides the count results of each of these 7 models with the ratio of abstraction in the last column[3]. The results support the view that two-level models (GIRM, openEHR, EN13606) have less (10, 15, 30) attributes than the other models resulting from their high levels of abstraction. And resulting the size of model (74) of ISO/TS 22220 is highest proportion of all models which illustrates the high levels of constraint due to its domain-specific scopes but it has low flexibility as a fixed model. The resulting, ISO/TS 27527 has a modest figure (21) for the size of model compared to ISO/TS 22220 but it has the most 67 attributes as a cost of "fixed" model, which agrees with the analysis of disunity between abstraction and granularity in modelling.

Table 5.1 The results of counting the ratio of abstraction

Model	No. of Vertices	No. of Edges	No. of Attributes	Size of Model	Ratio of Abstraction
GIRM Universal	8	11	10	20	0.50
EN13606 Demographics	5	13	30	19	1.58
openEHR 1.0.2 Demographics	10	12	15	23	0.65
Hl7 v3 Entities	10	9	52	20	2.60
ISO/TS 22220	37	36	54	74	0.73
ISO/TS 27527	9	11	67	21	3.19
IXS v2 Entities	6	6	33	13	2.54

[2]"+" denotes the sum of the numbers.
[3]The final result of ratio of abstraction is shown with two digit after the decimal point.

As a result, Figure 5.1 shows a selection of prominent current identity models in increasing level of constraint. Recalling that level of abstraction is calculated according to equation 5.1 by measuring the number of attributes in each model. For example, the ISO/TS 22220 is a one level model that has less flexibility but has more details compared to two level models such as the GIRM proposed in this chapter and the openEHR Demographics model (Beale *et al.*, 2006). However, it is worth mentioning that these results reveal the ratio of abstraction a limited way as these models are fairly complicated to be precisely represented by using the UML class diagrams (Steimann and Vollmer, 2006). This investigation is intended to answer Q1.
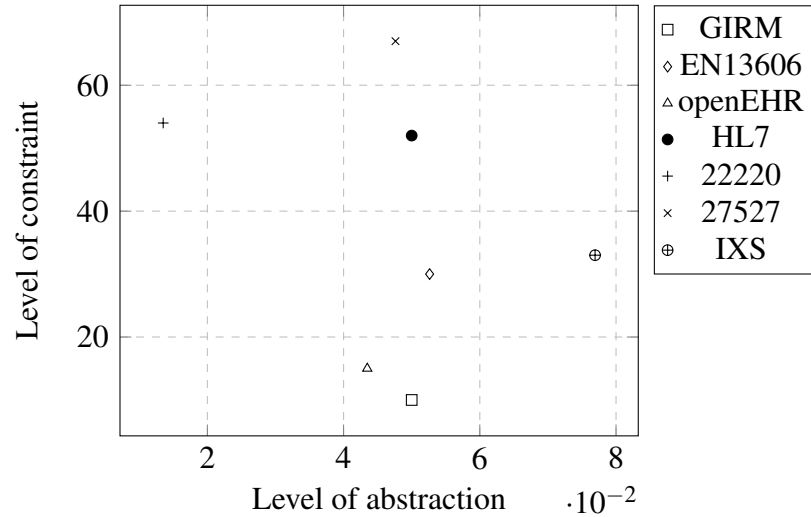
Figure. 5.1 Visualised results of abstraction and constraint in selected demographics models

5.2.3 *Investigation 2*: Expressing RM to support existing demographic models

In Figure 5.2, the schema representation of the GIRM is created to support an archetype editor "LinkEHR" (Maldonado *et al.*, 2009) which has been introduced in Chapter Three. For this work, the author has used the LinkEHR tool to create archetypes from the user defined schemas of GIRM which consist of the identified entity, the identity component
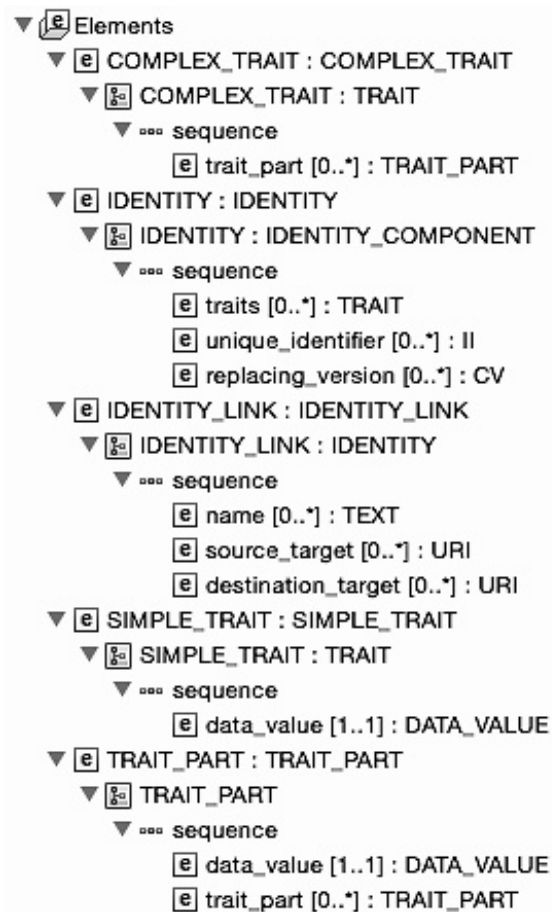
Figure. 5.2 The outline of the GIRM schema representation.

and the datatypes in support of CEN/TS 14796 (see the full XML schemas in Appendix A).

The rationale for archetype creation by using LinkEHR editor is that this tool supports import of created XSD based RM into the tool. This RM definition can be used to configure a set of the mandatory settings for creating a desired archetype that conforms to the GIRM mediated schema. After passing the tool's validation procedure for importing the user defined reference model, it is therefore a simple matter to create an archetype for example "Person" of EN13606 demographics model or indeed a GIRM compliant archetype. All the user needs to do is to select the required reference model from a drop down list when the archetype is first created. This investigation has described how the GIRM can be used with LinkEHR tool to replicate the original demographics models so answering Q2.

5.2.4 *Investigation 3*: Making EN13606 demographics extensible using GIRM

However, in particular, as mentioned in previous chapters, despite work in progress, the currently published the EN13606 demographics RM at time of writing is a static reference model which cannot be archetyped with flexibility in its current form which only includes four identified entities. Consequently it is not possible to extend this standardised model to permit new types of *identified entity* to which *traits* can be assigned. Figure 5.3 shows all concepts of identified entities in EN13606 demographics model that will be fully covered by the evaluation. The author constructed the original EN13606 demographics model using the GIRM and add a flexible placeholder with common features that is shown in Table 4.2 to express the different types of demographic entity.
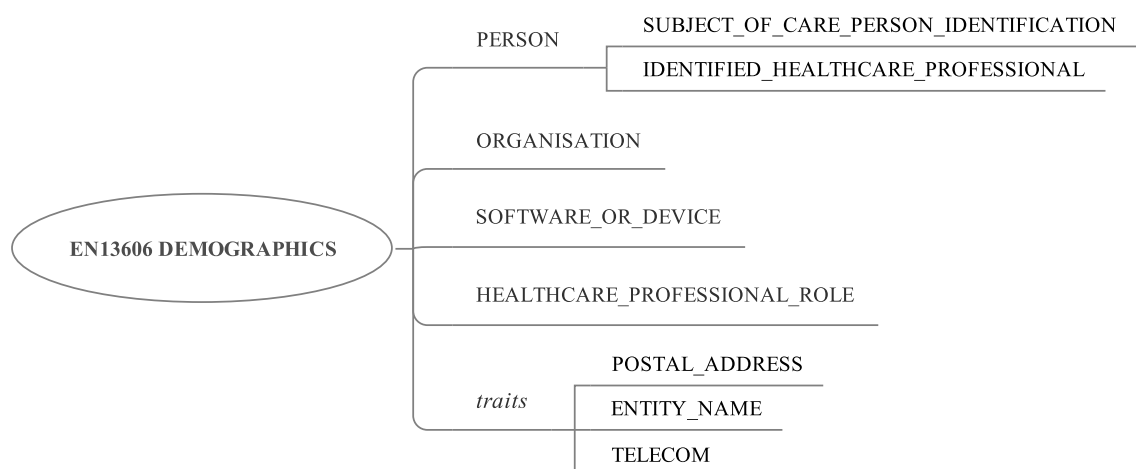
Figure. 5.3 The concepts of EN13606 demographics model

The raw codes of XML schema given in Listing 5.1, demonstrate that the CEN demographics RM can be constructed using the GIRM mediated schema to represent the *Entity_Name*. The fragment between line 2 and line 10 allows the EN13606 demographic model to be extended and archetype-able for implementing a two-level model which nevertheless has a fixed portion to add other types of the identities.

Listing 5.1 The raw XML schema of extended EN13606 Demographics RM

```
1  <!-- Generalised IDENTIFIED_ENTITY GROUP -->
2  <xs:element name="THIRD_PARTY_ENTITY" type="IDENTITY_COMPONENT" />
3  <xs:complexType name="THIRD_PARTY_ENTITY" abstract="true">
4    <xs:sequence>
5      <xs:element name="identity" type="IDENTITY" />
6      <xs:element name="identity_link" type="IDENTITY_LINK" />
7      <xs:element name="complex_trait" type="COMPLEX_TRAIT" />
8      <xs:element name="simple_trait" type="SIMPLE_TRAIT" />
9    </xs:sequence>
10 </xs:complexType>
11 ...
12 <!-- ENTITY_NAME -->
13 <xs:complexType name="ENTITY_NAME">
14   <xs:sequence>
15     <xs:element name="use" type="CV" />
16     <xs:element name="valid_time" type="IVLTS" />
17     <xs:element name="name_part" type="ENTITY_NAME_PART"
18       maxOccurs="unbounded" />
19   </xs:sequence>
20 </xs:complexType>
21 ...
```

As a further example, the author assessed feasibility of extension using the GIRM to facilitate a more flexible definition of identity in the EN13606 Demographics model. In Listing 5.2, a part of the ADL codes shows how a GIRM-derived archetype decomposes the original demographics information into complex trait and trait part that enables the addition of other subjects of secondary identities such as the body parts. In other words, the extended concept of "entity name" archetype of EN13606 can be employed to add more demographic information such as language, description, history. This investigation is intended to answer Q3.

Listing 5.2 A raw ADL file of EN13606 "SoC Person Identification'" archetype

```
1  COMPLEX_TRAIT[at0007] occurrences matches {0..*} matches {  -- entity_name
2    trait_part existence matches {0..1} cardinality matches {0..*; unordered; unique} matches {
3      TRAIT_PART[at0008] occurrences matches {0..*} matches {  -- entity_name_part
```

```
 4            trait_part existence matches {0..1} cardinality matches {0..*; unordered; unique} matches {
 5                TRAIT_PART[at0009] occurrences matches {0..*} matches {*}  -- entityPartName
 6                TRAIT_PART[at0010] occurrences matches {0..*} matches {*}  -- namePartType
 7                TRAIT_PART[at0011] occurrences matches {0..*} matches {*}  -- namePartQualifier
 8            }
 9        }
10        TRAIT_PART[at0012] occurrences matches {0..*} matches {*}  -- use
11        TRAIT_PART[at0013] occurrences matches {0..*} matches {*}  -- validTime
12    }
13 }
14 ...
```

5.2.5 *Investigation 4*: Building GIRM archetypes for the demographics model

As an example of building archetypes derived from ISO/TS 22220, Fig. 5.4(a) shows the "*SoC Name*" of the original ISO/TS 22220 model. A portion of the ISO/TS 22220 model is created in Figure 5.4(b) constructed by the GIRM in order to build the archetypes for ISO/TS 22220. The representation of "*Subject Of Care Name*" is created in the same way in order to constrain the demographics information of ISO/TS 22220 using the GIRM in the object diagram of Figure 5.4(b). For example, the name title group in ISO/TS 22220 is a complex trait and is required to have two sub-attributes, *Name Title*, *Name Title Sequence Number* as its *trait parts*. Listing 5.3 presents a raw ADL file of definition of *IDENTITY.SoC_Name* archetype constrains the demographics information of ISO/TS 22220 has been instantiated from the GIRM-based EN13606 RM as shown in Figure 5.4(b).

In the last step of constructing EN13606 demographics RM using the GIRM schema, the archetypes are mapped to XML schema to create an exchangeable identity instance to support a legacy EHR system. In the sample application that is developed to evaluate the work, the author has employed a Java open source XQuery implementation Saxon (Kay, 2008) to process the XML-based mapping between the GIRM-based schema and

(a) "SoC Name" class of ISO/TS 22220

(b) "SoC Name" transformation schema using GIRM
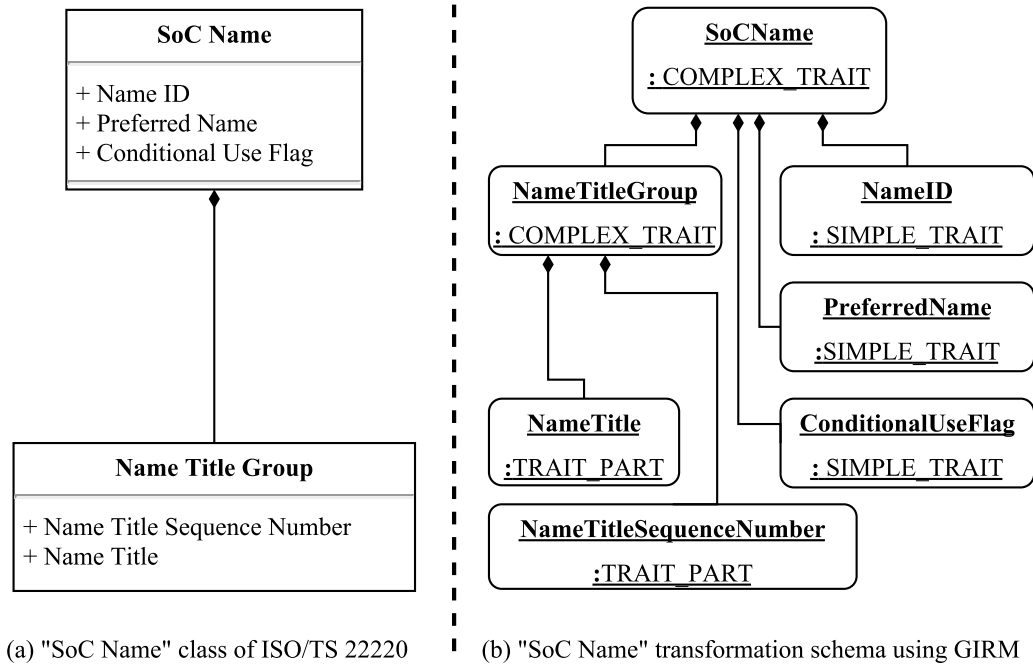
Figure. 5.4 The example of transformation of "SoC Name" by using GIRM.

Listing 5.3 The EN13606 "SoC Name" archetype with constraints of GIRM

```
1  IDENTITY[at0000] occurrences matches {1..1} matches { -- SoC Name
2    traits existence matches {0..1} cardinality matches {0..*; unordered; unique} matches {
3    SIMPLE_TRAIT[at0001] occurrences matches {0..*} matches {*} -- Name ID
4    SIMPLE_TRAIT[at0002] occurrences matches {0..*} matches {*} -- Preferred Name
5    SIMPLE_TRAIT[at0003] occurrences matches {0..*} matches {*} -- Conditional Use Flag
6    COMPLEX_TRAIT[at0004] occurrences matches {0..*} matches { -- Name Title Group
7    trait_part existence matches {0..1} cardinality matches {0..*; unordered; unique} matches {
8      TRAIT_PART[at0005] occurrences matches {0..*} matches {*} -- Name Title Sequence Number
9      TRAIT_PART[at0006] occurrences matches {0..*} matches {*} -- Name Title}}
10   ...
```

the archetype schema. It should be noted that archetype mapping and normalisation algorithms are out of the scope this work. The Poseacle project at university of Murcia does a lot of work on the development of those algorithms and tools that support mapping and data transformation between XML data sources and archetypes (Maldonado *et al.*, 2008).The example of the mapping creation for "*SoC Name*" are shown in Listing 5.4. The investigation provides further answer to Q2.

Listing 5.4 A sample mapping file of the GIRM schema and the archetype schema

```
1  /traits/SIMPLE_TRAIT[at0001] <-> <xs:element name="name_id" type="xs:string" minOccurs="0" />
2  /traits/SIMPLE_TRAIT[at0002] <-> <xs:element name="preferred_name" type="xs:string" minOccurs="0" />
```

```
3  /traits/SIMPLE_TRAIT[at0003] <-> <xs:element name="conditional_use_flag" type="xs:string" minOccurs="0" />
4  /traits/COMPLEX_TRAIT[at0004] <-> <xs:complexType name="NAME_TITLE_GROUP">
5  /traits/COMPLEX_TRAIT[at0004]/TRAIT_PART[at0005]
6  /traits/COMPLEX_TRAIT[at0004]/TRAIT_PART[at0005]
7  <->
8  <xs:complexContent>
9      <xs:extension base="IDENTIFIED_ENTITY">
10         <xs:sequence>
11             <xs:element name="name_title_sequence_number" type="xs:string" minOccurs="0" />
12             <xs:element name="name_title" type="xs:string" minOccurs="0" />
13         </xs:sequence>
14     </xs:extension>
15 </xs:complexContent>
```

5.2.6 The results of extended EN13606 demographics RM using GIRM

In a domain as complex as healthcare, the difficulties of modelling a large number of clinical concepts usually arise in a single-level modelling approach. The results show that single level (fixed) information models are complex but they can also employ an abstract model with lots of optionality. In the author's opinion, the high level of abstraction makes such models quite difficult to comprehend and implement. The fixed nature is also insufficient to allow domain experts to be directly involved in defining the semantics of clinical information system and it has less interoperability in the healthcare domain compared to the existing two level models such as openEHR.

In addition to looking at the feasibility of using the GIRM schema to extend the EN13606 demographics model, the level of constraint of the identified entities of the EN13606 demographics model were counted and the results of counting were compared as shown in Table 5.2. As a result, the level of constraint of the GIRM-based EN13606 demographics RM increases noticeably compared to original demographics RM by lifting the proportion of the size of model on their entire models to 242%.

The datatypes used in the GIRM schema employ the ISO/TS 14796 which has the datatypes (CS, CV, TS, QUANTITY, IVLTS, II, TEXT, URI, CODE, SIMPLE_TEXT,

DURATION) (CEN/TS, 2004). The evaluation only uses the minimum set of datatypes required by GIRM schema and it allows the creation and reuse of the complex datatypes just like reusing the archetypes.

Table 5.2 The level of constraint for each instance of extended EN13606 demographics RM

EN13606 demographics using GIRM	No. of Vertices	No. of Edges	Size of Model
Original:			
Organisation	5	4	10
Identified_healthcare_professional	9	8	18
Subject_of_care_person_identification	8	7	16
Software_or_device	6	5	12
Extended:			
Organisation	14	13	28
Identified_healthcare_professional	23	22	46
Subject_of_care_person_identification	22	21	44
Software_or_device	9	8	18

The set of the GIRM-based demographics archetypes that extended the EN13606 demographics model was shown in Table 5.3. In other words, the GIRM features were added into an EN13606 demographics model and this adapted reference model was loaded into LinkEHR as an XSD rendered RM. The EN13606 extended archetypes were created to conform to all entities within the EN13606 demographics model but with the flexible slots to support the demographics details from ISO/TS 22220 and ISO/TS 27527. The new GIRM archetypes that were developed based on ISO/TS 22220 and ISO/TS 27527 complement the EN13606 demographics model by creating a new flexible modelling extension. The results show that constructing the GIRM by extending the EN13606 demographics model is feasible way to fully cover all the elements in ISO/TS 22220 and ISO/TS27527 that are identified through this example. The overall initial evaluation according to the investigations 1 to 4 provides the answer to Q3.

