ARCHETYPE BASED DOMAIN MODELING FOR HEALTH INFORMATION SYSTEMS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS
OF
THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

KORAY ATALAĞ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
THE DEPARTMENT OF INFORMATION SYSTEMS

JULY 2007

Approval of the Graduate School of Informatics

_____

Assoc.Prof.Dr. Nazife Baykal

Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Doctor of Philosophy.

_____

Assoc.Prof.Dr. Yasemin Yardımcı

Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Doctor of Philosophy.

_____

Prof.Dr. Semih Bilgen

Supervisor

Examining Committee Members

Prof.Dr. Sedat Boyacıoğlu	(MESA Hospital)	_____
Prof.Dr. Semih Bilgen	(METU, EEE)	_____
Assoc.Prof.Dr. Onur Demirörs	(METU, II)	_____
Assoc.Prof.Dr. Gürden Gür	(Güven Hospital)	_____
Assoc.Prof.Dr. Ü. Erkan Mumcuoğlu	(METU, II)	_____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name : Koray Atalağ

Signature : _____

ABSTRACT

ARCHETYPE BASED DOMAIN MODELING FOR HEALTH INFORMATION SYSTEMS

Atalağ, Koray

Ph.D., Department of Information Systems

Supervisor: Prof.Dr. Semih Bilgen

July 2007, 163 pages

A major problem to be solved in health informatics is high quality, structured and timely data collection. Standard terminologies and uniform domain conceptual models are important steps to alleviate this problem which are also proposed to enable interoperability among systems. With the aim of contributing to the solution of this problem, this study proposes novel features for the Archetypes and multi-level modeling technique in health information and knowledge modeling. The study consists of the development of a research prototype for endoscopic data management, and based on that experience, the extension of Minimal Standard Terminology in Digestive Endoscopy (MST). A major contribution of the study consists of significant extensions to the modeling formalism. The proposed modeling approach may be used in the design and development of health information systems based on archetypes for structured data collection, validation and dynamic user interface creation. The thesis work is aimed to make considerable contribution to the emerging Electronic Health Records (EHR) standards and specifications.

Keywords: Information systems, health information systems, domain modeling, archetypes, endoscopy.

ÖZ

SAĞLIK ENFORMASYON SİSTEMLERİ İÇİN ARKETİP TABANLI ALAN MODELLEMESİ

Atalağ, Koray

Doktora, Bilişim Sistemleri Bölümü

Tez Yöneticisi: Prof.Dr. Semih Bilgen

Temmuz 2007, 163 sayfa

Sağlık bilişiminin çözülmesi gereken önemli problemlerinden biri kaliteli, yapısal ve zamanında veri toplanmasıdır. Standart terminolojiler ve kavramsal alan modelleri bu problemi hafifletmek için önemli adımlardır ve aynı zamanda sistemler arasında birlikte çalışabilirliği sağlayabilecekleri öne sürülmektedir. Bu problemin çözümü amacıyla, bu çalışma sağlıkta enformasyon ve bilgi modellemesi için Arketipler ve çok-seviyeli modelleme tekniğine özgün katkılar önermektedir. Çalışma endoskopide veri yönetimi için bir araştırma prototipi geliştirilmesi ve edinilen tecrübeyle Gastrointestinal Endoskopide Minimal Standart Terminoloji'nin (MST) geliştirilmesini içerir. Önerilen modelleme yaklaşımı yapısal veri toplanması, doğrulama ve dinamik kullanıcı arayüzleri geliştirmek için arketip tabanlı sağlık enformasyon sistemleri tasarlanması ve geliştirilmesinde kullanılabilir. Bu tez çalışmasının yeni gelişmekte olan elektronik sağlık kayıtları (ESK) standart ve spesifikasyonlarına önemli katkılarda bulunması hedeflenmiştir.

Anahtar kelimeler: Enformasyon sistemleri, sağlık enformasyon sistemleri, alan modellemesi, arketipler, endoskopi.