Table 5.3 Summary of extended archetypes of the EN13606 Demographics to support ISO/TS 22220 and ISO/TS 27527

Demographics Archetypes	Source
Extended archetypes using the GIRM:	
CEN-EN13606-DEMOGRAPHIC-Entity_Name.v1.adl	EN13606
CEN-EN13606-DEMOGRAPHIC-Healthcare_Professional_Role.v1.adl	EN13606
CEN-EN13606-DEMOGRAPHIC-Identified_Healthcare_Professional.v1.adl	EN13606
CEN-EN13606-DEMOGRAPHIC-Individual_Professional.v1.adl	EN13606
CEN-EN13606-DEMOGRAPHIC-Organisation.v1.adl	EN13606
CEN-EN13606-DEMOGRAPHIC-Person.v1.adl	EN13606
CEN-EN13606-DEMOGRAPHIC-Postal_Address.v1.adl	EN13606
CEN-EN13606-DEMOGRAPHIC-Software_or_Device.v1.adl	EN13606
CEN-EN13606-DEMOGRAPHIC-Subject_of_Care_Person_Identification.adl	EN13606
New archetypes using the GIRM:	
ISO-TS22220-IDENTIFIED_ENTITY-Additional_Demographic_Data.v1.adl	ISO/TS 22220
ISO-TS22220-IDENTIFIED_ENTITY-SoC_Address.v1.adl	ISO/TS 22220
ISO-TS22220-IDENTIFIED_ENTITY-SoC_Biometric_Identifier.v1.adl	ISO/TS 22220
ISO-TS22220-IDENTIFIED_ENTITY-SoC_Electronic_Communication.v1.adl	ISO/TS 22220
ISO-TS22220-IDENTIFIED_ENTITY-SoC_Identifier.v1.adl	ISO/TS 22220
ISO-TS22220-IDENTIFIED_ENTITY-SoC_Linkage_Identifier.v1.adl	ISO/TS 22220
ISO-TS22220-IDENTIFIED_ENTITY-SoC_Name.v1.adl	ISO/TS 22220
ISO-TS27527-IDENTIFIED_ENTITY-Individual_Provider_Demographic_Details.v1.adl	ISO/TS 27527
ISO-TS27527-IDENTIFIED_ENTITY-Individual_Provider_Field_of_Practice.v1.adl	ISO/TS 27527
ISO-TS27527-IDENTIFIED_ENTITY-Health_Care_Provider_Organisation.v1.adl	ISO/TS 27527
ISO-TS27527-IDENTIFIED_ENTITY-Individual_Provider_Name.v1.adl	ISO/TS 27527
ISO-TS27527-IDENTIFIED_ENTITY-Provider_Address_Details.v1.adl	ISO/TS 27527
ISO-TS27527-IDENTIFIED_ENTITY-Provider_Electronic_Communication.v1.adl	ISO/TS 27527
ISO-TS27527-IDENTIFIED_ENTITY-Provider_Identifier.v1.adl	ISO/TS 27527

5.3   *Application 1*: Exchanging archetype-constrained identity data

An overview of the message exchange process at an instance level has been presented in Figure 5.5.

1. The first step of this implementation process is to express the legacy reference models using XML schema in this work.

2. The GIRM definitions were coded in a mediating XML schema in order that LinkEHR editor that was used in this work could be used to perform the transformation to create archetypes.

3. A flexible definition of identity can be produced by domain experts and IT experts, who would reconstruct the GIRM by choosing the desired features of demographics information. This also allows users to create a GIRM-based schema.

4. After the archetype mappings were created, the author created a Java program to allow the transformation between ADL and XSD based on these mapping rules (Chen, X, 2015).

It is perhaps worth noting that the normalisation process requires the use of a data type schema which may or may not be specified by a standardised RM. For example as mentioned in the previous section, the author has employed the CEN data types (CEN/TS, 2004) defined in EN13606 for the ISO/TS 22220 demographics model. The final step in this process is to instantiate the XML schema with the data value so that the identities are exchanged in a XML message.

A complete example of transforming the HL7v2 PID (Person Identification) message to the GIRM-based ISO EN13606 Demographic Person SoC identification message as well as other archetypes of the surveyed demographics models covered in this chapter are

Figure. 5.5 Instance message exchanging process.

provided online at (Chen, X, 2015). In addition to exchanging data, The implementation in this work also seeks to build and populate the identity relational database based on the GIRM schema to persist the demographics data that support both EN13606 demographics model and the relevant demographics aspects of the HL7 RIM v2. The data source of this experiment is used from the relational database of an industrial partner in the EHR*land* project. Figure 5.6 visualises a sample identity object with two sets of attributes and the data value using UML object diagram. The relational database design is tested based on the patient table with 53 attributes within the legacy system.

This Java implementation is developed to demonstrate step by step how a particular identity is represented and mapped between two models as follows.

1. Represent the PID segments of an HL7v2 message in XML format for data communication.
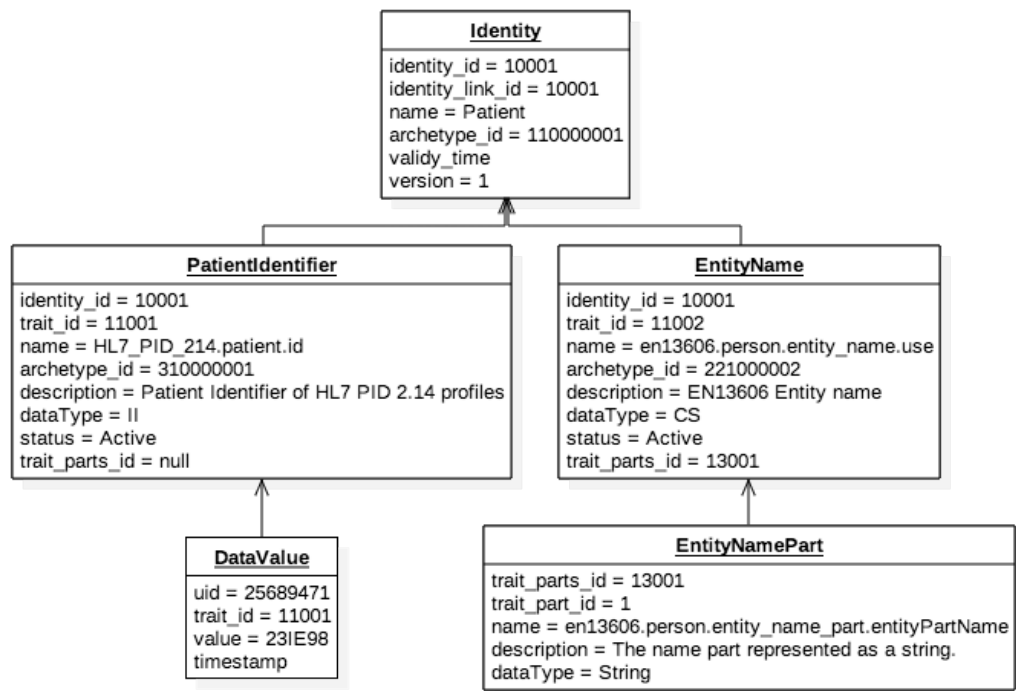
Figure. 5.6 Object diagram of sample identity data

2. Create a general GIRM demographics archetype represented in XML format for transformation between HL7v2 and GIRM.

3. Define the XQuery mapping of the HL7 PID RIM and GIRM-based EN13606 RM.

4. Generate the results message based on the GIRM-based EN13606 demographics RM by transforming HL7 message.

5.3.1   Observations

However, apart from any other considerations, the key work to enable transforming heterogeneous identity data requires the development of manual mappings before any healthcare organisation would even begin to consider employing the GIRM approach. For example, the mapping of *patient_address* between HL7 v2 RM and EN13606 demographics RM covers 5 PID.11 (Patient Identification) segments, the whole patient identity part of

the message to be added as shown in Listing 5.5. Table 5.4 shows the matching of concept. Most of the missing fields are for concepts that it is difficult to agree on at international level. The manual mappings have been tested on 21 PID segments from PID.1.1 to PID.11.5 and none of the transforming processes have caused the misinterpretation of archetypes within the test data. Results show that, regardless of data transforming implementation, the issues of semantic ambiguity, duplicated semantics between two chosen models still exist, which require the domain experts or terminologist to facilitate the terminology problems. Another possible reason for those mismatch concepts (12 out of 30) is due to different legislation. Thus, the specialised archetypes according to the different regulations (national or regional) of healthcare can be defined based on such requirements.

Listing 5.5 An example of HL7 v2 ADT message to EN13606 demographics

```
1  <patient_address>
2    <street_address>
3      {/HL7Message/PID/PID.11/PID.11.1/text()}
4    </street_address>
5    <other_designation>
6      {/HL7Message/PID/PID.11/PID.11.2/text()}
7    </other_designation>
8    <city>
9      {/HL7Message/PID/PID.11/PID.11.3/text()}
10   </city>
11   <state_or_province>
12     {/HL7Message/PID/PID.11/PID.11.4/text()}
13   </state_or_province>
14   <zip_or_postal_code>
15     {/HL7Message/PID/PID.11/PID.11.5/text()}
16   </zip_or_postal_code>
17 </patient_address>
```

The overall experience of this approach was positive as it was possible to generate EN13606-compliant archetypes using only graphical user interfaces: LinkEHR guarantees the compliance with the chosen archetypes so answering to Q4. Although the use of Graphical User Interface helps in producing the mapping XQuery, it does not change the

Table 5.4 A summary of mapping between PID segments and EN13606 demographics traits

PID	Segment Name	EN13606 Attribute
1	Set ID – Patient ID	id: SET<II>
2	Patient ID (External ID)	id: SET<II>
3	Patient ID (Internal ID)	id:SET<II>
4	Alternate Patient ID – PID	id:SET<II>
5	Patient Name	entityPartName: String
6	Mother's Maiden Name	entityPartName: String
7	Date/Time of Birth	birthTime: TS
8	Sex	administrativeGenderCode: CS
9	Patient Alias	entityPartName: String
10	Race	
11	Patient Address	addressLine: String
12	Country Code	
13	Phone Number – Home	TELECOM
14	Phone Number – Business	TELECOM
15	Primary Language	
16	Marital Status	
17	Religion	
18	Patient Account Number	id: SET<II>
19	SSN Number – Patient	id: SET<II>
20	Driver's License Number - Patient	id: SET<II>
21	Mother's Identifier	id: SET<II>
22	Ethnic Group	
23	Birth Place	
24	Multiple Birth Indicator	
25	Birth Order	birthOrderNumber: Integer
26	Citizenship	
27	Veterans Military Status	
28	Nationality	
29	Patient Death Date and Time	deceasedTime
30	Patient Death Indicator	

fact that the overall mapping process is a lengthy one. However with the right amount of training and an upgraded version of the LinkEHR tool, it is expected to dramatically speed up the process. Unfortunately it remains a manual process that cannot be completely automated. As this experiment being realised in a theoretical environment one can expect additional hurdles when using this approach in a real-life environment. The wide variety of HL7 messages could make this a daunting task. For the whole scenario to succeed, it is therefore advisable to agree on both the set of HL7 messages (e.g IHE PIX and PDQ profiles) to be produced by healthcare professionals and the archetypes to be used for producing the record extracts. Only with such an agreement can the time-consuming XQuery generation task be limited while ensuring a consistent approach to health record data transmission for a wide variety of users.

For an implementation note Trait Part in the GIRM, it was noted in Section 4.3.3 that Trait Part class was included to represent constructs such as Name Part and Address Part, i.e. data elements that are "sub-atomic" not full traits in their own right but form part of a trait. Conceptually, Trait Part has a role to play in the GIRM. However, from an implementation point of view, this role can be filled by extending the role of simple trait to include "sub-atomic" data elements as well as self contained simple traits that do not have multiple components. This leads to the UML diagram shown in Figure 5.7. There are a couple of pragmatic advantages to this approach in that the resulting model is simpler, and is more compatible with the reference model "design pattern" of recursive aggregation that is applied in openEHR and EN13606. But then, the entire GIRM could also be represented by reusing constructs that are already available within openEHR and EN13606 reference models.

However, it is not author's intention to make the work compatible with either EN13606 or openEHR, but only to show how the concept of a generalised identity model can be realised using a reference model and archetypes. In order to preserve this conceptual aspect
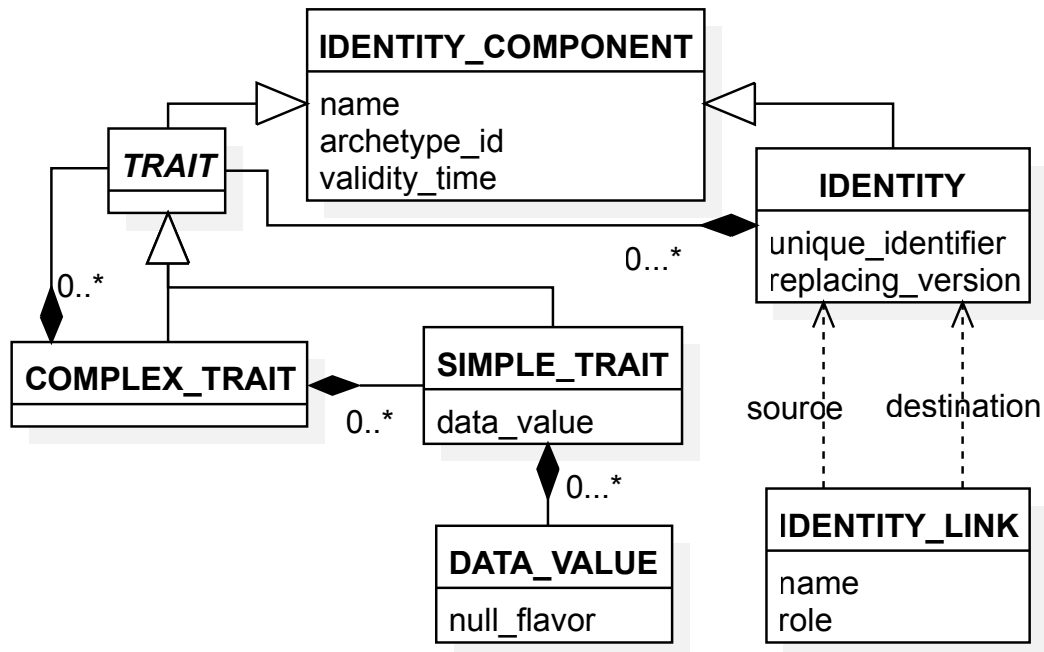
Figure. 5.7 The "simplified" class diagram of the GIRM prototype

of the GIRM that has been adopted this work, this class has not been removed within this work.

5.4   Summary

This chapter presented the evolution process of constructing the GIRM and using the GIRM to exchange identity information cross the working systems. The evaluation takes examples of both the fixed models and the two-level models approach for subjects of care which were discussed in Chapter Four. In particular, the author analysed and compared similarities and differences derived from the demographic packages of these modelling approaches. The chapter also describes the investigation and production procedures for a customised GIRM-based mediated schema. This schema was created to assist in the "baseline" task of identity exchanging in the presence of differing underlying demographic models or schemas where one schema employs two-level models and archetypes. The procedure emphasised the provenance of the chosen features utilised. The approach proposed in the

initial evaluation are offered by means of developing an extensible two-level information model for demographics and identity in healthcare. In various modern EHR systems demographics and health information are separated for security reasons.

CHAPTER SIX

APPLICATION OF THE GIRM FOR IDENTIFYING LABORA-

TORY INFORMATION

6.1 Overview

Let us return once more to the main domain example that has been used to help illustrate

the need for a GIRM – biomedical science as practiced in a contemporary acute care

setting. Biomedical scientists working in a modern clinical laboratory in a modern hospital

rarely come into direct contact with patients. Actually, from the author's experience of

working in such laboratories, it is more common for biomedical scientists to deal mostly

with specimens, orders and medical devices. In some cases, information about the patient

such as patient weight, gender, pregnancy status and age may provide context information

for the work of biomedical scientists and the array of clinical analysers that surrounds

them, but the analytical focus is primarily on the specimen as a subject of investigation,

rather than a patient as a subject of care, while the entity that is doing the analysis is often

a machine such as a clinical analyser - rather than a health care professional.

Laboratory Information Management System (LIMS) records information about spec-

imens and orders as identified entities. But when information from the laboratory is

transferred from a LIMS into a standards-based electronic health record (EHR), current

practice demands that the focus of documentation shifts so that the patient becomes the key identified entity. Meanwhile, the intended convention in ISO and openEHR EHR standards is that information is "signed in" to the EHR by the healthcare professional. And the focus for them is very strongly on the patient as subject of information. Information models for electronic health records are quite strict on this matter. Electronic health records are recorded about the health of the patient. Figure 6.1 presents a sample request form for blood transfusion of children's university hospital and it fairly obvious to find that record of information within the request form is about the blood not the patient. However, the EHR records are not intended to be recorded about (World 1) donors, specimens, orders, blood products, organs, medical devices, pharmaceutical products, or (World 2) episodes of care and orders, or indeed about (world three) medical images, sampled physiological signals and "omic" structures and many other potential identified entities that are of interest to various allied health professionals who are involved in the care of the patient.

Figure. 6.1 A sample of blood transfusion request form

All of these entities can themselves be subjects of healthcare information outside of the record. But if information relating to these entities is to be included in the record, then it must be related directly to the patient as sole subject of information. That is just for

134

care-flows that involve allied health professionals. Where the patient is participating in their own care, they may also need alternative identified entities other than themselves (limb, organ, specimen, pharmaceutical product, prescription) to focus on. The mismatch in patient oriented and this multi-subject oriented healthcare documentation needs to be explored in order to understand how these perspectives can be harmonised.

In addition, with the growing realisation of the value of secondary use of EHR information in fields of medical research, there are also problems with the traditional focus on *patient* and *health professional* as identified entities. This includes areas such as bio-banks to support cancer research, where the focus is exclusively on the sample. For the purposes of research, the patient is often secondary. The question arises, why not allow information in feeder systems that is destined for the electronic health record to be recorded about the patient, but also about other identified entities such as the specimen and the order?

6.2   Motivation

In Section 5.1.2, the idea of specimens as subjects of information was mentioned as a reason for having a generalised identity model. In the author's view, the reference models and the associated archetypes would allow specimens and medical devices to be more readily indexed and identified, thus allowing more effective linkage between LIMS and EHRs. While an entire design for a two-level LIMS is beyond the scope of this work, the author shows how a generalised identity model the GIRM that was described previously to be applied to develop a "two-level" LIMS information model that includes both specimen and medical device as archetype-able "identified entities". Within the bounds of a laboratory, it is appropriate to engage in "specimen-centred" care while always recalling that each specimen must retain a tangible tie to a patient.

If the LIMS can be developed using two level models, why not the medical devices also? The application implemented in this section, if it was used in communications with

analysers and other medical devices, devices would mean that those medical devices and EHRs would effectively "speak the same archetype language". This would lessen the need for expensive translation between formats that exists today. A relevant study of applying the archetype approach to bio-bank information system has shown that semantic conflicts between the existing archetypes and the biomedical information systems are fairly ubiquitous (Späth and Grimson, 2011). Furthermore, this study stated that the existing modelling approach for the sample entity which is the subject of sample-centric information is cumbersome in bio-bank information system. The GIRM offers is the possibility of creating a general "lightweight" schema that is nevertheless capable of comprehensively representing diverse, complex and evolving laboratory concepts.

For these reasons, the further evaluation will be based on the application of integrating the GIRM to the domain of laboratory. And it will show that the GIRM allows laboratory information that is destined for the electronic health record to be recorded about the patient, but also about the specimen and the order. If such identities are allowed to be represented even outside the record using two-level models, a general identity model such as the one presented in this thesis is required. As a general identity model within LIMS, the author will show how the GIRM could be applied to the scenario like specimen, instrument alongside patient demographics, as subjects of information. This evaluation will show how the sample parent-child hierarchy and actions and observation on the samples can be easily modelled using this approach.

6.3   Methods of applying GIRM to LIMS

A complete evaluation study of two level LIMS involving the whole reference model and archetypes approach would need to be considered in order to comprehensively investigate the viability of the idea but in keeping to the scope of this work, the author is looking only at the separation of (archetype-able) identity model and (archetyped) clinical information.

If the reader is interested in other aspects of this idea, an investigation of the clinical information modelling for a LIMS without the benefit of a generalised identity model, has been investigated by (Kopanitsa and Taranik, 2015).

In the context of designing data model for LIMS, a two-level approach may be taken by separating the clinical knowledge and the identity information. The ISO EN13606 suggested the reference models based on the patient centric subject of care. However, not only because of the privacy principle of designing the modern information system but also because of updating usage and last maintenance, the patient centric reference models are not sufficient and extendable to store and manage information based on laboratory results (Smith and Ceusters, 2006).

A generalised identity model provides the rich meanings and the functions for the identity representation in many EHR systems but only a few approaches can be taken to make heterogeneous systems simple to integrate a mutual identification process. In order to apply the two-level approach in the LIMS, a specimen centric entity can be used to replace the approach of the demographics model part of the original EHRcom reference models. However, the two-level approach cannot be applied without the support of identified specimens and probably identified orders also, as subjects of information. A solution is needed to model the identified specimen that is a primary actor rather than a patient within the LIMS based on a generalised identity model.

The first example application demonstrates how a generalised two level model that has been previously developed and validated by the author in Section 5.2, can be used to investigate the relationship between the differing perspectives of clinicians, laboratory professionals, and patients, and to provide different identified entities against which health information can be indexed and organised to suit these differing perspectives (namely specimen).

The second example application describes how a generalised identity model can bridge the gap of the identity models for EHRs that are used in healthcare, and those that are

required in research where the focus is more on the sample. The author believes that adopting this more general approach to identity will allow EHRs to be used for secondary use and for electronic records where research effort is around the specimen rather than the patient.

6.3.1 Evaluation Setting

For utilising the GIRM approach to develop the archetypes that can be used in a LIMS, the same process of constructing the GIRM by observing the main existing specifications of the analytical instruments is required. In this study, the author investigated the ISO 11073 and ASTM 1394-97 standards which has been mainly used for the development of medical devices in particular (Markey and Berry, 2010). Given that they are describing the same problem domain although with a gap of 15 years between the two sets of specifications. So it is likely to find some overlapping concepts between these two sources of metadata so that the evaluation illustrates the difference and similarity between the information models of ASTM 1394 and ISO 11073:90101 which helps to define the subject of laboratory information for the record.

Through analysis of the concepts defined in the ASTM models of laboratory information, the author developed a set of archetypes that fully cover all the fields of specimen, order/request and instrument using the GIRM within LIMS - but where traits that allow human or computer identification of the identified entities to be kept separate. Furthermore, the mapping and overlapping between the fields of specimen and instrument using the GIRM are presented to compare with the openEHR archetypes.

With the development of GIRM-based laboratory archetypes, the author surveyed three open source laboratory information systems, Bika Lab Systems (Bika, 2002), Labkey LIMS (LabKey, 2005), OpenLIMS (OpenLIMS, 2013) and a medical record system OpenMRS (OpenMRS, 2004) with the support of exchanging laboratory information as the data sets. Next, the evaluation process follows the investigation of the database structures of

these LIMS to define the GIRM-based archetypes for identifying the subject of clinical laboratory information within those models.

6.4 Results

6.4.1 *Investigation 5*: Difference and similarity between 90101 and ASTM

To demonstrate the difference and similarity between the existing information models, the author investigated both ISO 11073:90101 and ASTM 1394 to seek the candidate identified entities for modelling. Table 6.1 lists 26 concepts from ISO 11073:90101 which consists 101 attributes that were compared to the same concepts within the ASTM 1394 (a specification for clinical laboratory instruments) excluding the non-identified entities. For example, Device, Operator, Patient Service, Access point, Specimen, Order, Request and Directive are potential identified entities based on the author's experience.

By the comparison of these properties between test suites, the coverage of the attributes reaches 26% in total. This means that there are the big gaps in the test coverage of clinical laboratory information. However, the overlapping discoveries of these properties reveal the mandatory characteristics of clinical laboratory identity that are recorded in both specification (e.g., Header, Device, Access control, Operator, Patient, Service, Observation, Specimen, Order/Request, Note). Table 6.1 presents a list of the main identified entities that focus on device, specimen and order within ISO 11073 and ASTM 1394. The entities 3 to 7, 10, 22, 23 can be the part of record of the "DEVICE" which is the primary subject in the domain of laboratory. Similar to the *SPECIMEN*, the record of specimen can consist of the entities 9,11 and both the "SPECIMEN" and the "TERMINATION" could also be the part of record of the "REQUEST" and the "ORDER". In addition to the results, both ISO 11073:90101 and ASTM 1394 provide more detailed record components for the

Table 6.1 Investigated concepts of laboratory within ISO 11073:90101

	ISO 11073:90101 Property (The number of attributes)
1	HEADER (5)
2	DEVICE (10)
3	DEVICE STATUS (8)
4	DEVICE CAPABILITIES (2)
5	DEVICE STATIC CAPABILITIES (4)
6	ACCESS CONTROL (5)
7	DEVICE EVENT (3)
8	OPERATOR (2)
9	PATIENT (8)
10	SERVICE (5)
11	OBSERVATION (8)
12	ACCESS POINT (2)
13	SPECIMEN (4)
14	TERMINATION (2)
15	UPDATE ACTION (1)
16	END OF TOPIC (1)
17	ORDER (3)
18	REQUEST (1)
19	ACKNOWLEDGE (4)
20	DIRECTIVE (1)
21	ESCAPE (3)
22	REAGENT (3)
23	CONTROL CALIBRATION (5)
24	TIME (2)
25	TIMEZONE (4)
26	NOTE (1)

specific entities such as specimen, order, request which are extracted from these models as shown in Table 6.2. The empty columns in this table denote the status of not present which represents the mismatch properties in ASTM 1394. Further, the extra properties extracted from ISO 11073 and ASTM 1394 were summarised in Table 6.2 that contributed to the archetypes creation.

In this investigation, 3 new archetypes were created as shown in Table 6.4 that fully cover all the fields of specimen, order/request and instrument by support of ISO 11073:90101 and ASTM 1394 using the GIRM-based model to represent the traits of

Table 6.2 Properties of the identified entity between ISO 11073 and ASTM 1394

ISO 11073:90101 Property	ASTM 1394 Property
DEVICE	
device_id	Instrument identification
vendor_id	No ASTM equivalent
model_id	No ASTM equivalent
serial_id	No ASTM equivalent
manufacturer_name	No ASTM equivalent
hw_version	No ASTM equivalent
sw_version	No ASTM equivalent
device_name	No ASTM equivalent
vmd_name	No ASTM equivalent
vmd_id	Instrument Section ID
SPECIMEN	
specimen_dttm	Specimen Collection Date and Time
specimen_id	Specimen ID
source_cd	SpecDesc.SpecimenSource
type_cd	SpecDesc.SpecimenType
ORDER	
universal_service_id	No ASTM equivalent
ordering_provider_id	Ordering Physician
order_id	No ASTM equivalent
REQUEST	
request_cd	No ASTM equivalent

Table 6.3 A summary of the properties within ISO 11073:90101 and ASTM 1394

	EXTRA ISO 11073:90101 PROPERTIES		EXTRA ASTM 1394 PROPERTIES
1	RESULT RECORD	1	ORDER
	Record Type ID		Record Type ID
	Sequence Number		Sequence Number
	Units		Collector ID
	Nature of Abnormality Testing		Action Code
	Date of Change in Instrument Normative Values or Units		Danger Code
			Relevant Clinical Information
2	SPECIMEN RECORD		Date/Time Specimen Received
	Universal Test ID		Ordering Physician
	Priority		Physician 's Telephone Number
	Requested/Ordered Date and Time		Users Field No. 1
	Collection End Time		Users Field No. 2
	Collection Volume		Laboratory Field No. 1
	Specimen Service		Laboratory Field No. 2
	Specimen Institution		Date/Time Results Reported or Last Modified
			Instrument Charge to Computer System
3	REQUEST INFORMATION RECORD		Report Types
	Record Type ID		Reserved Field
	Sequence Number		Location or Ward of Specimen Collection
	Starting Range ID Number:		Nosocomial Injection Flag
	Ending Range ID Number		
	Universal Test lD		
	Nature of Request Time Limits		
	Beginning Request Results Date and Time		
	Ending Request Results Date and Time		
	Requesting Physician Name		
	Requesting Physician Telephone Number		
	User Field No. 1		
	User Field No. 2		
	Request Information Status Codes		

the identified entity. The results, in particular, show the effectiveness of modelling the identified entities in the sample-centric LIMS using the GIRM so answering Q5.

Table 6.4 New archetypes of Specimen, Instrument and Order created by the GIRM

Clinical laboratory Archetypes	Source
GIRM-IDENTIFIED_ENTITY.Specimen.v1.adl	ISO 11073:90101
GIRM-IDENTIFIED_ENTITY.Instrument.v1.adl	ISO 11073:90101
GIRM-IDENTIFIED_ENTITY.Order.v1.adl	ISO 11073:90101 & ASTM 1394

6.4.2 *Application 2*: Mapping/overlapping between the fields of specimen

Then some key results need to be clarified here, a specimen has identity but **a specimen can also have a "record"** just as a patient can have a record. Only some properties of a specimen can be considered as "traits" that are there for the purposes of identifying the specimen. The rest of the properties are part of the clinical information as the "record" of the specimen that tell the biomedical scientist about the *health/status* of the specimen and they are not used for identification of the specimen. The same definition for constructing the GIRM for laboratory information could be said for other identified entities such as instruments, limbs, organs, medical devices, blood products.

For example, Figure 6.2 presents the identity characteristics of the specimen from the ASTM 1394 and ISO 11073 that are archetyped using the GIRM approach. Further, the author defined which attributes in the GIRM-based laboratory information model are part of **the identity of the specimen**, and which are part of the "clinical information" of the specimen based on these principles so answering to Q7.

1. The trait set includes a unique identifier or relevant information which can be used to identify the specimen by LIMS or human such as specimen ID, collector ID, etc.

2. The trait has the attribute that can be recognised by human to identify the specimen such as time, colour, weight, volume, etc.

3. A trait has an attribute that can only be recognised by system would need to be handled with extreme caution. For example, the *Specimen Collection Time* is like the "date of birth" of a patient and it is important to help to identify the "birth" status of the specimen.

The various parts of Universal Test ID and Specimen Descriptor are identity information as well as the collection time which is like "date of birth" of specimen. The rest of the data
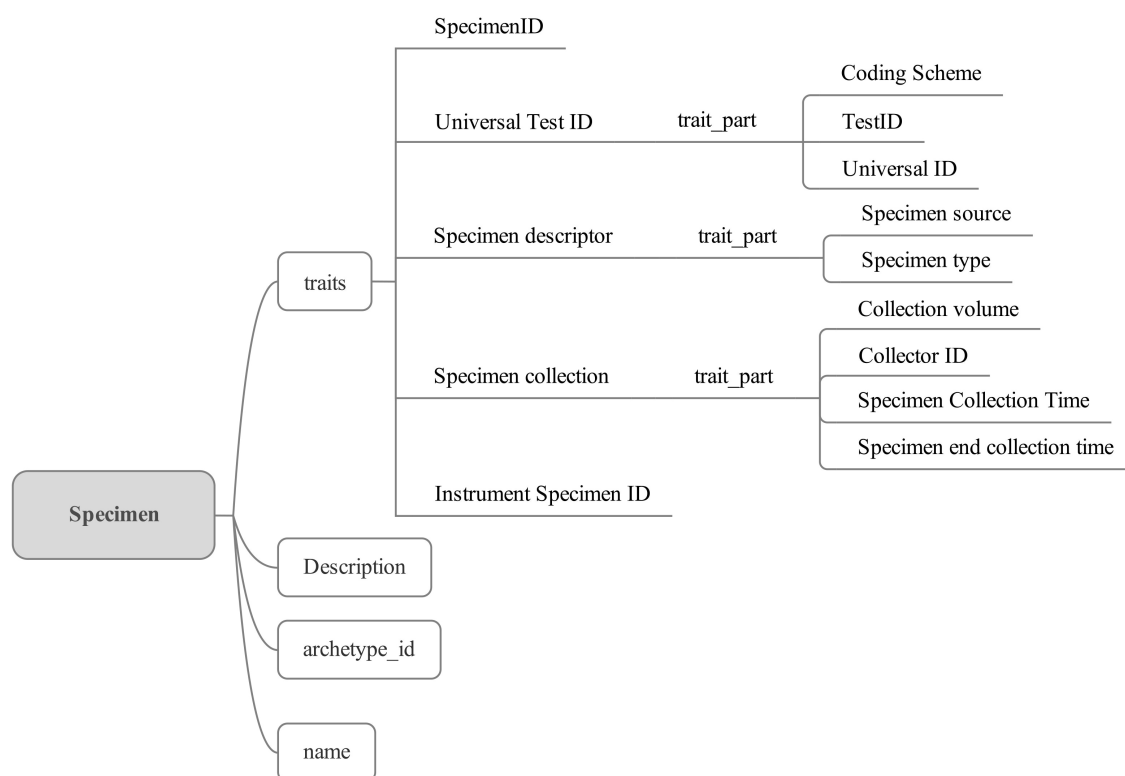
Figure. 6.2 Specimen concepts of ISO 11073 constructing by GIRM

items are part of the "documentation" about the specimen. So the identity model should only be used to represent the identity information. The rest of the information should be represented using openEHR reference model *Clusters* and *Items* based on the principles above. Also the author expects that every item in the openEHR specimen archetype as shown in Figure 6.3 should also be in the record of specimen except only a small number of the properties are identity "traits" or duplicated terminologies.

There is another consequence of using the GIRM in the laboratory domain. Using this approach is possible to create the archetypes not only for the identity and for care but also the management information that is only interested in the feeder system and its local domain. In the LIMS, this would include the information on the category of specimen, QC or medical device. Such information would be part of the EHR but it is still important in the lab domain. In this author's view, the third category of information should be recorded using openEHR archetypes and associated identity entities such medical device, specimen

and request. Therefore, this example application alongside the investigation in the previous section provides the answer to Q5 and Q6.
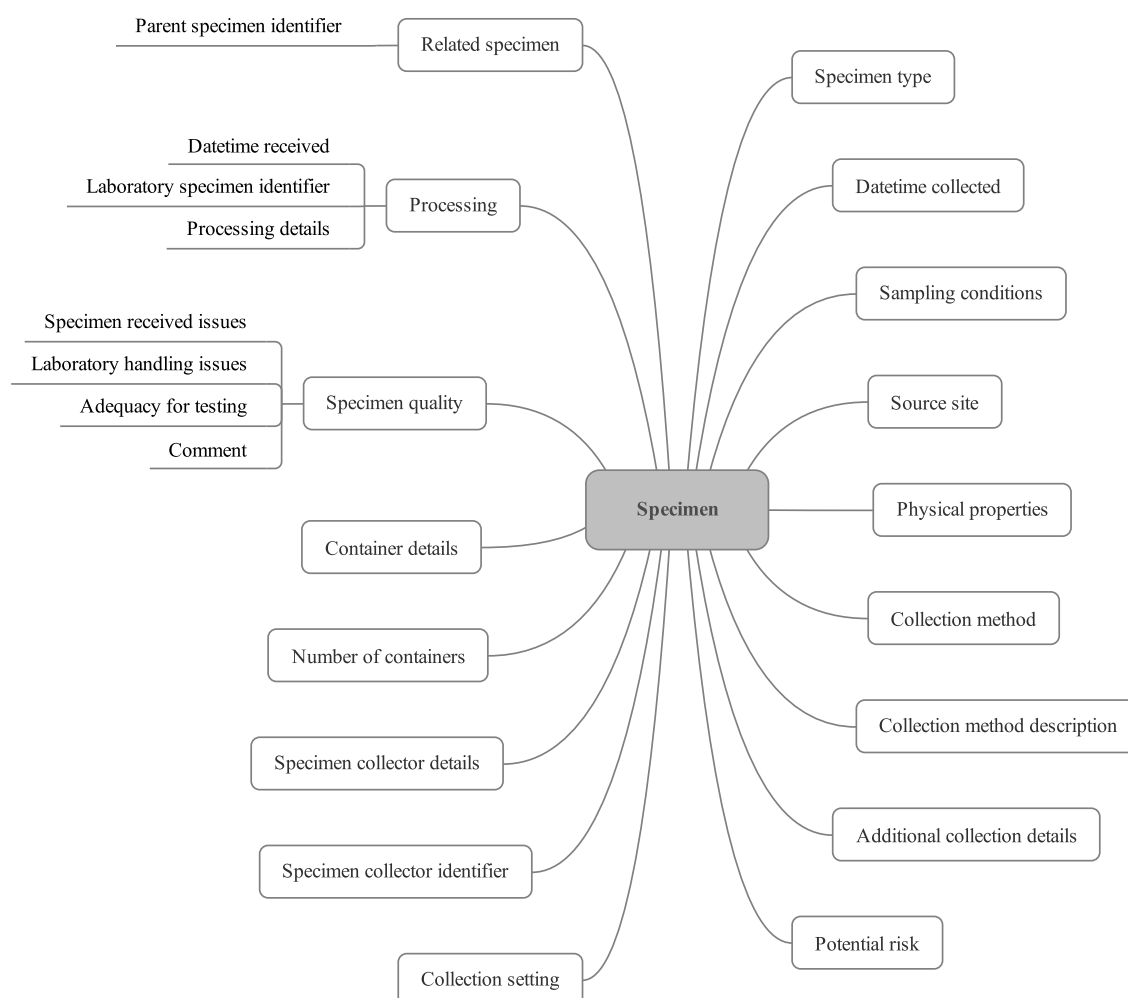
Figure. 6.3 The concepts and structure of the openEHR specimen archetype

6.4.3   *Application 3*:   Applying   the   GIRM   to   the   database   of   open source LIMS

There are 22 open source laboratory information management systems introduced on the LIMSwiki.org which is dedicated to the laboratory informatics community (LIMSwiki, 2015). Four working laboratory information systems and medical software applications

Table 6.5 A summary of four data sources used in the study

	Bika LIMS	LabKey	openLIMS	openMRS
Written in	Python	Java	PHP	Java
Database	PostgreSQL	PostgreSQL/ Microsoft SQL Server	PostgreSQL	MySQL
Type	Laboratory informatics software	Laboratory informatics software	Laboratory informatics software	Patient-based medical record software (with LIMS Module)

with laboratory components were investigated in this study due to ease of access to database as shown in Table 6.5.

- Bika LIMS by Bika Lab Systems (approximately 7 installations according to the LIMSwiki) is primarily intended for laboratory data management (LIMSwiki, 2015).

- Labkey is a popular LIMS (approximately 36 countries and 350 installations) which focuses on biomedical research data management including proteomics, flow cytometry, and observational study management (Nelson *et al.*, 2011).

- openLIMS is intended to provide a open source web-based laboratory information systems with different laboratory phases such as planning experiment, entering data and sharing data.

- openMRS (approximately 120 installations) is an open sourced healthcare documentation application which has a laboratory information system component.

As it turns out, the working database models of these information systems are complicated to examine. Therefore, the conceptual schemas of underlying relational databases that present the entity relationship were used. In relation to the discovery of laboratory information within the entire database schema, Bika, Labkey, openLIMS and openMRS have 49 classes, 66 classes, 122 classes and 101 classes respectively.