To Atalağ Family ☺

especially my father, Mehmet

and

my daughter, Karya

ACKNOWLEDGMENTS

I would like to express my sincerest gratitude to my thesis supervisor, Dr. Semih Bilgen who helped this heavy ship (Ph.D. studies) out in the stormy waters (business life) and brought it back to the calm waters (academia). If he did not assure me that we could make it through the end in each and every meeting, I would not be writing these lines. I really appreciate his warm and excellent guiding, and especially for his ability to find an appropriate time slot for me in his busy schedule. It is still a wonder for me how he could read and edit my tens of pages of thesis material in just a matter of hours and always provide me with written feedback. Thank you again…

Thanks to Dr. Onur Demirörs for helping me to establish a solid understanding of software engineering discipline and to Dr. Erkan Mumcuoğlu for supporting my studies.

I also would like to thank Dr. Sedat Boyacıoğlu for getting me involved with gastrointestinal endoscopy domain and providing me with valuable knowledge. I am also grateful to Dr. Gürden Gür for his generous support during development of the research prototype. I still appreciate their continuing support, especially for participating to my thesis progress meetings and also for accepting to appear in the thesis jury.

Uninterrupted support and close collaboration of Thomas Beale and Dr. Sam Heard from openEHR have broadened my vision in this field and provided me with the knowledge and tools for conducting my research. I highly appreciate and feel deeply indebted to them.

I also acknowledge Philippe Ameline for sharing me with his technical knowledge on endoscopy and also supporting me with his genuine friendship.

Thanks to Dr. Louis Korman who always provided me with excellent insight and practical knowledge on both MST and informatics.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF BOXES

LIST OF ACRONYMS and ABBREVIATIONS

ADL:	Archetype Definition Language
ADT:	Admission, Transfer or Discharge
AM:	Archetype Model
ANSI:	American National Standards Institute
AOM:	Archetype Object Model
ASC:	Accredited Standards Committee
ASGE:	The American Society for Gastrointestinal Endoscopy
ASTM:	American Society for Testing and Materials
ATC:	Anatomic, Therapeutic, Chemical Classification System
CAP:	College of American Pathologists
CASE:	Computer Aided Software Engineering
CCR:	Continuity of Care Record
CDA:	Clinical Document Architecture
CEN:	European Committee for Standardization
CIS:	Clinical Information System(s)
CME:	Continuing Medical Education
CMET:	Common Message Element Type(s)
COAS:	Clinical Observations Access Service
CORI:	Clinical Outcomes Research Initiative
CPR:	Computer-based Patient Record(s)
CPT:	Current Procedural Terminology
CRM:	Common Reference Model
CT:	Computerized Tomography
CUI:	Concept Unique Identifier (UMLS)
DDL:	Data Definition Language
DICOM:	Digital Imagind and Communications in Medicine
DIM:	Domain Information Model(s)
DRG:	Diagnosis Related Groups
DSA:	Digital Subtraction Angiography
DSM:	Domain Specific Modeling
DTD:	Document Type Definition(s)
DVD:	Digital Video Disc
EBM:	Evidence Based Medicine
ebXML:	Electronic Business XML
ECG:	Electro Cardiography
ECRI:	Emergency Care Research Institute

EGD:	EsopgahoGastroDuodenoscopy
EHCR:	Electronic Health Care Record(s)
EHR:	Electronic Health Record(s)
EIS:	Endoscopic Information System
EMB:	IEEE Engineering in Medicine and Biology Society
EMR:	Electronic Medical Record(s)
EPR:	Electronic Patient Record(s)
ER:	Entity Relationship
ERCP:	Endoscopic Retrograde Cholangio Pancreaticoduodenoscopy
ESGE:	The European Society of Gastrointestinal Endoscopy
EU:	European Union
FOPL:	First-Order Predicate Logic
GALEN:	Generalized Architecture for Languages, Encyclopaedias and Nomenclatures in Medicine
GDP:	Gross Domestic Product
GMDN:	Global Medical Device Nomenclature
GP:	General Practicioner(s)
GRAIL:	GALEN Representation and Integration Language
GUI:	Graphical User Interface
HCP:	Healthcare Provider(s)
HDTF:	OMG Health Domain Taskforce
HILS:	Health Information Locator Service
HIS:	Health Information System(s)
HL7:	Health Level 7
HMD:	Hierarchical Message Definition(s)
HMO:	Health Management Organization(s)
ICD:	International Classification of Diseases
ICT:	Information and Communications Technology
IEEE:	Institute of Electrical and Electronic Engineers, Inc.
IFOMIS:	The Institute for Formal Ontology and Medical Information Science
IHE:	Integrating the Healthcare Enterprise
INN:	International Nonproprietary Names
IOM:	Institute of Medicine
IS:	Information System(s)
ISO:	International Organization for Standardization
IT:	Information Technology
kV:	Kilo Volt
LES:	Lower Esophageal Sphincter
LIS:	Laboratory Information System(s)
LOINC:	Logical Observation Identifiers Names and Codes
LQS:	Lexicon Query Service
mA	Mili Ampere
MEDIX:	Medical Data Interchange Standard
MEDLINE:	MEDlars onLINE (a comprehensive literature database of life sciences and biomedical information)
MESH:	Medical Subject Headings
MIB:	Medical Information Bus
MRI:	Magnetic Resonance Imaging
MST:	Minimal Standard Terminology for Digestive Endoscopy
NLM:	The US National Library of Medicine
NLP:	Natural Language Processing
NLU:	Natural Language Understanding
NMIS:	Nuclear Medicine Information System(s)