Many of the classes within these LIMS models are out of the direct scope of this work as these applications are designed to be multi-purpose and not just for frontline healthcare. Therefore they operate for other perspectives such as biobank, sequence
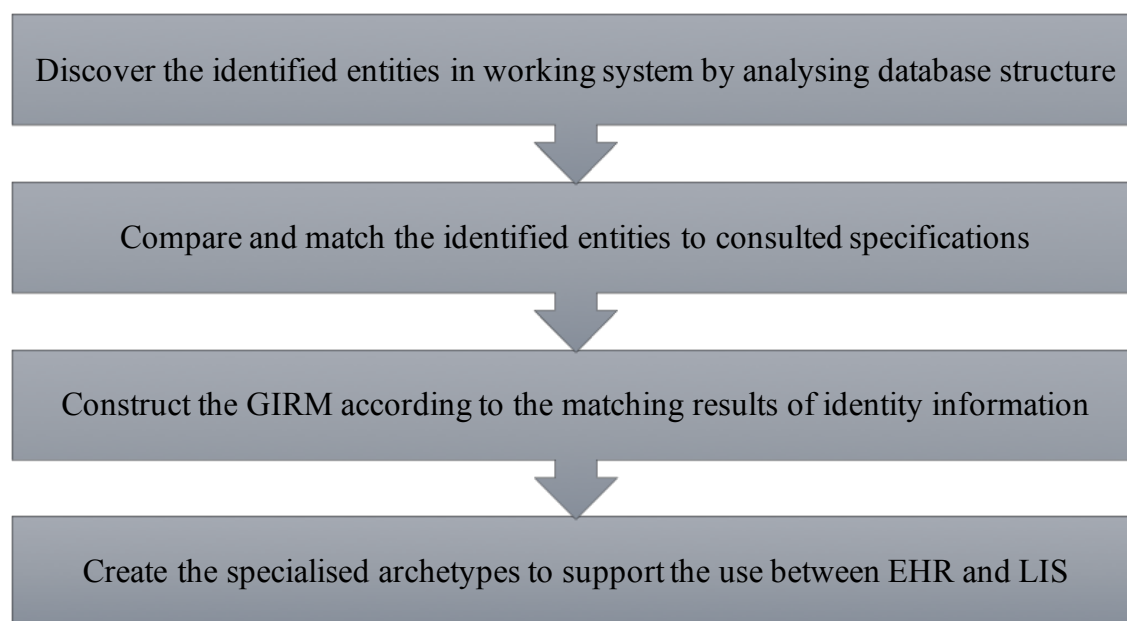
Figure. 6.4 Flowchart of summarsing the GIRM modelling approch applied on clinical laborarty information

analysis, and administration functions. However, the identification of the subject of laboratory information at the time of writing is still a time consuming process especially within these heterogeneous systems. According to the principles of identifying clinical laboratory entities that are defined in Section 6.4.2, at first, any result that appears not to be "relevant" identified entity is removed from the data sets.

Figure 6.4 visualised the process of applying the GIRM to model the identified entities that are considered as the subject of laboratory information. As a result, a prime difference between the GIRM approach and existing archetypes (openEHR) is that, the GIRM is to model the identity information for the subject of clinical laboratory but it also supports reuse of the archetypes for clinical information as the records of those subjects.

Figure 6.5 illustrates the conceptual schema of the main relationships between identified entities in the Bika database. Those entities highlighted in yellow are categorised by the author as candidate identified entities that are the commonly encountered subjects of clinical laboratory information in LIMS. For instance, *Samples*, *Analysis request*, *Reference Samples*, *Worksheet* and *Instruments* are identified according to the principles defined in
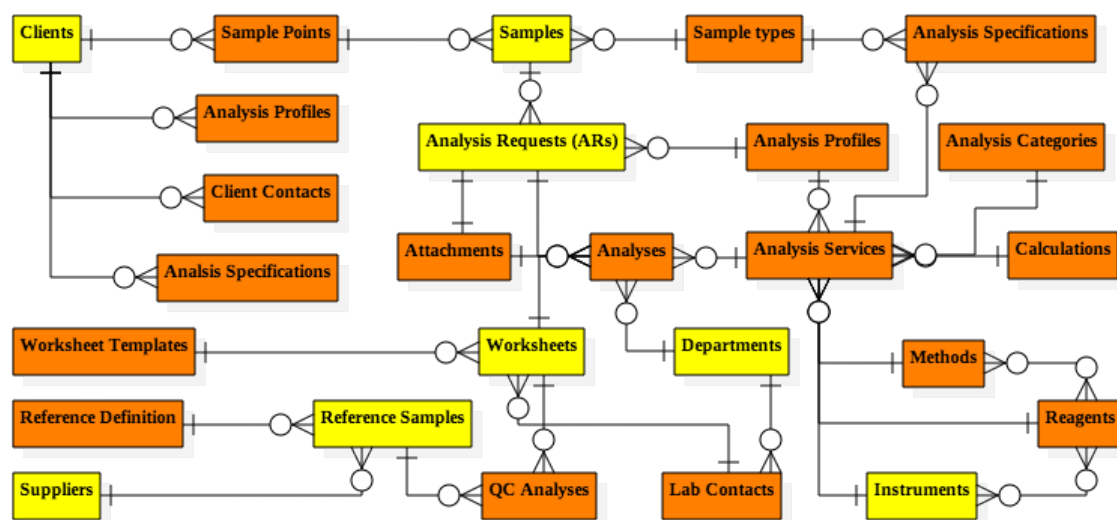
Figure. 6.5 Overview of the entity relationship schema of laboratory identities of Bika

Section 6.4.2. The rest of the entities highlighted in orange in Figure 6.5, are the associated attributes or the demographics traits such as department name, client contact details, lab contact details.

The overview conceptual schema of Labkey data entry is shown in Figure 6.6. There are 6 identified entities highlighted in yellow such as the identities *Participant* (identified by "*Participant IDs*") and *Visits* (identified by "*Visit IDs*" or "*SequenceNums*") that are identified based on the GIRM approach.

Figure 6.7 and Figure 6.8 illustrate the similar results of identifying the GIRM-based entities from those laboratory database schema. Further, the author continued to define the archetypes to cover all those identified entities in the previous evaluation. Such demographics identities (*Participant*, *Provider*, *Patient*, *Users*, *Suppliers* and *Organisation*) are defined by reusing the archetypes of EN13606 and ISO/TS 27527 (which are created by the GIRM in the initial evaluation). The main clinical identified entities (*Specimen*, *Request*, *Instruments*, *Order*) are represented by the archetypes whose traits have been extracted from ASTM and ISO 11073:90101 that are modified to add the specialised unique identifiers according to the underlying LIMS. There are 4 identified entities *Worksheet*,
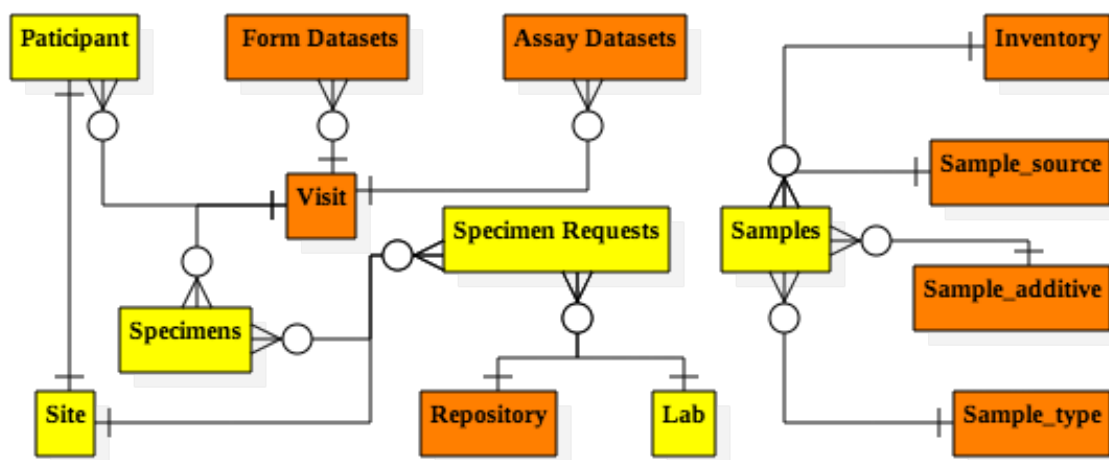
Figure. 6.6 Overview of the entity relationship schema of laboratory identities of Labkey
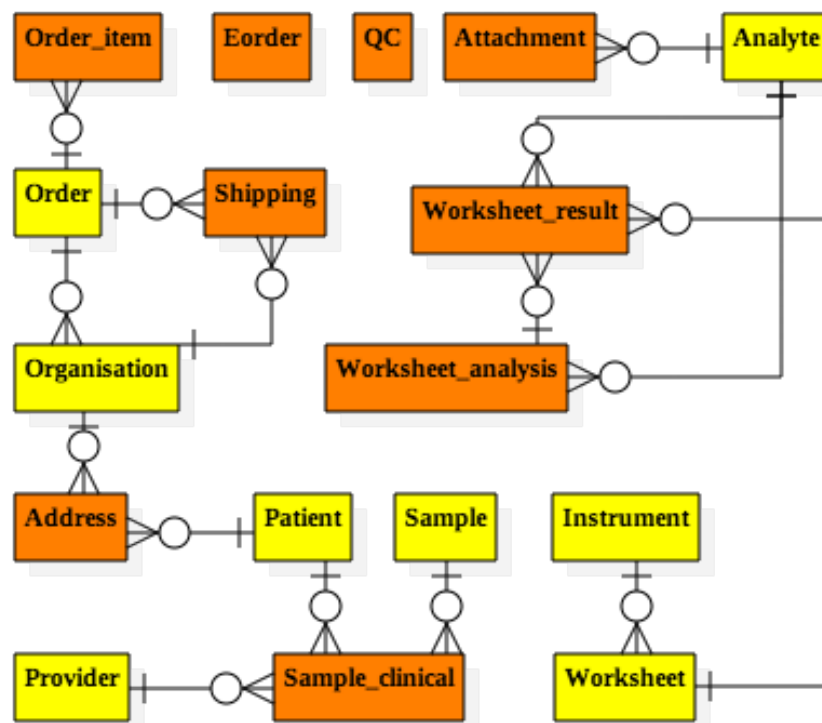
Figure. 6.7 Overview of the entity relationship schema of laboratory identities of Open-LIMS
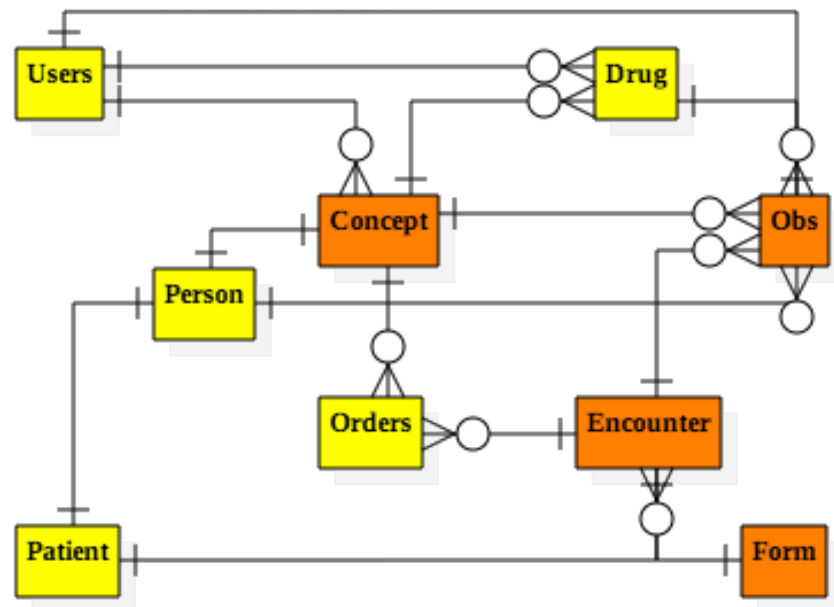
Figure. 6.8 Overview of the entity relationship schema of laboratory identities of OpenMRS

Inventory, *Analyte* and *Drug* that have no match for the surveyed specifications as each of these is defined to play a part in its own system. A possible solution for these entities is to define a new customised archetype just by adding their own identifiers for modelling by the GIRM.

For example, the following Table 6.6 can be constructed by the GIRM archetypes that covers all identified entities in the example application. There are 2 new archetypes which are added for "Site" and "Worksheets" and 6 archetypes are required to modify in order to fit the test database. 10 other demographics archetypes (has been created by the initial evaluation) can be reused in this case and this provides the answer to Q6.

6.5   Discussion

From surveyed literature about identity services in Chapter Two, such services traditionally seem to focus on identity management to support mainly access control to online resources. Of course as indicated in chapter four, in the healthcare domain, the definition of identity

Table 6.6 Overview of selected non-demographic archetypes of identified entities within the selected database schemas

Clinical laboratory Archetypes	Source
New archetypes:	
GIRM-ASTM-IDENTITY.body_site.v1.adl	ASTM 1394
GIRM-ASTM-IDENTITY.Worksheets.v1.adl	ISO 11073:90101
Modified archetypes:	
GIRM-IDENTIFIED_ENTITY.Specimen.v1.adl	ISO 11073:90101
GIRM-IDENTIFIED_ENTITY.Laboratory_request.v1.adl	ISO 11073:90101
GIRM-IDENTIFIED_ENTITY.instruments.v1.adl	ISO 11073:90101
GIRM-IDENTIFIED_ENTITY.Analyser_Information_Request.v1.adl	ASTM 1394
GIRM-IDENTIFIED_ENTITY.Analyser_subject_information.v1.adl	ASTM 1394
GIRM-IDENTIFIED_ENTITY.Order.v1.adl	ASTM 1394

service would be broader, as it encompasses at the very least, the identities of both the user (healthcare professional) and the patient (record). Traditional identity services seem to focus on identity management to support mainly access control to online resources.

In this work, the author's intended scope for identity management and identity services is broader still and it applies over the EHR ranging from the actor (such as patient, healthcare professional) to the resource that they are trying to access the record for e.g., specimen, limb, organ, medical device, health information system.

The set of archetypes that the doctor is looking at would be different but following the principle of harmonisation, the difference would be simply that a subset of the LIMS-archetype data is required in the EHR-archetype. So the integrated system does not have to perform a translation between getting the result and providing the report for signing in to the EHR.

During the course of modelling sample-centric identities, other standards and specification of specimen information and structure is required to be consulted. The adoption of standards is an essential requirement for achieving the interoperability among diverse healthcare systems. Harmonising between the different specifications would also improve and facilitate interoperability between these systems. A mapping between the experimental models was achieved in most cases but sometime a direct mapping was not possible. In

particular, for applying the GIRM approach to existing LIMS, it requires a process of reusing , extending and creating archetypes to specify the realistic situation with respect to documentation. It has been noted that the best design in practice for identifiers should not be coded in text data types such as DV_TEXT. The basic principle of identifiers should be more explicit to be unique ID with a formal mechanism for validation and identification such as OID, DV_URI and UUID. Beyond that, the two-level identity modelling also require a wider set of datatypes other than demographics datatypes and the clinical datatypes.

Experience of implementing two-level identity model helps to refine the GIRM. For example, the recursive function of the complex trait within the GIRM is added to support the nested trait and its trait part, which is derived by implementing the complex clinical archetypes. The application provides the identity modelling guidelines for those wishing to utilise the GIRM to improve the integration between the working EHR information model and demographics or generalised identities. It is anticipated that this technical assessment of implementing a two-level generalised identity model provides evidence for the impact of the GIRM on clinical archetypes, by for example allowing multiple subjects of information and also more integrated and multi-subject secondary use of the associated data.

Other experience of implementing two-level identity service model has been discussed in earlier articles of the author (Berry *et al.*, 2010a; Chen *et al.*, 2008). As a part of the EHR*land* project deliverables, a large-scale two-level identity service architecture that conforms to the EHRcom standards is reported to make recommendations for HIQA. (see Appendix B).

6.5.1 Consequence for access control of using two-level identity models

The implementation has demonstrated the operation of the two-level identity service architecture through matching of working LIMS schemas to the GIRM-style modelling approach. This approach leads to the possibility of generalised organisation-wide access control of information about a wide range of identified entities either in healthcare or indeed in other domains. Two level models are not tied to documenting information about patients. Using this more generalised identity model they could also be applied to document a "care record" for an elevators and escalators, aircraft, road vehicles, or buildings. In this approach, each user of the system is provided with a set of permissible roles which will allow them to access the record. As part of the care process, users in roles that are associated with care delivery in the current health context of the patient will be provided with access to the EHR of the subject of care.

One of the key benefits of the GIRM approach is to allow access for certain users to be limited to the subjects of information that concern them, such as specimens or XRays as subject of information instead of access to entire patient records. Also, if a set of complex archetypes can be defined for different types of identified entities such as healthcare professional, medical device, specimen, patient, then templates can be used to profile and simplify the visibility of data sets for individual locations. Sites get to have some flexibility about recording and making visible certain traits.

Role based access control is a proven technology, but the aspect that requires additional investigation is the idea of access control exceptions. Exceptions may apply if the care giver has been explicitly refused access to the record by the patient because they are related to the patient, or live nearby. This introduces the idea of excluding access on the basis of identifiable information which is likely to be feasible by the GIRM approach as well as for the specimen record.

6.6   Summary

In this chapter, the evaluation work seeks to go further with the use of the GIRM and employ two-level models to describe non-demographic entities by developing a two-level identity model for the sample-centric identities that can be used to document health related processes where the focus is not specifically on the patient.

This evaluation has demonstrated how the two-level clinical modelling approach that is used for EHRs can be adapted and reapplied to ease identity-related system integration problems in laboratory domain. It is shown that this approach provides a highly flexible and expressive means for representing subjects of information in the laboratory domain. The proposed approach facilitates cross-referencing and disambiguation of certain standards and data models to represent additional clinical identified entities such as specimen and order alongside patient demographics, as subjects of information, while still permitting reuse of relevant published standard-based archetypes.

DISCUSSION AND CONCLUSIONS

7.1 Background

In previous chapters, the author has introduced some of the challenges of identity exchange in healthcare surveyed some of the existing standard identity models in the health domain. Then the author investigated and outlined the approaches taken by some relevant healthcare related identity services. Next, a two-level identity model was described and evaluated in Chapter Four as well as the applications of utilising the generalised approach for other domains of health. This chapter summarises the main findings of the author's work grouped together with the details of the contributions.

7.2 Overview of the work

In Section 2.2.2, some of the "upper" ontologies such as BFO and Popper's three worlds were described and discussed. In particular, Popper's three worlds and the author's vision of an two level identity model are compatible. The GIRM has a similar purpose in some respects to the "ontology of credentials" from NIST. However, where ontologies model "what is possible", the GIRM is intended to support identification of entities that can be subject of documentation. As such the GIRM and associated archetypes can be used to

describe the small subject of entities in an ontology that require documentation as part of the health care process.

To achieve identity interoperability using the two-level modelling approach, identity management should enable safe and convenient identity exchange between different domains and according to different standards. This will enable EHRs to be conveniently integrated. Works by Kilic and Dogac (Bicer *et al.*, 2005; Kilic *et al.*, 2005) have demonstrated that EHR systems using two-level modelling will support the exchange of instances from the native form of different RMs based on different standards. This is equally true of systems using RMs like the GIRM, which in this case can transform identities from a source instance to a target instance. Additionally, if a system has two-level model functionality, it seems sensible to use the power of that functionality to (for instance) support higher quality demographics models.

Two level models for identity represents the direction that the author believes the two level clinical information modelling community needs to go in. In the author's opinion, legacy system integration will continue to be a painful problem for two level models until the perspectives of feeder information systems are acknowledged as the next step really beyond EHR perspectives. "No man is an island" and this is also true for EHRs. Therefore, this thesis has presented identity modelling associated with the implementation of applying the GIRM to identify the clinical laboratory information. Table 7.1 provides a summary of the evaluation and the key finding for each. The methods and experiments that have been described in this thesis were designed to provide evidence in support of dynamic two-level modelling of identity, to improve the efficiency, interoperability and expandability based on existing unstructured or single model approach in the domain of health.