OCL:	Object Constraint Language
OCL:	Object Constraint Language
OETS:	Order/Entry Tracking Service
OMED:	World Association of Endoscopic Gastroenterology
OMG:	Object Management Group
OO:	Object Oriented
OR:	Object Relational
ORCA:	Open Record for Care
OSHCA:	Open Source Healthcare Alliance
OSI:	Open Systems Interconnection
PBL:	Problem Based Learning
PET:	Positron Emission Tomography
PIDS:	Person Identification Service
PIS:	Pathology Information System(s)
PubMed:	A service of NLM that provides access to over 15 million MEDLINE citations
QVT:	Queries/Views/Transformations
RAD:	Resource Access Decision Service
RIM:	Reference Information Model(s)
RIS:	Radiology Information System(s)
RM:	Reference Model(s)
RMIM:	Restricted Message Information Model(s)
RM-ODP:	The ISO Reference Model for Open Distributed Processing
SDE:	Structured Data Entry
SM:	Service Model(s)
SNOMED:	Systematized Nomenclature of Medicine
SPECT:	Single Photon Emission Computed Tomography
SQL:	Structured Query Language
TC 215:	ISO Technical Committee 215
TC 251:	CEN Technical Committee 251
TOM:	Template Object Model
TQS:	Terminology Query Service
UI:	Unique Identifier (UMLS)
UMDNS:	Universal Medical Device Nomenclature System
UML:	Unified Modeling Language
UMLS:	Unified Medical Language System
WHO:	World Health Organization
XDS:	Cross-Enterprise Document Sharing
XML:	Extensible Markup Language

CHAPTER 1

INTRODUCTION

Medicine is one of the oldest professions in our civilization dealing with human life which deeply affects our individual and societal wellbeing. It should be expected that such an important field has to be backed up with every possible science and technology. Information and communications technology (ICT) is no exception with a good track of success in many fields such as finance and tourism. However it is obvious that this is not the case in healthcare in spite of the vast amount of literature on the proven benefits of ICT for tackling the fundamental problems of healthcare. The identified reasons for this paradox are many fold but they mainly focus on added difficulty to IS development due to the complexity and volatility of medical concepts. Another important reason roots from the inherent subjective or non-deterministic nature of medicine. Not only is the body of knowledge highly variable but also the practice changes from time to time and place to place. This means a particular instance of a medical situation may not be valid at all times. The consequences of this during HIS development are disastrous. Medical authorities concentrate on establishing a standardized or at least core medical curriculum in medical education and also try to realize a common medical terminology for practicing physicians to alleviate these problems. Considering the dynamic and fast track advance in ICT, we believe it will be more efficient to attack the former problem while hoping the latter one to progress over time. This is the general motivation for this study.

Gastrointestinal endoscopy, where the author previously had extensive experience, proved to be an appropriate domain for the study for a number of reasons. First a very comprehensive and high quality terminology (Minimal Standard Terminology for Digestive Endoscopy – MST) with an embedded structure for modeling was available. Second it is a manageable niche domain with well defined boundaries in terms of both medical and administrative processes. Third substantial amount of literature existed on the use of ICT in endoscopy including a large scale validation study of MST which we could build upon existing knowledge and benchmark our results.