A key contribution of this thesis is summarised as follows,

This work has described the development and initial evaluation of **a two-level generalised identity model and computational approach called GIRM for the management of shared identity information and knowledge in healthcare for validating and**

Table 7.1 A summary of the overall evaluation

Evaluation	Name (Page No.)	Description	Key finding
	Investigation 1 (P113)	Measuring the level of abstraction and constraint of GIRM	The GIRM has a high level of abstraction and is simple to instantiate to create GIRM-derived archetypes that are quite complex in modelling clinical information.
Part I	Investigation 2 (P116)	Expressing RM to support existing demographic models	This assesses the potential and effectiveness for extension of the experimental models using archetypes.
	Investigation 3 (P118)	Making EN13606 demographics extensible using GIRM	The recursive aggregation feature to enable more flexibility when constructing the EN13606 demographics entities.
	Investigation 4 (P120)	Making EN13606 demographics extensible using GIRM	This validates suitability of its use in modelling demographics information from other "fixed" models.
	Application 1 (P125)	Exchanging archetype-constrained identity data	The issues of semantic ambiguity, duplicated semantics between two chosen models still exist, which require the domain experts or terminologist to facilitate the terminology problems.
Part II	Investigation 5 (P137)	Difference and similarity between 90101 and ASTM	There are the big gaps in the test coverage of clinical laboratory information and this also shows the effectiveness of modelling the identified entities in the sample-centric LIMS using the GIRM.
	Application 2 (P141)	Mapping/overlapping between the fields of specimen	Principal distinctions between identity information and clinical information that is derived from the idea of the specimen as the "patient".
	Application 2 (P143)	Applying the GIRM to the database of open source LIMS	This allows multiple subjects of information and also more integrated and multi-subject secondary use of the associated data.

157

improving the identification of legacy health information in relation to archetype-based EHR systems. The basic idea behind the GIRM is same as for the two-level modelling approaches that is separation of knowledge and information: give developers just one generalised identity model on which to identify the demographics identities or other entities, and make it possible for these identities to be shared on any other clinical domains for the purposes of documentation across a community of care. This approach also supports the use of generalised identity to facilitate the secondary use of clinical data for purposes other than patient care.

7.3 Details of the contributions

7.3.1 Explore the "gap" of between EHR and identified entity

This work explores the state of the art in shared identity management architectures that provide a combination of processes and technologies to share identity information in the context of large scale national EHR systems and accompanying information systems. The author has provided a literature review of the most relevant shared identity and demographics models for healthcare, EHR standards and identity applications in the health care domain, to examine the level of interoperability, re-usability, functionality and complementarity that they provide.

A large body of literature is available in biomedical domains about identity sharing, identification, and identity matching. However, many works do not address the issues of identity sharing and identity matching across healthcare provider organisations. The identity management architecture is a key to address the problems associated with overloading of identity terminology, errors of patient identity matching, misidentification, unnecessary duplicate identities and "the lack of conceptual clarity". On the other hand, at the present, demographics and health information are quite rightly separated in many modern EHR systems. Interoperability between heterogeneous identity domains to support identity

sharing would require identity management features to be in place in order to make this a working system.

However, modern identity models are limited to the demographics information but as this work has shown, there are other important entities in the healthcare processes contributing to the EHR, such as devices, specimens that also need to be positively and securely identified. The existing demographics models such as the EN13606 demographics model and the openEHR demographics model, are very much driven by the view of clinicians who interact with the patient on the front line of care. Other allied health professionals (AHP) are not treated with the care due to an important actor and subject in the same way that their equivalents in "front-line care", healthcare professional and patient are. These concepts that exhibit varying levels of complexity and emphasise different features are very much "in the shadows" of an electronic health record. For this reason, the two-level identity modelling is considered to be useful to patients and healthcare providers, for flexible use of EHR standards, and reuse of data for two-level models, other subjects of care and subjects of information. This introduced the first contribution of this work:

Contribution 1: This work has explored the state of the art in shared identity model architectures that provide a combination of processes and technologies to share identity information in the context of large scale national EHR systems and accompanying information systems.

7.3.2   Investigation of the existing demographics information models

This work investigates a selection of existing demographics information models from health informatics standards with a view to modelling these features in a generalised and archetype-able identity mediated schema. The author compared and analysed the features of existing approaches by employing a categorisation approach to different types of identity reference models. The analytical data is chosen from selected EHR reference models that are CEN EN13606 Demographics, OpenEHR Demographics, HL7 RIM v2.14, ISO/TS

22220 and ISO/DTS 27527. The findings of this investigation step categorised a selection of prominent current identity models in increasing level of constraint.

The creation of a reference model such as the GIRM is only the beginning of a long process, as two-level models do not immediately cover the entire conceptual space of a domain. The first reference model level (in this case the GIRM) provides the main generic concepts of the domain. The second level (in this case consisting of GIRM archetypes) allows domain experts to combine and constrain generic domain concepts into more specific concepts over time, thereby allowing continuous experimentation and development of rich and evolving semantic descriptions of domain concepts. However, it must be noted that the evidence also shows sometimes this flexibility comes at a cost of additional complexity in the systems that use this dual model approach. This focus on the modelling approaches leads to the second contribution:

Contribution 2: The author has compared and analysed the features of the existing approaches to demographics and identity by employing a categorisation approach to different types of the identity models. The analysis in this work has provided evidence in support of the dynamic two-level modelling approach to improve the efficiency, interoperability and expandability based on existing unstructured or single model based approaches in e-Health.

7.3.3 Development of the GIRM approach and evaluation

This thesis has described the development of an approach to create a generalised identity reference model (GIRM) for sharing archetype-constrained identity information between diverse identities, models and services using two-level models within the health community. This work demonstrated that two level models enables two different logical EHR demographics or identity standards that have been expressed using the same RM to be mapped to each other by using a mediated schema approach, archetypes, and mapping

tools. The GIRM is constrained by an EHR developer while preserving the flexibilities of an archetype-able reference model.

This aspect of the work can be seen as an implementation instance of a domain compatible specialisation / enhancement EAV/CR representation. Like the EAV/CR representation, the GIRM leverages the concepts of Object-Oriented Programming (OOP) to facilitate the flexibility of data processing as well as moderating granularity of data. Unlike the EAV/CR however, two level models are domain specific in nature, containing domain specific features in the reference model. In principle, although the domain specificity is admittedly quite lightweight, the GIRM has both the advantages of the simple logical-information approach and the complex domain-knowledge approach. Furthermore, the GIRM is flexibly implementable at an instance level by creating a representation of the reference model using selected aspects of the demographics standards such as EN13606, ISO/TS 22220. Thus, the course of implementing the GIRM and the results based on the primarily evaluation in Chapter Five provide the contribution:

Contribution 3: This thesis has described the development of an approach to create a generalised identity reference model for sharing archetype-constrained identity information between diverse identities, models and services using two-level models within the health community.

It must be noted that some simple identified entities (with small number of traits) may end up as "fixed" model entities.

On the aspects of data integration, the mediated schema as the instance of the GIRM helps in mapping attributes and definitions back to their origins when the mediated schema of two EHR standards is derived from the same GIRM. In relative enhancement of domain knowledge; changing attributes of continually generated biomedical data allows much higher frequency reference information model updates. The results of the initial evaluation in Section 5.2.6 have shown that the use of the mediated schema for instantiating the GIRM facilitates the efficiency and reusability of the static identity reference model and semantic

interoperability even though it requires more cost of efficiency on basic database operation. Therefore, the contribution can be stated as follows:

Contribution 4a: This approach will allow EHR systems to utilise two level models for identifying the multiple demographics entities in the different specification perspectives. This should make system integration process easier and more streamlined through fewer transformations and potentially safer too.

On the aspects of exchanging identity information, existing standard demographics models lead to distributed systems which can be difficult to extend as changes to the fixed information model can lead to costly software changes. They do not tend to support consensus-based development of concepts by the domain community. For instance, they can produce multiple descriptions of the same domain-specific concept but ambiguity is still a problem for the meta data mapping according to differing standards. Consequently, Section 6.4.2 has demonstrated that the GIRM is extendable to support other non-demographics entities from health domain such as laboratory information, with examples of a set of modified archetypes that fully cover all the fields of specimen using two level models. As a result of the preliminary evaluation of the GIRM, this leads to the contribution:

Contribution 4: The author has presented an integrated approach to enhance identification of the patient-centred record and to also enable records that focus on other subjects of medical information such as samples, or individual body parts so that these secondary identities can be used for certain clinical studies.

7.3.4 Categorisation of identity information and record information

Another aspect of this approach is to preserve the unique relationship between patients and health professionals but allows other perspectives and types of identities to be represented.

In addition to restricting the usage of the EHR system by users, supporting identification systems should maximise the availability of cross-referencing of healthcare-related identities that help parts of an EHR to be conveniently integrated accessed and identified by the users.

So for instance for the purposes of clinical trials it is useful to create anonymised, synthesised or composite identities (Friedman *et al.*, 2010) and to make samples of various types as subjects of information. A number of other demographics or identity trait set model that can be used to represent other models and concepts from the specifications based on the methodology of two-level modelling and the derivations of the main existing identity models. Another consequence of this work could benefit biomedical research in both anonymous and identifiable settings. Clinical trials, consented patients, sensitive genomic datasets all require careful management of identities in both clinics and research labs. The author demonstrates the applicability of the GIRM approach, by providing examples of integrated applications of two-level identity modelling in the domain of the clinical laboratory. The application presented in Section 6.4.2 has been demonstrated to extract the identity information from the patient-centric EHR (reusing archetypes) to enable identification for the subject of sample-centric activity within the other health domains. These two applications provide the evidence of the contribution:

Contribution 5: The work has supported ease of evolution for laboratory information systems as well as EHR systems. Secondary use of more detailed feeder system data (i.e. LIS data) could benefit from the data quality enhancement that comes with a harmonised but also multi-view approach. Secondary use could also benefit from this seamless link between record and specimen or order views.

7.4 Implications for the identity information modelling

In the author's opinion based on experience in this work, the following advantages can be associated with applying a two level identity model:

1. It makes the reference model generic-simple-small and the domain model specific-complex-large. This also helps to make the demographics / identity schema of EHR feeder systems and EHR system more changeable and sharable.

2. From the point of view of the reference model, it is lightweight applying the two-level models that allow the combined identity model to be extendable; meanwhile, it is compatible with both a mediated identity schema and original schema to facilitate identity federation.

3. Existing identity schemas (e.g., CEN EN13606, ISO/TS 22220, ISO/TS 27527) do not allow a high level of flexibility when it comes to entering data about demographic entities or other entities or resources. The mediated schema that is employed in this thesis is designed on two levels. This allows more flexibility when using identity archetypes to support additional identity traits, or to add new types of identity without a requirement for development of additional code for the schema.

4. It facilitates more natural documentation in two level modes for allied health professions and for associated information systems that act as EHR feeder systems.

Arising from the point 4, the following consequences could also contribute the domain of biomedical laboratory and possible other AHP domains based on the author's experience.

Using the GIRM, AHPs will get access to richer context-linked information that is harmonised with the information rich resources in the EPR. For example, if a clinician wishes to do research, they are more likely through this approach to be able to determine which assay was used to generate each report. Archetype provides this richness of the new

level details to help clinician to make better use of the big data set. In another instance, in modern predictive QC practices when a QC incident occurs, it is necessary to trace specimens / samples that went through particular devices back to the point when the QC problem developed and then the information is traced back to the record in order to flag and validate the related lab results. QC is now in the middleware but not in the LIS.

Furthermore, evolution and richness of archetypes is a good match with lab medicine, which is also continuously evolving while some lab systems are designed (hospital culture choice) around the patient, they also have a view for indexing, searching, managing specimens and orders. For example, medical devices directive trends not to change information in any way as each change in format could leads to medical device directive headaches.

7.5   Future work

Admittedly, while the mechanism of building a RM is simple, it is not so simple a matter to build the right RM. The GIRM is a simple RM that is intended to demonstrate the idea of generalised identity with high level of abstraction to be instantiated to create the GIRM-derived archetypes that decompose the demographics information into *Complex_Trait* and *Trait_Part*. The recursive feature of the traits allow the certain flexibility to create or reuse the user defined archetypes but it should be noted that this flexibility comes at a cost of complexity. For this reason, despite the care that the author has taken with the development of the model, this GIRM can only be presented as a single step in a journey towards a robust and extendable identity model. SDOs can reuse existing model classes to reproduce this functionality.

Further work will be based on the evaluation that orients the identity model to make it fully compatible with modern ontological approaches to EHR creation. It is clear that special emphasis must be placed in EHR architectures on patients and healthcare

providers, for flexible use of EHR standards, and reuse of data for two-level models, other subjects of care and subjects of information should be considered. In addition, further work will be intended to develop an independent two-level identity service that provides the functions of identity management to allow certain types of users to retrieve identities and exchange identities for multiple types of identified entities within the health domain. Each domain has its own correlated identities and the identities need to be exchanged in both local domain and cross-referenced environment. Another aspect that was not covered in this work but which would be an important consideration is the ethical, legal, security implications of allowing non-demographic identified entities to be treated in a similar way to demographic entities.

7.6 Conclusions

In the author's view, for reasons of convenience and efficiency, it can make sense to have a simple single level demographics model particularly where the associated information systems do not need to interoperate. However, the flexibility of two-level models facilitates the interoperability of identity management between heterogeneous mutually self-governed EHR systems. Archetypes can be used to specialise traits, trait parts or identified entities to allow representation of other demographics models using a single general reference model.

The addition of this approach would also allow demographics models such as the EN13606 model to be extended using archetypes to encompass ISO/TS 22220 and ISO/DTS 25757 specifications and to facilitate other technical perspectives on the information contained in a patient-centred EHR. This flexible modelling approach does not need to be restricted to demographics as it exists in local domains. It could also be used to extend the application of this model beyond demographics, into other types of (human) identifiable entity, including samples, episodes, orders, organs, bones and wounds. Even more generally, the introduction of a generalised identity model, frees two level models

to be used in radically different environments (construction, aviation, transport and many others) to help to help to support the identification of a wide variety of valuable resources and to document their care.

As a first step towards fully validating the GIRM, this work has described the procedure of investigation and production of the mediated schema of a GIRM. The procedure emphasised the provenance of the chosen features of the model. It also described an initial proof-of-concept that evaluates the viability of a GIRM.

The overall results can also be considered as a means of addressing the problem of low data quality in demographics services and patient master indexes (PMIs), by:

- allowing standardised but adaptable identity models to be developed as the author presents in this work,

- allowing archetypes constraints to be made tighter to produce the data quality filters at the point of data entry (Berry, 2011). This in turn would encourage developers to support higher data quality entry practice.

At the present, for security reasons, demographics and health information are separated in many modern EHR systems. However, interoperability between heterogeneous identity domains to support matching of identity would require access control and other key security features to be in place in order to make this a working system but that is not the focus of this thesis. As noted earlier, this work has presented what could be considered as a identity model for standardisation. In order for a model such as the GIRM to be used generally by the health informatics community, it would need to be further discussed and more rigorously evaluated in a similar fashion to EN13606 and openEHR reference models by the community of domain experts.

Finally, the GIRM is not a final model but the thesis is just intended to show two-level identity modelling approach as a technique. It is hoped to suggest this two-level approach to identity modelling as a possible work item for CEN TC251 or another standards body. In

the author's opinion, this type of trial implementation or evaluation-before-standardisation, is an approach that should be incorporated more into standards development work.

Bibliography

Abbott, R. (1997). Information transfer and cognitive mismatch: a Popperian model for studies of public understanding. *Journal of Information Science*, 23(2):129–137.

Abiteboul, S. (1997). Querying semi-structured data. In Afrati, F. N. and Kolaitis, P., editors, *Database Theory—ICDT'97*, volume 1186 of *Lecture Notes in Computer Science 1186*, pages 1–18. Springer.

Agassi, J. (1974). Objective knowledge: An evolutionary approach. *Philosophia*, 4(1):163–201.

AMA (1970). Current procedural terminology (CPT). *JAMA : the journal of the American Medical Association*, 212:873–874.

Anderson, A. and Lockhart, H. (2004). SAML 2.0 profile of XACML. *OASIS, September*, (February):1–22.

Araujo, T. V., Pires, S. R., and Paiva, P. B. (2013). Development of archetypes of radiology for electronic health record.

Arellano, M. G. and Weber, G. I. (1998). Issues in identification and linkage of patient records across an integrated delivery system. *Journal of healthcare information management : JHIM*, 12(3):43–52.

B-RES (2011). Minas Gerais region EHR project. http://sres.saude.mg.gov.br/documentacao/listar. Online; Accessed: 2015-09-30.

Bachman, C. W. (1980). The Role Data Model Approach to Data Structures. *ICOD*, page 18.

Bachman, C. W. and Daya, M. (1977). The role concept in data models. In *Proceedings of the third international conference on Very large data bases*, pages 464–476.

Banton, C. L. and Filer, D. L. (2014). The impact of multiple master patient index records on the business performance of health care organizations. *Journal of Technology Research*, 5:1–10.

Bawden, D. (2002). The three worlds of health information. *Journal of Information Science*, 28(1):51–62.

Beale, T. and Heard, S. (2007). openEHR - Architecture Overview. *The OpenEHR Foundation*, pages 1–79.

Beale, T., Heard, S., Kalra, D., Lloyd, D., T Beale, S. H., and D Kalra, D. L. (2006). The openEHR Information Model (Revision 5.0). *Review Literature And Arts Of The Americas*, pages 1–25.

Becker, M. Y. (2007). Information governance in NHS's NPfIT: A case for policy specification. *International Journal of Medical Informatics*, 76(5-6):432–437.

Beeler, G. W. (1998a). HL7 Version 3 - An object-oriented methodology for collaborative standards development. *International Journal of Medical Informatics*, 48(1-3):151–161.

Beeler, G. W. (1998b). HL7 Version 3—An object-oriented methodology for collaborative standards development. *International Journal of Medical Informatics*, 48(1):151–161.

Bellamy-McIntyre, J., Luterroth, C., and Weber, G. (2011). OpenID and the Enterprise: A Model-Based Analysis of Single Sign-On Authentication. In *2011 IEEE 15th International Enterprise Distributed Object Computing Conference*, pages 129–138. IEEE.

Bermejo, J. (2007). A simplified guide to create an ontology. *Madrid University*, DRAFT:1–12.

Berry, D. (2011). *Towards the use of Archetypes to Ensure the Quality of Data in Electronic Health Records*. PhD thesis.

Berry, D., Camous, F., Chen, X., Sheng, Y., Henry, P., Pardon, S., O'Neill, A., McGrory, J., Clarke, F., Chrzanowski, K., and Garus, G. (2010a). EHRland Final Report 1.0.

Berry, D., Ven, J., Freriks, G., and Moner, D. (2010b). Use of OIDS and IIs in EN13606. Technical report, Dublin Institute of Technology.

Bicer, V., Kilic, O., Dogac, A., and Laleci, G. (2005). Archetype-based semantic interoperability of web service messages in the health care domain. *International Journal on Semantic Web and Information Systems IJSWIS*, 1(4):1–23.

Bika (2002). Open Source laboratory information management systems. http://www. bikalabs.com/. Online; Accessed: 2015-06-30.

BioPortal (2015). BioPortal - repository of biomedical ontologies. http://ncicb.nci.nih.gov/ xml/owl/EVS/Thesaurus.owl#C68806. Online; Accessed: 2015-06-30.

Bird, L., Goodchild, a., and Tun, Z. (2003). Experiences with a Two-Level Modelling Approach to Electronic Health Records. *Journal of Research and Practice in Information Technology*, 35(2):121–138.

Blobel, B. (2004a). Authorisation and access control for electronic health record systems. *International journal of medical informatics*, 73(3):251–7.

Blobel, B. (2004b). Authorisation and access control for electronic health record systems.

Blobel, B., Goossen, W., and Brochhausen, M. (2014). Clinical modeling-A critical analysis. *International journal of medical informatics*, 83(1):57–69.

Blobel, B., Nordberg, R., Davis, J. M., and Pharow, P. (2006). Modelling privilege management and access control. *International Journal of Medical Informatics*, 75:597–623.

Boone, K. (2013). CDA is Dead. http://motorcycleguy.blogspot.co.uk/2013/09/cda-is-dead. html. Online; Accessed: 2015-09-30.

Botsis, T., Hartvigsen, G., Chen, F., and Weng, C. (2010). Secondary Use of EHR: Data Quality Issues and Informatics Opportunities. *AMIA Summits on Translational Science proceedings AMIA Summit on Translational Science*, 2010:1–5.

Brickley, D. and Guha, R. (2004). RDF Vocabulary Description Language 1.0: RDF Schema.

Brookes, B. C. (1975). The fundamental problem of information science. *Informatics*, 2:42–49.

Brookes, B. C. (1980). The foundations of information science: Part I. Philosophical aspects.

Camp, L. (2004). Digital identity. *IEEE Technology and Society Magazine*, 23(3):34–41.

CDISC (2000). Clinical Data Interchange Standards Consortium. www.cdisc.org. Online; Accessed: 2015-09-30.

CDISC (2015). Biomedical Research Integrated Domain Group (BRIDG) Model. www.cdisc.org/bridg. Online; Accessed: 2015-09-30.

CEN/TC 251 (2012). ICEN/TC 251 Health informatics. http://standards.cen.eu/dyn/www/f?p=CENWEB:6:::NO:::. Online; Accessed: 2015-09-30.

CEN/TS (2004). CEN/TS 14796:2004 Health Informatics - Data Types.

Ceusters, W. and Smith, B. (2005). Tracking referents in electronic health records. *Studies in health technology and informatics*, 116:71–6.

Ceusters, W. and Smith, B. (2006). Strategies for referent tracking in electronic health records. *Journal of Biomedical Informatics*, 39(3):362–378.

CHA (2006). Continua Health Alliance. www.continuaalliance.org/. Online; Accessed: 2015-09-30.

Chamberlin, D. (2002). XQuery: An XML query language. *IBM Systems Journal*, 41(4):597–615.

Chen, X., Berry, D., and Grimson, W. (2008). Identity Management to Support Access Control in E-Health Systems. In *IFMBE Proceedings*, volume 22 of *IFMBE Proceedings*, pages 880–886. Springer Berlin Heidelberg.

Chen, X (2015). A source code repository of the author's work. https://goo.gl/P1lh9h. Online; Accessed: 2015-06-30.

Christensen, C. and Albert, R. (2006). Using graph concepts to understand the organization of complex systems. *Int. J. Bifurcation and Chaos*, 17:38.

CIMI (2011). Clinical Information Modeling Initiative. www.opencimi.org. Online; Accessed: 2015-09-30.

CIMI (2014). CIMI History and Overview. pages 1–22.

CIMI (2014). CIMI Repository. https://github.com/opencimi. Online; Accessed: 2015-09-30.

Cornet, R. and de Keizer, N. (2008). Forty years of SNOMED: a literature review. *BMC medical informatics and decision making*, 8 Suppl 1:S2.

Crompton, P. (2007). The National Programme for Information Technology–an overview. *Journal of visual communication in medicine*, 30(2):72–7.

Demner-Fushman, D., Chapman, W. W., and McDonald, C. J. (2009). What can natural language processing do for clinical decision support? *Journal of Biomedical Informatics*, 42(5):760–772.

Dolin, R. H., Alschuler, L., Boyer, S., Beebe, C., Behlen, F. M., Biron, P. V., and Shabo (Shvo), A. (2006). HL7 Clinical Document Architecture, Release 2. *Journal of the American Medical Informatics Association*, 13(1):30–39.

Eastern Research Group (2006). ERG Final Report - UNIQUE IDENTIFICATION FOR MEDICAL DEVICES. Technical Report 223.

Eichelberg, M., Aden, T., Riesmeier, J., Dogac, A., and Laleci, G. B. (2005). A survey and analysis of Electronic Healthcare Record standards. *ACM Computing Surveys*, 37(4):277–315.

Elmasri, R., Weeldreyer, J., and Hevner, a. (1985). The category concept: An extension to the entity-relationship model. *Data & Knowledge Engineering*, 1(1):75–116.

Farfán Sedano, F. J., Terrón Cuadrado, M., Castellanos Clemente, Y., Serrano Balazote, P., Moner Cano, D., and Robles Viejo, M. (2011). Patient Summary and medicines reconciliation: application of the ISO/CEN EN 13606 standard in clinical practice. *Studies in health technology and informatics*, 166:189–196.

FHIR (2014). Fast Health Interoperable Resources. http://hl7.org/implement/standards/fhir/index.html. Online; Accessed: 2015-09-30.

Florescu, D. (1999). A performance evaluation of alternative mapping schemes for storing XML data in a relational database. *Language*, pages 1–24.

Forslund, D. W., Smith, R. K., and Culpepper, T. C. (2000). Federation of the Person Identification Service between enterprises. *Proceedings / AMIA ... Annual Symposium. AMIA Symposium*, pages 240–4.

Frade, S., Freire, S. M., Sundvall, E., Patriarca-Almeida, J. H., and Cruz-Correia, R. (2013). Survey of openEHR storage implementations. *Proceedings of CBMS 2013 - 26th IEEE International Symposium on Computer-Based Medical Systems*, pages 303–307.

Friedman, C. (2009). Using Natural Language Processing and Mining of the Electronic Health Record. *Prevention*, 5651:1–5.

Friedman, L. M., Furberg, C. D., and Demets, D. L. (2010). *Fundamentals of clinical trials*.

Gao, S., Sperberg-McQueen, C. M., Thompson, H. S., Mendelsohn, N., Beech, D., and Maloney, M. (2009). W3C XML Schema Definition Language (XSD) 1.1 Part 1: Structures. *World Wide Web Consortium*, 3.

Garde, S., Knaup, P., Hovenga, E., and Herd, S. (2007a). Towards Semantic Interoperability for Electronic Health Records. *Methods of Information in Medicine*, 46(3):332–43.

Garde, S., Knaup, P., Hovenga, E. J. S., and Heard, S. (2007b). Towards semantic interoperability for electronic health records: Domain knowledge governance for openEHR archetypes. *Methods of Information in Medicine*, 46(3):332–343.

Goossen, W., Goossen-Baremans, A., and van der Zel, M. (2010). Detailed Clinical Models: A Review. *Healthcare Informatics Research*, 16(4):201.

Gregor, W. M. and Dutcher, W. (2006). An Ontology of Identity Credentials, Part 1: Background and Formulation. *Laboratory, National Institute of Standards and*, page 70.

Grenon, P. and Smith, B. (2004). SNAP and SPAN: Towards dynamic spatial ontology. *Spatial cognition and computation*, 1(March):69–103.

Grieve, G. (2013). On the future of CDA. http://www.healthintersections.com.au/?p=1699. Online; Accessed: 2015-09-30.

Grimson, W., Berry, D., Grimson, J., Stephens, G., Felton, E., Given, P., and O'Moore, R. (1998). Federated healthcare record server - the Synapses paradigm. *International Journal of Medical Informatics*, 52(1-3):3–27.

Guarino, N. (1992). Concepts, attributes and arbitrary relations. *Data & Knowledge Engineering*, 8:249–261.

Halpin, T. (1998). Object role modeling: An overview. *white paper, (online at www.orm.net)*.

Halpin, T. (2010). Object-Role Modeling. *International Journal of Information System Modeling and Design*, 1(1):33–57.

Hammond, W. (2003). Making the boundaries clearer: revisiting information systems with fading boundaries. *International Journal of Medical Informatics*, 69(2-3):99–104.

Hannan, T. (1991). Medical informatics-an Australian perspective. *Australian and New Zealand Journal of Medicine*, 21(3):363–378.

Hardt, D. (2012). The OAuth 2.0 Authorization Framework. Technical report.

Hillestad, R., Bigelow, J., Bower, A., Girosi, F., Meili, R., Scoville, R., and Taylor, R. (2005). Can electronic medical record systems transform health care? Potential health benefits, savings, and costs. *Health affairs (Project Hope)*, 24(5):1103–17.

Hillestad, R., Bigelow, J., and Chaudhry, B. (2008). Identity Crisis: an Examination of the Costs and Benefits of a Unique Patient Identifier for the U.S. Health Care System. page 95.

HIMMS (2014). Strategic Interoperability in Germany Spain and the UK.

HIQA (2009). Recommendations for a Unique Health Identifier for Individuals in Ireland. *Health Technology Assessment*, (March).

HIQA (2014). Health Identifiers Act.

HIQA (2015). Health Information and Quality Authority - health information. www.hiqa. ie/healthcare/health-information. Online; Accessed: 2015-09-30.

HL7 (2014). HL7 Version 3 Standard : Identification Service (IS), Release 1 Service.

HL7 (2015). Health Level Seven International. www.hl7.org. Online; Accessed: 2015-09-30.

HSSP (2005). Healthcare Services Specification Program. https://hssp.wikispaces.com. Online; Accessed: 2015-09-30.

HSSP (2012). Healthcare Services Specification Project. *Framework*, (2012-02-23).

Hu, J. and Peyton, L. (2009). Integrating Identity Management With Federated Healthcare Data Models. In *Lecture Notes in Business Information Processing*, volume 26 LNBIP, pages 100–112.

IBM (2009). Integrating Patient Medical records in pursuit of the EMR.

IHE (1998). Integrating the Healthcare Enterprise. www.ihe.net. Online; Accessed: 2015-09-30.

IHE (2008). IHE IT Infrastructure (ITI) Technical Framework Volume 1 Integration Profiles. *International Journal of Healthcare Technology and Management*, 1:1–177.

Inmon, W. H. (1996). *Building the data warehouse*.

ISO/IEEE (2008). ISO/IEEE 11073 Personal Health Data (PHD) standards. http://www.iso.org/iso/home/search.htm?qt=11073&published=on. Online; Accessed: 2015-09-30.

ISO/JTC (2011). Information technology - Security techniques – A framework for identity management - Part 1: Terminology and concepts.

ISO/TC 215 (1998). ISO/TC 215 Health informatics. www.iso.org/iso/iso_technical_committee?commid=54960. Online; Accessed: 2015-09-30.

ISO/TC 215 (2005). ISO/TR 20514:2005 - Health informatics - Electronic health record - Definition, scope and context.

ISO/TC 215 (2006). Health informatics - Electronic health record communication - Part 1 : Reference model.

ISO/TC 215 (2008). Health informatics - Electronic health record communication - Part 2: Archetype interchange specification.

ISO/TC 215 (2009a). Health informatics - Electronic health record communication - Part 3 : Reference archetypes and term lists.

ISO/TC 215 (2009b). Health informatics - Electronic health record communication - Part 4: Security.

ISO/TC 215 (2010). ISO/TS 27527:2010 Health informatics - Provider identification. page 76.

ISO/TC 215 (2011a). ISO 21090:2011 Health informatics – Harmonized data types for information interchange.

ISO/TC 215 (2011b). ISO/TS 22220:2011 Health informatics – Identification of subjects of health care. (88).

ISO/TS (2002). Health Informatics - Public key infrastructure.

Kay, M. (2008). Ten Reasons Why Saxon XQuery is Fast. *IEEE Data Engineering Bulletin*, 31(4):65–74.

Keller, W. (1997). Mapping Objects to Tables - A Pattern Language. *Proc Of European Conference on Pattern Languages of Programming Conference EuroPLOP 97*.

Kilic, O., Bicer, V., and Dogac, A. (2005). Mapping Archetypes to OWL. Technical report.

Klein, G., Klein, G., Ohyama, N., and Ohyama, N. (2003). Health informatics - Privilege management and access control - Part 3: Access control management.

Klein, G. O. (2002). Standardization of health informatics–results and challenges. *Methods Inf Med*, 41(4):261–270.

Klyne, G. and Carroll, J. J. (2004). Resource Description Framework (RDF): Concepts and Abstract Syntax. *W3C Recommendation*, 10:1—-20.

Kopanitsa, G. and Taranik, M. (2015). Application of ISO 13606 Archetypes for an Integration of Hospital and Laboratory Information Systems. volume 538 of *Communications in Computer and Information Science*, pages 29–36. Springer International Publishing, Cham.

Kwak, Y. S. (2005). International standards for building Electronic Health Record (EHR). In *Proceedings of the 7th International Workshop on Enterprise Networking and Computing in Healthcare Industry, HEALTHCOM 2005*, pages 18–23. IEEE.

LabKey (2005). LabKey Server. https://www.labkey.com/. Online; Accessed: 2015-06-30.

Laires, M. F., Ladeira, M. J., and Christensen, J. P. (1995). *Health in the New Communications Age: Health Care Telematics for the 21st Century*. IOS Press.

Lenson, C. M. (1998). Building a successful enterprise master patient index: a case study. *Topics in health information management*, 19(1):66–71.

Levy, A. Y., Rajaraman, A., and Ordille, J. J. (1996). Querying Heterogeneous Information Sources Using Source Descriptions. *Proceedings of 22th International Conference on Very Large Data Bases*, 1:1–26.

LIMSwiki (2015). LIMSwiki.org, the laboratory, health, and science informatics encyclopedia. http://www.limswiki.org/index.php/Main_Page. Online; Accessed: 2015-06-30.

Lincoln, A. M. (2010). EHR Implementation. In *15th Annual Conference and Scientific Symposium*.

Lippi, G., Blanckaert, N., Bonini, P., Green, S., Kitchen, S., Palicka, V., Vassault, A. J., Mattiuzzi, C., and Plebani, M. (2009). Causes, consequences, detection, and prevention of identification errors in laboratory diagnostics. *Clinical chemistry and laboratory medicine : CCLM / FESCC*, 47(2):143–53.

LiU (2006). LiU Archetype Editor for openEHR Archetypes. www.imt.liu.se/mi/ehr/tools. Online; Accessed: 2015-09-30.

Lloyd, D. and Kalra, D. (2003). EHR requirements. *Studies in health technology and informatics*, 96:231–7.

Lovis, C., Spahni, S., Cassoni, N., and Geissbuhler, A. (2007). Comprehensive management of the access to the electronic patient record: towards trans-institutional networks. *International journal of medical informatics*, 76(5-6):466–70.

Lundell, K. (2010). The Swedish ehealth approach - Symposium on Swedeish ehealth strategy. Technical report, Center for eHealth in Sweden and the National Board of Health and Welfare.

Magnussen, G. and Stavik, S. (2006). *Access Control in Heterogenous Health Care Systems: A comparison of Role Based Access Control Versus Decision Based Access Control.* PhD thesis.

Maldonado, J. a., Moner, D., Bosca, D., Fernandez, J. T., Angulo, C., and Robles, M. (2008). Semantic upgrade and normalization of existing EHR extracts. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 2008:1466–1469.

Maldonado, J. a., Moner, D., Boscá, D., Fernández-Breis, J. T., Angulo, C., and Robles, M. (2009). LinkEHR-Ed: A multi-reference model archetype editor based on formal semantics. *International Journal of Medical Informatics*, 78(8):559–570.

Mannos, D. (2003). NCPS Patient Misidentification Study : A Summary of Root Cause Analyses. *TIPS National Center for Patient Safety*, 3(1):1–4.

Manzorr, S., Ceusters, W., and Rudnicki, R. (2007). Implementation of a Referent Tracking System. *International Journal of Healthcare Information Systems and Informatics*, 2(4):41–58.

Markey, B. and Berry, D. (2010). A quiet success story in the laboratory: survey of 30 implementations of the ASTM 1394-97 standard for analyser interfaces.

Martin, W (2015). New health identification number for every individual. www.irishtimes. com/news/health/new-health-identification-number-for-every-individual-1.2135560. Online; Accessed: 2015-09-30.

Masolo, C., Borgo, S., Gangemi, A., Guarino, N., and Oltramari, A. (2003). WonderWeb Deliverable D18. *Communities*, 2003:343.

Masolo, C., Borgo, S., Gangemi, a., Guarino, N., Oltramari, a., and Schneider, L. (2002). {DOLCE}: a descriptive ontology for linguistic and cognitive engineering. *WonderWeb Project, Deliverable D17 v2*, 1(3.2):2–3.

McGuinness, D. L. and van Harmelen, F. (2004). OWL Web Ontology Language Overview. Technical report.

Moehr, J. R. (1994). Health informatics - a scientific challenge? *Methods of information in medicine*, 33(3):250–3; discussion 282–4.

Moreno-Conde, a., Moner, D., Cruz, W. D. D., Santos, M. R., Maldonado, J. a., Robles, M., and Kalra, D. (2015). Clinical information modeling processes for semantic interoperability of electronic health records: systematic review and inductive analysis. *Journal of the American Medical Informatics Association*, 22(4):925–34.

Nadkarni, P. M., Marenco, L., Chen, R., Skoufos, E., Shepherd, G., and Miller, P. (1999). Organization of Heterogeneous Scientific Data Using the EAV/CR Representation. *Journal of the American Medical Informatics Association*, 6(6):478–493.

National E-Health Transition Authority (2006). Privacy Blueprint - Unique Healthcare Identifiers.

NCIT (2015). National Cancer Institute Thesaurus. http://ncicb.nci.nih.gov/xml/owl/EVS/ Thesaurus.owl#C68806. Online; Accessed: 2015-06-30.

Nelson, E. K., Piehler, B., Eckels, J., Rauch, A., Bellew, M., Hussey, P., Ramsay, S., Nathe, C., Lum, K., Krouse, K., Stearns, D., Connolly, B., Skillman, T., and Igra, M. (2011). LabKey Server: an open source platform for scientific data integration, analysis and collaboration. *BMC bioinformatics*, 12(1):71.

NHS (2008). Personal Demographics Service (PDS) - A guide for general practice.

NHS (2011). Personal Demographics Service (PDS) Version 3. 44(0).

Noumeir, R. (2008). Integrating the healthcare enterprise process. *International Journal of Healthcare Technology and Management*, 9(2):167.

Noy, N. F., Crubezy, M., Fergerson, R. W., Knublauch, H., Tu, S. W., Vendetti, J., and Musen, M. a. (2003). Protégé-2000: an open-source ontology-development and

knowledge-acquisition environment. *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium*, page 953.

Nugroho, A. and Lange, C. F. J. (2008). On the relation between class-count and modeling effort. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5002 LNCS:93–104.

Ocean Informatics (2013). Ocean Informatics Archetypes Editor. https://oceaninformatics. com/solutions/knowledge_management. Online; Accessed: 2015-09-30.

OMG (1989). Object Management Group. www.omg.org. Online; Accessed: 2015-09-30.

OMG (2000). The Person Identification Service (PIDS) Specification. (March).

OMG (2009a). Entity Identification Service (EIS) Specification. (April).

OMG (2009b). Identity Cross-Reference Service (IXS) Version 1 (EIS).

openEHR (1999). OpenEHRan open domain-driven platform for developing flexible e-health systems. www.openehr.org. Online; Accessed: 2015-09-30.

OpenEHR (2005). Archetype Definition Language - ADL 2. *The openEHR foundation*, pages 1–117.

OpenEHR (2006). Archetype Definition Language - ADL 1.4. *The openEHR foundation*, 167(1-3):911–915.

openEHR (2008). Node+Path Persistence. https://openehr.atlassian.net/wiki/pages/ viewpage.action?pageId=6553626. Online; Accessed: 2015-09-30.

openEHR (2014). ADL Workbench release 2.0.5. www.openehr.org/downloads/ ADLworkbench/home. Online; Accessed: 2015-09-30.

openEHR (2015). Modelling Toolsopenehr website. http://www.openehr.org/downloads/ modellingtools. Online; Accessed: 2015-09-30.

OpenLIMS (2013). OpenLIMS Web Site. http://www.open-lims.org/. Online; Accessed: 2015-06-30.

OpenMRS (2004). Open Medical Record System Web Site. http://openmrs.org/. Online; Accessed: 2015-06-30.

Pilgrim, J. v. (2007). Measuring the Level of Abstraction and Detail of Models in the context of MDD. In *Models in Software Engineering - Workshops and Symposia at MoDELS 2007, Nashville, TN, USA, September 30 - October 5, 2007, Reports and Revised Selected Papers*, number October, pages 105–114.

Powell, J. and Buchan, I. (2005). Electronic health records should support clinical research. *Journal of medical Internet research*, 7(1):e4.

Protégé (2011). *The Protégé Ontology Editor*, volume 19.

Reed, A. (2002). The Definitive Guide to Identity Management. *Rainbow Technologies*.

Regan, F. and Taylor, C. (2002). Blood transfusion medicine. *BMJ (Clinical research ed.)*, 325(7356):143–7.