A research prototype (GASTROS) was built based on MST terms and structure by using classical object-relational (OR) methodology. This was extremely useful for refining initial requirements and also for validating MST. When the prototype became mature enough to be used in data collection and installed at a large university hospital endoscopy unit, it helped us to identify points for user acceptance and improvement areas in both MST and the HIS. After three years of live clinical usage, a validation study on the Turkish translation of MST was performed (Atalağ, Bilgen, Gür and Boyacıoğlu, 2007b). The results turned out to be in accordance with the previous international study with considerable improvements over the first version of MST (Delvaux 2000). However from the ICT point of view we learned important lessons: First the need to add, delete or change domain concepts turned out to be much higher than expected which resulted in considerable development effort during maintenance. Second it became obvious that more domain knowledge was necessary for building HIS because MST does not explicitly state whether certain terms are mandatory or optional, or that they may repeat or occur only once. So in order to achieve terminology and HIS standardization more work was definitely needed. These formed the additional motivation for conducting the successive phase of the study.

After having validated MST and elicited requirements for the domain, we conducted a rigorous literature survey on current and emerging methodologies addressing the needs of our research. We have clearly observed that separation of domain knowledge from underlying software code by making explicit highly volatile domain knowledge versus stable operational or technical knowledge was a novel and promising approach for handling complex and rapidly changing requirements. We have selected openEHR Multilevel Modeling and Archetypes methodology which also guaranteed a high level of interoperability among HIS.

The specific problems being addressed in the study are:

1) Validity of the Turkish translation of MST content and structure for developing an HIS in digestive endoscopy,

2) Difficulties in maintenance of HIS due to changing domain knowledge in the MST reflected as changing requirements. Changes in database schema and recoding, testing and deployment were cumbersome,

3) Problems in maintaining consistency and validity of collected data due to changing database schema resulting from changing requirements,

4) Difficulties in consolidation and sharing of collected data due to non-conforming to a particular data model,

5) Difficulties to enable interoperability with other HIS.

This study mainly tackles the challenges of:

1) Information system modeling for better handling of complex and changing requirements,

2) Separation of domain knowledge from software code and database schema,

3) Identifying approaches for good user acceptance and wide usage of HIS,

4) Enabling structured and high quality data collection in medicine,

5) Sharing of collected data and establishing a high level of interoperability among HIS,

6) Building computationally usable and valid domain ontology.

1.1 Description of the Research Domain

While studying HIS and especially EHR modeling in the IS domain, we have conducted our research in the gastrointestinal endoscopy medical domain. Gastrointestinal endoscopy is a relatively new field in medicine which depends on visualization of the gastrointestinal tract for both diagnostic and therapeutic purposes.

By using an endoscope, upper gastrointestinal organs (esophagus, stomach and duodenum), lower gastrointestinal organs (colon and ileum) and in a special type of examination (ERCP) pancreas and biliary system can be assessed. While being a small, manageable and niche field in medicine, gastrointestinal endoscopy is a highly specialized, technology oriented and critical area. Because it is quite an invasive procedure and that important clinical decisions depend on it, results need to be reliable, complete and unambiguous. The gastrointestinal endoscopy community has started a terminology standardization initiative as early as 1984 which resulted in the comprehensive OMED (World Association of Endoscopic Gastroenterology) terminology. This has more recently been followed by the publication of the Minimal Standard Terminology for Digestive Endoscopy (MST). The second version of MST has already been translated into 11 languages including Turkish. Also recently it has been integrated with the National Library of Medicine's Unified Medical Language System (UMLS). MST contains a "minimal" list of terms and structure that could be utilized by HIS to fully record the results of an endoscopic examination.

The small size and manageability of gastrointestinal endoscopy domain with the availability of an appropriate domain terminology already translated into Turkish had strongly influenced our decision to conduct our research in this domain.

1.2 Objective and Goals of the Study

The objective of this study is to identify areas of improvement in problematic areas of HIS development and then find methods to alleviate them by first performing an in-depth analysis of the research domain and then determination of specific goals.

These specific goals are:

1) To develop a research prototype for the purpose of evaluation of MST content and structure and also for refining initial requirements for building an endoscopic HIS,

2) To provide, as much as possible, remedies for problems related with content and structure of MST that will be identified during development and usage of research prototype.

3) To study and contribute to the problem of maintainability at the level of modeling methodology. In particular, contribute to openEHR Archetypes and Multi-level

modeling formalism which is a promising methodological framework for alleviating this problem by explicit separation of domain knowledge from information.