Reinwald, B., Lehman, T. J., Pirahesh, H., and Gottemukkala, V. (1996). Storing and using objects in a relational database. *IBM Systems Journal*, 35:172–191.

Rossi Mori, A. and Consorti, F. (1998). Exploiting the terminological approach from CEN/TC251 and GALEN to support semantic interoperability of healthcare record systems. *International journal of medical informatics*, 48(1-3):111–24.

Sauermann, L., Elst, L., and Dengel, A. (2007). PIMO - A Framework for Representing Personal Information Models. *Proceedings of I-Semantics' 07*, pages 270–277.

Schultz, A., Matteini, A., Isele, R., Bizer, C., and Becker, C. (2011). LDIF - Linked Data Integration Framework. *Proceedings of the 11th Interational Semantic Web Conference ISWC2011*, 782:1–6.

Sermersheim, J. (2006). Lightweight Directory Access Protocol (LDAP): The Protocol.

Shaw, M. J. and Mahoney, J. T. (2000). Information Rules: A Strategic Guide to the Network Economy. *Academy of Management Review*, 25(2):441–443.

Sheth, A. P. and Larson, J. A. (1990). Federated database systems for managing distributed, heterogeneous, and autonomous databases. *ACM Computing Surveys*, 22(3):183–236.

Shirky, C. (2005). Ontology is overrated: Categories, links, and tags.

Slezak, T., Wagner, M., Yeh, M., Ashworth, L., Nelson, D., Ow, D., Branscomb, E., and Carrano, A. (1995). A database system for constructing, integrating, and displaying physical maps of chromosome 19. *Proceedings of the Twenty-Eighth Annual Hawaii International Conference on System Sciences*, 5:14–23 vol. 5.

Smith, B. (2012). On Classifying Material Entities in Basic Formal Ontology. In *Proceedings of the Third Interdisciplinary Ontology Meeting*, pages 1–13. Interdisciplinary Ontology.

Smith, B. and Ceusters, W. (2006). HL7 RIM: an incoherent standard. *Studies in health technology and informatics*, 124(August):133–8.

Smith, B. and Grenon, P. (2006). Basic formal ontology (bfo). *INFOMIS Reports*.

Späth, M. B. and Grimson, J. (2011). Applying the archetype approach to the database of a biobank information management system. *International Journal of Medical Informatics*, 80(3):205–226.

Steimann, F. (2000a). On the representation of roles in object-oriented and conceptual modelling. *Data and Knowledge Engineering*, 35(1):83–106.

Steimann, F. (2000b). On the representation of roles in object-oriented and conceptual modelling. *Data and Knowledge Engineering*, 35(1):83–106.

Steimann, F. and Vollmer, H. (2006). Exploiting practical limitations of UML diagrams for model validation and execution. *Software and Systems Modeling*, 5(1):26–47.

Sucurovic, S. (2007). Implementing security in a distributed web-based EHCR. *International journal of medical informatics*, 76(5-6):491–6.

UMBC (2015). A ontology that describe a person. http://daml.umbc.edu/ontologies/ittalks/person. Online; Accessed: 2015-06-30.

Urquhart, C. (1998). Personal knowledge: a clinical perspective from the Value and EVINCE projects in health library and information services. *Journal of Documentation*, 54(4):420–442.

van Harmelen, D. L. M. and Frank (2009). OWL 2 Web Ontology Language.

Webprotege (2015). A ontology of patient that describe patient information. http://webprotege.stanford.edu/#Edit:projectId=3c107af8-f61f-47f7-89c1-fbcdfe9b2b06. Online; Accessed: 2015-06-30.

WHO (2010). *International Statistical Classification of Diseases and Related Health Problems (International Classification of Diseases)(ICD) 10th Revision - Version:2010*, volume 1.

Wiederhold, G. (1992). Mediators in the architecture of future information systems. *Computer*, 25(3):38–49.

Xu, Y., Sauquet, D., Zapletal, E., and Degoulet, P. (1999). Using XML in a generic model of mediators. pages 697–705.

Yasnoff, W. a., O'Carroll, P. W., Koo, D., Linkins, R. W., and Kilbourne, E. M. (2000). Public health informatics: improving and transforming public health in the information age. *Journal of public health management and practice : JPHMP*, 6(6):67–75.

Zemach, E. M. (1970). Four Ontologies. *The Journal of Philosophy*, 67(8):231.

Glossary

Allied Health Professional A general term that refers to a number of health professional roles other than "front line" staff such as doctors or nurses. Included in this category are biomedical scientists, medical physicists, speech and language therapists, dieticians and others. Although some of these individuals do directly interact with patients, their day to day focus in many cases is on subjects of information other than the patient. For example, biomedical scientists are generally focused on specimens and orders.

American National Standards Institute A private non-profit organisation that is consists of the company, the government, and other members, is to coordinate and review U.S. standards.

American Society for Testing and Materials An SDO that was responsible for among other specifications HL7-style instrument messaging protocols where identity focused on patient but also other non human subjects such as specimen, order and instrument.

Archetype The second level of two level models. Archetypes can be developed by domain experts using a semi-graphical tool, to reflect some feature of healthcare documentation or health identity management. So for example, a valid "health data" archetype would be "Blood Pressure" while a suitable identity archetype could be "Blood Product Vendor Contact Details".

Archetype Definition Language A formal language that is used to define the archetypes..

Archetype Object Model An object model that is used to describe the structural properties of archetypes.

Archetype-able It means that the information is able to be expressed by the archetype approach.

Basic Formal Ontology an ontology that acts as a very abstract root of other ontologies and is closely related to the idea of identified entity that is adopted in this work.

Biomedical Research Integrated Domain Group A "static" Domain Analysis ModeI (DAM) which is developed by CDISC in 2003 aims to make standards of clinical research data that can be exchanged with HL7.

CEN TC251 The CEN technical committee that is responsible for developing eHealth standards in Europe. At time of writing, most CEN TC251 work also comes under the banner of ISO TC215 which is the equivalent technical committee of ISO.

Clinical Document Architecture The CDA, which provides an exchange model for clinical documents (such as discharge summaries and progress notes) and CDA documents are encoded in XML as a basis for describing the structure of medical records though the development of schema by acting as a data definition standard.

Clinical Information Modelling Initiative A cross-Atlantic initiative for developing detailed clinical models that are intended for use in a numner of different formats.

Committee European Normalisation European committee for standardisation, the European SDO. The nationsl standards bodies of all member states feed into this body.

Data Types The categories of data that are contained in standard information model. For example, the II is defined as a data type for identifiers in CEN 13606.

Demographic Service The service constitutes the discrete components of demographics application that define loosely message-based interfaces for access by users across the networks.

Document Type Definition A set of tags of the grammar rules which is part of authentication mechanism of the XML documents.

EHR Service A set of components of applications that include all EHR data management processing and accessibility functions associated with a community of EHR resources.

EHR System This term is used with similar meaning to an EHR Service but focuses on the concept of integrated whole of EHR applications in particular.

EHR*land* A 30 month HIQA-funded research project that investigated the use of EN13606 for a future Irish EHR system. The author's work was partially conducted as part this project.

EHRcom A commonly used name for the ISO EN13606 family of specifications for EHR communication..

Electronic Health Record *"A repository of information regarding the health of a subject of care in computer processable form, stored and transmitted securely, and accessible by multiple authorised users. It has a commonly agreed logical information model which is independent of EHR systems"* (ISO/TC 215, 2005).

Enterprise Master Patient Index A system that organises client identity across multiple systems to collect and store specific person identified demographic information from the source system (track new IDs, track changes of existing IDs).

Entity Attribute Value A mechanism that has been used among other things to store information "dynamically" in a relational database, without requiring a schema.

Entity Identification Service The second evolution of the OMG/HSSP identity specifications. Following the relative success of PIDS, tt was the first incarnation of these specifications to include the idea of generalised identity.

Extensible Stylesheet Language Stylesheet Language that can be used for displaying XML documents.

Extensive Markup Language A cross platform language and it describe the data content on the network and structure of the standard.

Health and Information Quality Authority of Ireland The Health and Information Quality Authority of Ireland, responsible for overseeing good practice in healthcare and in use of healthcare data in Ireland..

Health Level Seven International One of the main international Healthcare SDOs.

Healthcare Services Specification Program A collaboration arising out of a memorandum of understanding between HL7 and OMG to jointly develop ehealth interoperability standards.

HL7 Reference Information Model A static information model that was introduced as part of HL7 V3.

HL7 version 2 A wide-ranging set of specifications for messages containing healthcare information.

HL7 version 3 A revision of the HL7 v2 that depends on an object model to describe the subjects of information.

Hospital Information System A system that manages among other things, the logistics of patient visits to hospital, including room, bed and appointment management. As such, a HIS typically encapsulates an EMPI in addition to a number of identity domains relating to hospital resources.

Identification Service The most recent in a series of specifications for managing generalised identities in healthcare that was developed as part of the HL7 / OMG HSSP initiative.

Identity Cross-Reference Service The second iteration of the OMG/HSSP identity specifications.

Identity Domain An independent storage area is to assign the identities such as hospitals, medical institutions or standardised organisations etc.

Identity Linking A technology or a service provides the linkage function among the participants in healthcare with same unique health identifier or additional demographic data.

Identity Merging A technology or a service allows the different forms of identities to be unified into one for reducing the identity redundancy.

Identity Service A service that allows for management of identity.

Information Model A conventional representation of information used to define the method that reuses the managing data for different applications.

Integrating the Healthcare Enterprise An organisation who promotes healthcare interoperability by identifying practical interoperability use cases in healthcare and developing practical profiles and guidance about implementing ehealth interoperability standards. They have developed the PIX and PDQ profiles for identification.

Interoperability The ability of different computerised system, networks, operating systems and applications work together, and to share information.

Laboratory Information Management System Also called a Laboratory Information System is an information system that is used to help oversee manage the process of analysing specimens and reporting the results of those investigations. In the clinical laboratory, specimens are typically biological in nature such as treated blood, urine or faecal material.

Laboratory Information System see Laboratory Information Management System.

Master Person Index Also called Patient Master Index - provides an index referencing for patients in computer-based medical system.

National Health Service The national health service of the United Kingdom. Many features of the Irish ehealth system are modelled on those found in the NHS.

Object Management Group An SDO who develop specifications about object-based interoperability. They developed PIDS and collaborated with HL7 in the development of the IXS and IS specifications.

Patient Identification segment A set of patient identification information that forms part of HL7 message. It typically contains different profiles containing traits for identifying subjects of care. It consequently facilitates the transfer of demographic information according to the requirements of a compliant ehealth system.

Patient Master Index see Patient Master Index.

Person Identification Service The original distributed identity service in the series that led to the HSSP IS. It was developed by OMG Corbamed, later rcalled the OMG health domain taskforce and focused on identification of human subjects only.

Personal Demographics Service a national demographics service that was introduced by the NHS.

Picture Archiving and Communication System Somewhat overlaps with RIS (Radiology Information System). A PACS manages and provides access to medical images such as x-rays, and MRI images.

Referent Refers to all participates in healthcare such as objects, actions, persons, organisations, device etc..

Semantic Signifier An associated set of validation instructions for the EIS Traits that apply to a particular type of entity in a particular context.

Simple Object Access Protocol An XML-based communication protocol of information network based on decentralised and distributed environment for network information. Under the agreement, software components or applications communicate each other through the standard HTTP protocol.

Synapses Object Model and Synapses Object Dictionary The object model and object dictionary of Synapses project. They correspond to reference model and Archetypes respectively.

Trait An individual characteristics of demographic information such as name, address, telephone number etc.

Trait Set A set of traits which is usually assigned by certain domains or organisations for standardised data exchange and can be applied to a particular type of entity.

Unified modelling Language A set of standardised modelling formats for representing static and dynamic aspects of object models. This specification is managed by OMG.

Uniform Resource Identifier A general format for holding links to physical or abstract resources. A uniform resource locator (URL) that refers to web resource is a type of URI.

Web Ontology Language A derivation of RDF that is used to express ontologies for the web.

XML schema definition A XML structure definition and is also a substitute for DTD.

THE MEDIATED SCHEMAS OF GIRM REPRESENTATION

Listing A.1 A raw XML schema of GIRM "Identified Entity"

```xml
<!-- Generalised Identity Reference Model 1.0.1 XML schema -->
<xs:include schemaLocation="GIRM−identity_component.xsd" />

<xs:element name="IDENTITY" type="IDENTITY" />
<xs:complexType name="IDENTITY">
  <xs:complexContent>
    <xs:extension base="IDENTITY_COMPONENT">
      <xs:sequence>
        <xs:element name="traits" type="TRAIT" minOccurs="0"
          maxOccurs="unbounded" />
        <xs:element name="unique_identifier" type="II" minOccurs="0"
          maxOccurs="unbounded" />
        <xs:element name="replacing_version" type="CV" minOccurs="0"
          maxOccurs="unbounded" />
      </xs:sequence>
    </xs:extension>
  </xs:complexContent>
</xs:complexType>

<xs:element name="IDENTITY_LINK" type="IDENTITY_LINK" />
<xs:complexType name="IDENTITY_LINK">
  <xs:complexContent>
    <xs:extension base="IDENTITY">
      <xs:sequence>
        <xs:element name="name" type="TEXT" minOccurs="0"
```

```
26              maxOccurs="unbounded" />
27          <xs:element name="source_target" type="URI" minOccurs="0"
28              maxOccurs="unbounded" />
29          <xs:element name="destination_target" type="URI"
30              minOccurs="0" maxOccurs="unbounded" />
31        </xs:sequence>
32      </xs:extension>
33    </xs:complexContent>
34  </xs:complexType>
35
36  <xs:complexType name="TRAIT" abstract="true">
37    <xs:complexContent>
38      <xs:extension base="IDENTITY_COMPONENT">
39      </xs:extension>
40    </xs:complexContent>
41  </xs:complexType>
42
43  <xs:element name="SIMPLE_TRAIT" type="SIMPLE_TRAIT" />
44  <xs:complexType name="SIMPLE_TRAIT">
45    <xs:complexContent>
46      <xs:extension base="TRAIT">
47        <xs:sequence>
48          <xs:element name="data_value" type="DATA_VALUE" />
49        </xs:sequence>
50      </xs:extension>
51    </xs:complexContent>
52  </xs:complexType>
53
54  <xs:element name="COMPLEX_TRAIT" type="COMPLEX_TRAIT" />
55  <xs:complexType name="COMPLEX_TRAIT">
56    <xs:complexContent>
57      <xs:extension base="TRAIT">
58        <xs:sequence>
59          <xs:element name="trait_part" type="TRAIT_PART"
60              minOccurs="0" maxOccurs="unbounded" />
61        </xs:sequence>
62      </xs:extension>
63    </xs:complexContent>
64  </xs:complexType>
65
66  <xs:element name="TRAIT_PART" type="TRAIT_PART" />
```

```
67  <xs:complexType name="TRAIT_PART">
68    <xs:sequence>
69      <xs:element name="data_value" type="DATA_VALUE" />
70      <xs:element name="trait_part" type="TRAIT_PART" minOccurs="0"
71        maxOccurs="unbounded" />
72    </xs:sequence>
73  </xs:complexType>
74
75  </xs:schema>
```

Listing A.2 A raw XML schema of GIRM "Identity Component"

```
1   <?xml version="1.0" encoding="utf-8" ?>
2     <!-- GIRM Identity_Componenet 1.0.1 XML schema -->
3     <!-- Authored by TeaPOT in DIT 2010.09.14  -->
4   <xs:schema version="v1.0.2" targetNamespace="http://schemas.ehrland.ie/v1"
5     elementFormDefault="qualified" xmlns:xs="http://www.w3.org/2001/XMLSchema"
6     xmlns="http://schemas.ehrland.ie/v1" id="GIRM−identity_component.xsd">
7     <xs:include schemaLocation="GIRM−dataTypes.xsd" />
8
9     <xs:element name="identity_component" type="IDENTITY_COMPONENT" />
10    <xs:complexType name="IDENTITY_COMPONENT" abstract="true">
11      <xs:sequence>
12        <xs:element name="name" type="TEXT" />
13        <xs:element name="archetype_id" type="xs:string"
14          minOccurs="0" maxOccurs="1" />
15        <xs:element name="validity_time" type="TS" minOccurs="0" />
16      </xs:sequence>
17    </xs:complexType>
18  </xs:schema>
```

Listing A.3 A raw XML schema of GIRM datatypes

```
1   <?xml version="1.0" encoding="UTF-8"?>
2     <!-- GIRM DataTypes 1.0.1 XML schema -->
3   <xs:schema version="v1.0.2" targetNamespace="http://schemas.ehrland.ie/v1"
4     elementFormDefault="qualified" xmlns:xs="http://www.w3.org/2001/XMLSchema"
5     xmlns="http://schemas.ehrland.ie/v1" id="GIRM−dataTypes.xsd">
6
7     <xs:complexType name="CODE">
8       <xs:sequence>
9         <xs:element name="value" type="xs:string" />
```

```xml
10      </xs:sequence>
11    </xs:complexType>
12
13    <xs:complexType name="DATA_VALUE" abstract="true">
14      <xs:sequence>
15        <xs:element name="nullFlavor" type="CS" minOccurs="0" />
16      </xs:sequence>
17    </xs:complexType>
18
19    <xs:complexType name="CS">
20      <xs:complexContent>
21        <xs:extension base="DATA_VALUE">
22          <xs:sequence>
23            <xs:element name="codingScheme" type="OID" minOccurs="0" />
24            <xs:element name="codingSchemeName" type="xs:string"
25              minOccurs="0" />
26            <xs:element name="codingSchemeVersion" type="xs:string"
27              minOccurs="0" />
28            <xs:element name="codeValue" type="xs:string" minOccurs="0" />
29          </xs:sequence>
30        </xs:extension>
31      </xs:complexContent>
32    </xs:complexType>
33
34    <xs:complexType name="OID">
35      <xs:sequence>
36        <xs:element name="oid" type="xs:string" minOccurs="0" />
37      </xs:sequence>
38    </xs:complexType>
39
40    <xs:complexType name="CV">
41      <xs:complexContent>
42        <xs:extension base="DATA_VALUE">
43          <xs:sequence>
44            <xs:element name="codingScheme" type="OID" minOccurs="0" />
45            <xs:element name="codingSchemeName" type="xs:string"
46              minOccurs="0" />
47            <xs:element name="codingSchemeVersion" type="xs:string"
48              minOccurs="0" />
49            <xs:element name="codeValue" type="xs:string" minOccurs="0" />
50            <xs:element name="displayName" type="xs:string"
```

```xml
                  minOccurs="0" />
          </xs:sequence>
        </xs:extension>
      </xs:complexContent>
    </xs:complexType>
    <xs:complexType name="TS">
      <xs:complexContent>
        <xs:extension base="DATA_VALUE">
          <xs:sequence>
            <xs:element name="time" type="dateTime-time" />
          </xs:sequence>
        </xs:extension>
      </xs:complexContent>
    </xs:complexType>
    <xs:simpleType name="dateTime-time">
      <xs:union memberTypes="xs:dateTime xs:time" />
    </xs:simpleType>
    <xs:complexType name="DURATION">
      <xs:complexContent>
        <xs:extension base="QUANTITY">
          <xs:sequence>
            <xs:element name="days" type="xs:integer" minOccurs="0" />
            <xs:element name="hours" type="xs:integer" minOccurs="0" />
            <xs:element name="minutes" type="xs:integer" minOccurs="0" />
            <xs:element name="seconds" type="xs:integer" minOccurs="0" />
            <xs:element name="fractional_second" type="xs:double"
              minOccurs="0" />
            <xs:element name="sign" type="xs:integer" minOccurs="0" />
          </xs:sequence>
        </xs:extension>
      </xs:complexContent>
    </xs:complexType>
    <xs:complexType name="QUANTITY" abstract="true">
      <xs:complexContent>
        <xs:extension base="DATA_VALUE">
          <xs:sequence />
        </xs:extension>
      </xs:complexContent>
    </xs:complexType>
    <xs:complexType name="IVLTS">
      <xs:complexContent>
```

```
92        <xs:extension base="DATA_VALUE">
93          <xs:sequence>
94            <xs:element name="low" type="TS" minOccurs="0" />
95            <xs:element name="high" type="TS" minOccurs="0" />
96            <xs:element name="lowClosed" type="xs:boolean"
97              minOccurs="0" />
98            <xs:element name="highClosed" type="xs:boolean"
99              minOccurs="0" />
100           <xs:element name="width" type="DURATION" minOccurs="0" />
101         </xs:sequence>
102       </xs:extension>
103     </xs:complexContent>
104   </xs:complexType>
105   <xs:complexType name="II">
106     <xs:complexContent>
107       <xs:extension base="DATA_VALUE">
108         <xs:sequence>
109           <xs:element name="extension" type="xs:string" minOccurs="0" />
110           <xs:element name="assigningAuthorityName" type="xs:string"
111             minOccurs="0" />
112           <xs:element name="validTime" type="IVLTS" minOccurs="0" />
113           <xs:element name="root" type="OID" />
114         </xs:sequence>
115       </xs:extension>
116     </xs:complexContent>
117   </xs:complexType>
118   <xs:complexType name="TEXT" abstract="true">
119     <xs:complexContent>
120       <xs:extension base="DATA_VALUE">
121         <xs:sequence>
122           <xs:element name="originalText" type="xs:string"
123             minOccurs="0" />
124           <xs:element name="language" type="CS" minOccurs="0" />
125           <xs:element name="charset" type="CS" minOccurs="0" />
126         </xs:sequence>
127       </xs:extension>
128     </xs:complexContent>
129   </xs:complexType>
130   <xs:complexType name="SIMPLE_TEXT">
131     <xs:complexContent>
132       <xs:extension base="TEXT">
```

```
133          <xs:sequence />
134        </xs:extension>
135      </xs:complexContent>
136    </xs:complexType>
137    <xs:complexType name="URI">
138      <xs:complexContent>
139        <xs:extension base="DATA_VALUE">
140          <xs:sequence>
141            <xs:element name="value" type="xs:string" minOccurs="0" />
142            <xs:element name="scheme" type="xs:string" minOccurs="0" />
143            <xs:element name="path" type="xs:string" minOccurs="0" />
144            <xs:element name="fragment_id" type="xs:string"
145              minOccurs="0" />
146            <xs:element name="uri_query" type="xs:string" minOccurs="0" />
147            <xs:element name="literal" type="xs:string" minOccurs="0" />
148          </xs:sequence>
149        </xs:extension>
150      </xs:complexContent>
151    </xs:complexType>
152  </xs:schema>
```

A LARGE-SCALE TWO-LEVEL IDENTITY SERVICE ARCHITECTURE

Introduction

This appendix discusses topics associated with developing a large scale two-level identity service architecture. This leads to multiple issues that are associated with the requirements based on enterprise view point as described in the following sections.

A shared EHR system requires supporting components at both a national level at each EHR "site". In the proposed system, a regional EHR system while supporting archetype repositories and terminology services would provide access to registers of identities for patients (Regional Health Identity Service) and health professionals (Regional HCP Identity service). Each EHR system site would have access to these regional resources. In addition to the clinical information services at each EHR system site, the local identity services would support the mapping of the local patient and health care provider identities to the regional equivalents.

If the such systems need to be implemented, the identity service should include the following functionality:

- A reference implementation of EHRcom compatible Demographics Identity Service which uses the interfaces adopted from HSSP Identity Cross-Reference Service (IXS) specification (OMG, 2009b).

- Support for cross linking (or cross referencing) identities between disparate patient domains or disparate HCP domains for three types of entities: Patients, Health Professionals or Providers in a centralised national domain according to the CEN 13606 demographics model.

- The general identity management operations of register retrieve and query for data entities that allow the actors defined within the EHRcom standards to find, exchange and reference demographics entity data while saving and associating the data.

- Support for linking the patient identities to the identity of their GP, consultant, or hospital department that they visit.

Main actors in the business domain

It is supposed that the following participants which exist in most cases of health domains can be supported by this service architecture but only person entities related to health professional are included in this design of demographics identity service. However, identified entities incorporated into the service include,

- **Person** - Patient/Healthcare Professional (incorporating different EHRcom roles such as Composer, Committer, Responsible HCP)

- **Organisation** - GP/Hospital, Healthcare Provider, Healthcare research institution, Software or Medical device

Overview according to open distributed processing

The previous section provides a high level description a proposed architecture for a distributed electronic health record system that is based on the two-level modelling approach. This section presents the identity service design that is compatible with the enterprise viewpoint of Reference Model of Open Distributed Processing (RM-ODP) framework with the integration of two-level models.

RM-ODP includes Enterprise Viewpoint, Information Viewpoint , Computation viewpoint, Engineering viewpoint and Technology viewpoint. The main body of this document work corresponds to the ODP Information Viewpoint. This appendix focuses on the viewpoint that is typically site independent, namely the Enterprise. One key benefits of documenting according to ODP viewpoints in the author's view is to effectively deal with the inherent complexity of large and legacy systems. This service design has arisen out of experimentation and extensive discussion over the course of the work. The enterprise architecture describes some of the main information services that would be required at a national or regional level and those that would need to be merged with the existing relevant ICT resources at each participating site. The following services are considered to be necessary elements of a working design.

"National" Identity Service

This service would provide the functions of identity management that allow certain types of users to register identities, retrieve identities and manage identities for multiple entities within the regional health domain such as patients, HCPs, organisations and devices. In a jurisdiction like Ireland, where there has initially been no national system, each healthcare site will have a sub-domain of the main regional domain, each having its own legacy identities and it correlates these identities (e.g Patient, GP) to establish a standardised information view, which provides a basis for an EHR system. This service employs an

approach that is similar to the HSSP Identity Cross-reference Service (IXS) to extend the management of identity to provide cross-referencing query and administrative functions for an identified entity.

Local Identity Service

This component typically acts as a service to the composition committer while entering the new clinical data into the EHR. It is also notified of changes in the local demographics data by the demographic observer. Following changes in local demographics data, it is responsible for interacting with the national identity service and it may also store national IDs obtained from the national service.

Local Access Control Service

On top of national access control policy, which focuses essentially on role-based access control, the EHR provider may wish to enforce additional policies, based, for example, on the ID of the requester, whether it corresponds to a professional or an entire organisation. The role of a local access control service will specifically be to enforce these locally determined access control policies.

Demographic Observer

This component observes changes in the relevant demographics data sources and informs the local identity service. It is configured by the EHR provider configuration tools to interact with the local demographics data sources and the local identity service.

Record Identity Manager

This component is responsible for delivering unique UUIDs, or URIs, or indeed IIs with suitable root OIDS (depending on the approach to identifiers) and (for example) assigning authority names to the composition committer at committal time..

ODP - Enterprise view point

In chapters two and three of the main work, some of the fundamental concepts behind two-level models were described as well as an overview of some of the main classes used in the standard, in the reference model that is the focus of part one of the standard. Chapter four of the main document described what corresponds to the Information Viewpoint for an Identity Service. This section corresponds to the ODP Enterprise Viewpoint; it provides a set of use cases and use case descriptions and scenarios associated with implementation of the EN13606 standard.

National identity service use cases

The following sets of diagrams are provided to show how the author expects the identity service to be used. They should be entirely consistent with the interface descriptions above.

Overview

The use cases of demographics identity service can be classified into three tiers by the level of detail as shown in Figure B.1. Each use cases will be described from the tier of general roles to specific roles.

Description of actors:

- Cross-reference Agent: defined in EN13606 as an "Identified_Entity" which contains the sub-classifiers and it may be an organisation, person, or device or software. In this service, cross-reference agent is also defined as the consumer or user of demographics identity service.

- Cross-reference Provider: is defined in ISO/TS 27527 as an "Any person or organisation that is involved in or associated with the delivery of health services to a client, or
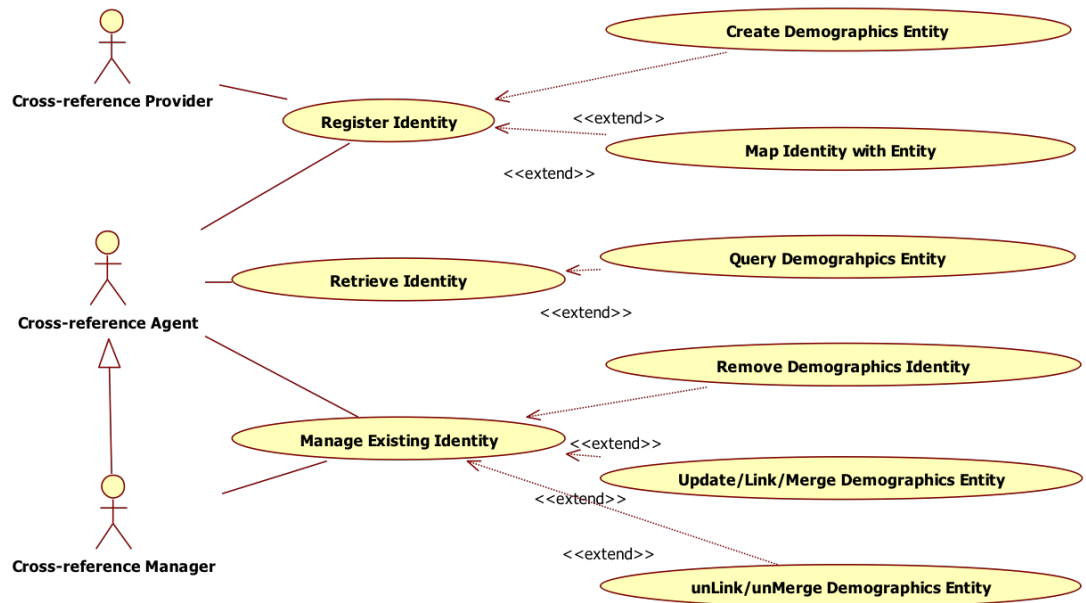
Figure. B.1 Overview of use cases

caring for client well-being" and it is also defined as demographics identity service provider in a position to register and query identified entity.

- Cross-reference Manager: is the function of the administration of demographics identity service and it has greater privilege for administrative operation such as remove, link, and update entities.
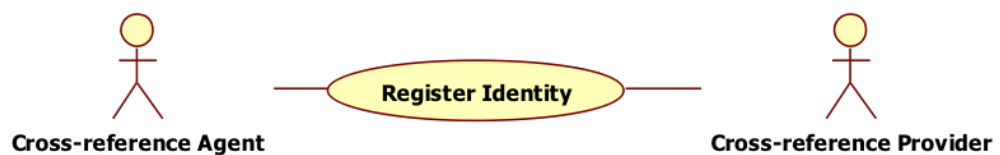
Primary Use Case 1: Register Identity

Figure. B.2 The use case of registering identity

Summary: the goal of this use case is to register an identity (EN13606 Identi-fied_Entity) from cross-reference provider as shown in Figure B.2. The main actors are the cross-reference agent and the cross-reference provider.

Primary scenario:

- The cross-reference agent sends a request to the cross-reference provider. The request contains the identification information defined in EN13606 demographics reference model. The cross-reference provider will create and store demographics entity and mapping between identifiers and entity IDs.

- The cross-reference provider authenticates the cross-reference agent.

- The cross-reference provider checks the access privileges of the cross-reference agent for the subject of care and the query.

- The cross-reference provider sends back an acknowledgement message with IXSSta-tus response.

Primary Use Case 2: Manage Existing Identity
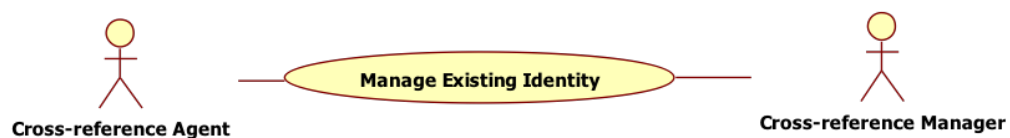
Figure. B.3 The use case of managing the existing identities

Summary: the goal of this use case is to manage existing identity (CEN EN13606 Identified_Entity) from a cross-reference manager as shown in Figure B.3. The main actors are the cross-reference agent and the cross-reference manager.

Primary scenario:

- The cross-reference agent sends a request to the cross-reference manager. The request contains the identification information defined in CEN EN13606 demographics reference model with the full or partial traits. The cross-reference manager links merges and activates the entity and also operates in the reverse.

- The cross-reference manager authenticates the audit log reviewer.

- The cross-reference manager checks the access privileges of the cross-reference agent for the administrative operation of demographics identity service.

- The cross-reference manager sends back an acknowledgement message with IXSStatus response.

Primary Use Case 3: Retrieve Identity
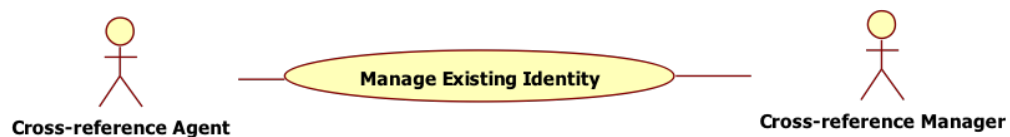
Figure. B.4 The use case of retrieving identitiy

Summary: the goal of this use case is to query (CEN EN13606 Identified_Entity) from a cross-reference provider as shown in Figure B.4. The main actors are the cross-reference agent and the cross-reference provider.

Primary scenario:

- The cross-reference agent sends a request to the cross-reference provider. The request contains the full or partial queried traits (identification information) defined in CEN EN13606 demographics reference model. The cross-reference provider returns a list of linked identifiers of patient records from all other domains that are linked to the queried information about one particular patient in a specified domain.

207

- The cross-reference provider authenticates the cross-reference agent.

- The cross-reference provider checks the access privileges of the cross-reference agent for the subject of care and the query.

- The cross-reference provider sends back an acknowledgement message with IXSStatus response.

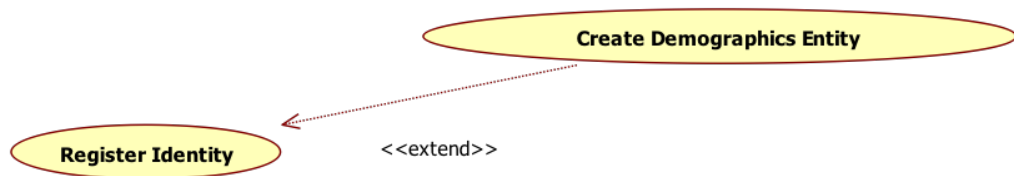**Extensions of Primary Use Case 1: Create Demographics Entity**

Figure. B.5 The use case of creating demographic entity

Primary scenario:

- The cross-reference agent sends the request to create a new patient entity (CEN EN13606 Identified_Entity) if the patient entity does not exist as shown in Figure B.5.

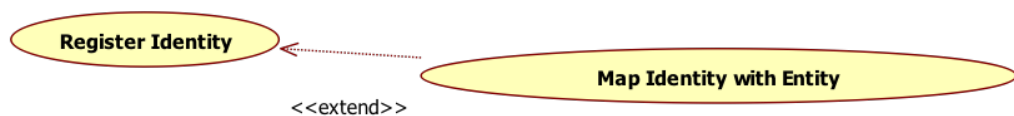**Extensions of Primary Use Case 2: Map Identity with Entity**

Figure. B.6 The use case of mapping identity with entity

Primary scenario:

- The cross-reference agent sends the request to create an ID mapping for a new patient identity with an entity (CEN EN13606 Identified_Entity) if the patient entity exists as shown in Figure B.6.

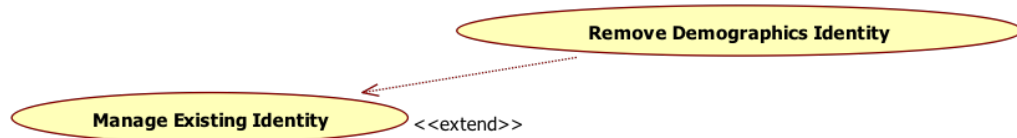**Extensions of Primary Use Case 3: Remove Demographics Identity**

Figure. B.7 The use case of removing demographics identity

Primary scenario:

- The cross-reference agent sends the request to delete an identity/entity pair and its associated traits from the repository and it is done by cross-reference manager as shown in Figure B.7.

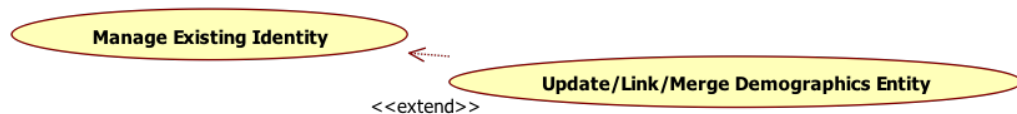**Extensions of Primary Use Case 4: Update / Link / Merge Demographics Identity**

Figure. B.8 The use case of updating/linking/merging demographics identity

Primary scenario as shown in Figure B.8:

- The cross-reference agent sends the request to update an entity with required demographics traits and it is done by cross-reference manager.

- The cross-reference agent sends the request to create an ID link between a supplied entity and target entity and it is done by cross-reference manager.

209

- The cross-reference agent sends the request to consolidate two duplicate entities and it is done by cross-reference manager.

Extensions of Primary Use Case 5: unlink/ unMerge Demographics Identity
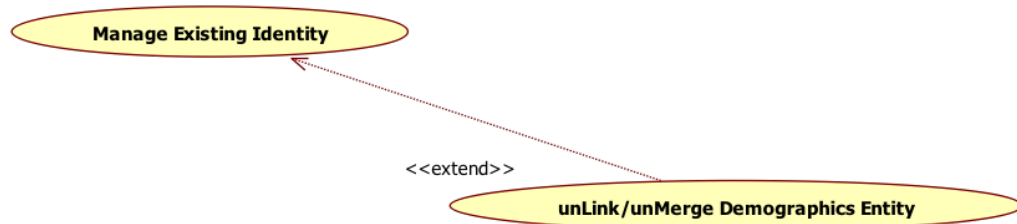

Figure. B.9 The use case of unlinking/unmerging demographics identity

Primary scenario as shown in Figure B.9:

- The cross-reference agent sends the request to reverse the link demographics entity process and it is done by cross-reference manager.

- The cross-reference agent sends the request to reverse the merge demographics entity process and it is done by cross-reference manager.

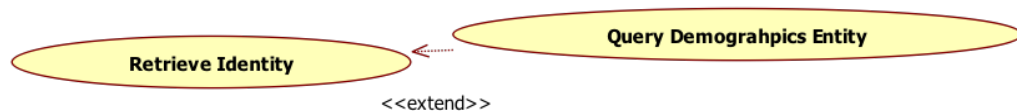**Extensions of Primary Use Case 6: Query Demographics Entity**

Figure. B.10 The use case of querying demographics intity

Primary scenario as shown in Figure B.10:

- The cross-reference agent sends the request to query a patient identity and cross-reference provider returns the single entity or the multiple entities associated with queried traits.

- The cross-reference agent sends the request to ask the national xIXS to give the national IHI that corresponds to the local ID.

- The cross-reference agent sends the request to query the national xIXS to give the national IHI that corresponds to the traits from the local entity.

- The cross-reference agent sends the request to query the national xIXS to give the full entity that matches a IHI.

LIST OF PUBLICATIONS

Chen, X., Berry, D., and Grimson, W. (2008). Identity Management to Support Access Control in E-Health Systems. In *IFMBE Proceedings*, volume 22, pages 880–886. Springer Berlin Heidelberg.

Camous, F., Chen, X., Yu, S., Grimson, W., and Berry, D. (2008). EHRcom-Compliance for EHR Exchange in Ireland: Elements of an Architecture. In *Proceedings of the 13th Annual Conference and Scientific Symposium of the Healthcare Informatics Society of Ireland (HISI)*.

Berry, D., Camous, F., Chen, X., Sheng, Y., Henry, P., Pardon, S., O'Neill, A., McGrory, J., Clarke, F., Chrzanowski, K., and Garus, G. (2010). EHR*land* Final Report version 1.0.

Chen, X., Berry, D., and Stephens, G. (2015). Using a generalised identity reference model with archetypes to support interoperability of demographics information in electronic health record systems. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pages 6834–6839. IEEE.