4) To identify semantically equivalent clinical domain models with alternative representations and propose methods to manage them by exploiting openEHR Archetypes for achieving a high level of semantic interoperability.

5) To model MST completely by incorporating extensions made during the study and also extra knowledge gained by consulting to domain experts with the extended methodology.

1.3 Research Methodology and Validation Approach

The study consists of the following steps:

1. Identification of digestive endoscopy as the research domain,
2. Investigation of domain knowledge and terminology (MST),
3. Extensions to MST and initial modeling,
4. Developing a prototype application based on MST model,
5. Validation of the content and structure of MST,
6. Analysis of the problem domain, in particular with respect to maintainability and interoperability,
7. Identification of areas for improvement: separation of domain knowledge from software code and database schema,
8. Research for determining appropriate solution(s) and selecting openEHR Archetypes and Multi-level modeling,
9. Extending the modeling formalism,
10. Modeling of MST, using the extended methodology and validation.

After analyzing the research domain and investigating MST we have developed a prototype application (GASTROS) to validate MST content and structure. This validation has been performed by measuring the usage of MST terms versus free-text

and also considering the usage of the prototype application by endoscopists during the study. Since a previous validation study has been performed in the EU, we were able to compare the results of our validation study.

We have used rapid prototyping to gather and refine user requirements. The fast and effective user feedback into the development process is believed to result in a user friendly and widely accepted application. The design allowed customization of certain application parameters and the user interfaces, especially MST based structured data entry (SDE) forms which have let the users to perform their tasks faster and easier. These were demonstrated in the research prototype GASTROS and has been validated by the exclusive preference of the prototype by all endoscopists for reporting of all cases during this phase.

The research prototype proved to be invaluable for further refining initial requirements for building a digestive endoscopy HIS. We were able to identify a serious problem related with the hierarchy of MST and then discover a better suited hierarchy for building the successive versions of GASTROS and also base our modeling work thereafter.

By consulting to domain experts, both the terms and structure of MST was extended to correctly represent domain knowledge. It was clearly observed that the domain knowledge contained in MST was not sufficient to build a working system. For example it was not stated explicitly in MST whether a term is allowed to repeat or not. There was also no knowledge about whether certain items were mutually exclusive or conditionally dependent on others. Translation errors were corrected and also improvements became obvious for certain words.

As the maintenance of GASTROS turned out to be extremely difficult due to rapidly changes in MST, the need for a substantial improvement was evident. As the main solution we have identified proper modeling of domain knowledge a priori that would be able to explicitly separate knowledge from information and thus stabilize software code and database schema. Alternative modeling formalisms were evaluated and then openEHR Archetypes and Multi-Level Modeling methodology was selected.

The openEHR modeling methodology, which is designed for use in EHR systems, was extended to accommodate the needs of our modeling task. In particular Archetype Definition Language (ADL) has been extended and also a new type of Archetype was proposed which made possible to model MST Findings.

It was also noticed that the interoperability of the systems based on such models would be enhanced because all domain model artifacts map onto a standardized EHR architecture and use common reference models such as data structures and types. In order to enable sharing of the highly detailed and domain specific data collected by a gastrointestinal endoscopy HIS and also during integration with hospital IS or regional HIS where data models are not compatible, standard EHR architectures were identified as the common denominator and then we selected openEHR. This method not only enables means for interoperability but also proposed to decrease cost, and increase flexibility and maintainability of HIS. We strongly believe that the endoscopic record is an important part of the patient's lifetime EHR and extremely valuable for clinical research if it can be linked to other parts of the health record. So our assumption is to bring up the sharing of endoscopic records to a wider context and achieve this in parallel with the interoperability of EHR

As the modeling work proceeded and the needs from the endoscopy unit were collected, similar domain models but with different structure and/or terms were observed. It was evident that prescribing a single domain model would not be acceptable so a solution to handle this paradigm was sought. We have defined semantic equivalence of clinical domain models and also proposed methods to manage them. This would clearly aid in achieving a higher level of interoperability of information and systems.

The extended MST and extra knowledge about domain terms and relationships gained during the initial prototype phase were modeled by using the extended openEHR Archetypes and Multi-level modeling methodology. Parts of the MST were dissected and mapped onto the prescribed EHR architecture in openEHR EHR Reference Model. Finally the links to external terminology/ontology systems were established. The modeling work resulted in a novel domain ontology which may be used to develop a complete gastrointestinal endoscopy HIS.

The evaluation of the MST Archetype model by using the extended hierarchy indicated that all MST terms and relationships were successfully represented. Its syntax and ontology section containing all terms and external terminology links were also validated by the openEHR Archetype Workbench.

1.4 Organization of the Thesis

Chapter 2 defines key concepts used in this study and reviews the pertinent literature organized in different sections. The first section covers the information systems and software in general without specifying any application domain and presents the fundamental problems of this field. Second section is about healthcare problems and e-health where we dig into the essential problems of healthcare not necessarily always related with e-health. After setting forth very clearly these problems, we then review pertinent literature about the role and promises of HIS to alleviate them. It also provides a historical perspective of the use of ICT in healthcare over the years and then depicts current trends including EHR. The third section covers domain modeling and describes RM and Archetypes. The last section reviews the use of IS gastrointestinal endoscopy domain and MST.

Chapter 3 covers the part of the research related with the gastrointestinal endoscopy domain. After a detailed analysis of MST, our contributions related with MST are presented and thoroughly discussed. The case study consisting of the development of the research prototype GASTROS is also described in this chapter.

Chapter 4 describes the modeling paradigm used in this study, presents alternative modeling methods and discusses the rationale for selecting openEHR Multi-level modeling and Archetypes. Our contributions to the modeling methodology are discussed and also the resulting MST Archetype model is presented.

Chapter 5 concludes the thesis and outlines the contributions of the study. After auto-critique of the study, areas of future research are suggested.

CHAPTER 2

LITERATURE REVIEW

This chapter presents a review of the literature that constitutes the background to this study. It will start with general information systems issues and proceed to topics on information systems in healthcare. The rest of the chapter will review the details of health information systems, electronic health records and new approaches in health informatics discipline that try to tackle problems of IS development in healthcare. The focus will then be the multi-level modeling of HIS via Archetypes and its application to gastrointestinal endoscopy which forms the very core of this study.

2.1 INFORMATION SYSTEMS and SOFTWARE

Information systems (IS), by definition are integrated systems for providing information to support operations, processes, management analysis and decision-making functions within an organization (DeLone and McLean, 1992) as cited in Özkan (2006). An IS normally includes hardware, software, information, data, applications, communications, and people. It is evident that the success of an IS depends on all components. Software is a major, if not only, component of an IS and is the major focus of this study. In the rest of the chapter, software should be understood within the context of IS, especially health information systems (HIS).

Software development by itself is a formidable task. It is knowledge intensive and abstract; formed by construct of concepts, relationships, algorithms, data representation, events and functions. Therefore it is difficult to understand, visualize and describe. Another inherent feature of software is its complexity. Software is more complex than other human productions because each and every item is unique. Software also has many different states than other complex objects, such as computers, which makes it hard to handle (Brooks, 1987). In software development, it is nearly impossible to capture all requirements initially and then implement, because requirements can not be completely elicitated before building most of the system (Yeh, 1991).

Cost of software is high compared to hardware technologies which have been declining in spite of better performance over the years. This is in part due to the essential difficulties of software itself but also can be attributed to incorrect or changing requirements. Software costs often dominate system costs and maintenance is usually more costly. According to Sommerville (2000) and Yeh (1991) software maintenance clearly exceeds development costs (around 80% of software cost) and may range from two times to 100 times of development costs.

Successful software delivers the required functionality and performance to the users and should be maintainable, dependable and usable (Sommerville, 2000). It is based on the realities of environment and the tasks of its users. So the development process is not only technical but also social to acquire the knowledge to realize a working system. In a way software can be thought of a storage medium for knowledge (Armour, 2000). The capture of knowledge is mainly done during requirements elicitation and analysis in software development process. It is very important as this step mainly determines what software to build and failure in this phase has been number one reason for most failed projects (Liebowitz, 1999; Sommerville, 2000 and Brooks, 1987). Serious difficulties exist for handling requirements. First all requirements can not be known in advance (Armour, 2000; Liebowitz, 1999; Brooks, 1987; Yeh, 1991). Second elicitated requirements may not turn out valid for the purpose of the software after deployment (Johan, Hoorn, Konijn, van Vliet and van der Veer, 2007). Sommerville (2000) reports that correcting requirements errors after deployment may cost up to 100 times more than fixing an implementation error. However the third difficulty is the most problematic for maintaining software:

Requirements later might change or new requirements emerge (Liebowitz, 1999; Sommerville, 2000; Brooks, 1987; Yeh, 1991; Johan et al., 2007).

In order to tackle these difficulties, several approaches had been developed: appropriate software process models to better fit requirements handling (spiral and evolutionary model), new paradigms such as reusable component based development and rapid prototyping, utilization of conceptual modeling notation and tools (UML, CASE tools) are examples. However in spite of all these efforts, new paradigms are still needed to ease problems in software development (Yeh, 1991; Boehm, 1988; Fowler, 2004).

2.2 Healthcare problems and e-Health

2.2.1 General Problems of Healthcare

Cost, Quality, Safety, Accessibility/Equity, Effectiveness

According to the World Health Organization (WHO) (2006) and other pertinent literature, the major problems of healthcare system globally in the 21st century have to do with the quality, safety, effectiveness, cost and accessibility/equity (Chaudhry et al., 2006; Stolberg, 2004; Tang, 2003; Godlee, Pakenham-Walsh, Ncayiyana, Cohen and Packer, 2004; IOM, 2001; President's Advisory Commission of Consumer Protection and Quality in the Healthcare Industry, 1998; Garson, 2004).

Regardless of economic and social status, all citizens are seeking for safe, high quality and effective healthcare services at a reasonable cost. In reality this is not satisfied and the reason is two fold: in the first place the resources are scarce as in other sectors such as the limited number of doctors or funding. Secondly, the high level of specialization and division of labor due to the size and complexity of healthcare coupled with the necessity to keep up with rapid advances in medical knowledge and technology necessitate effective management and coordination. Thus communications play a key role to deliver high quality services. However this is not the reality today because the system is poorly organized, managed and coordinated which seriously degrade the quality of healthcare. The efficiency of services is low, resources are not rationally utilized, and patient outcomes are not satisfactory because processes are not based on best-practices and standard pathways. The public health is also degraded because continuity of care can not be achieved. Other problems include

ineffectiveness, under-utilization (not receive sufficient care), overuse (receive unnecessary services) and highly variable provision of healthcare services (changes from one place to another is clear indication that not all services rely on best scientific knowledge) (Stolberg, 2004; President's Advisory Commission of Consumer Protection and Quality in the Healthcare Industry, 1998; IOM, 2001).

The safety of healthcare system is not even close to what it has to be. In U.S. only, it has been reported by the official governmental agency Institute of Medicine (IOM, 2000), that nearly 100,000 people die due to preventable medical errors each year. Those errors broadly happen due to failure of a planned healthcare action to be completed as intended or wrong planning.

Cost of healthcare is on the rise globally. In developed countries nearly 10% of GDP is spent for healthcare. In U.S. this figure is as much as 15% as of circa 2005 and constantly increasing (Charette, 2006). Although major part of this high cost is attributable to essential problems unlikely to be improved by known methods, rest is due to sick planning, bad management, lack of coordination, irrational spending of resources, and ineffective processes. The challenge is to decrease costs while increasing quality, safety and effectiveness (WHO, 2006; Chaudhry et al., 2006; Godlee et al., 2004; Menachemi and Brooks, 2006).

Today, even in developed countries, accessibility to healthcare is problematic. The reasons are related with geographic location, uneven distribution of facilities and resources, lack of health insurance, linguistic and cultural differences. Equity, inclusiveness and common access to healthcare services irrespective of culture, education, language, geographical location, physical and mental ability, age and gender, are major challenges (WHO, 2006).

Effectiveness of healthcare can be increased by providing preventive actions and medical services based on best scientific evidence and practical value, while making no significant trade-offs, meaning the benefits of services considerably outweigh the possible risks. Medical knowledge has been estimated to double every six years which poses a big challenge for healthcare workers to stay up-to-date and deliver effective services (Grimson, 2001). High quality and large volumes of data spanning a long period of time are needed to calculate health outcomes of healthcare services and products such as drugs. Since this information is extremely difficult to acquire, widespread and long-term effects of drugs and treatments or preventive measures can

not be assessed. This seriously degrades safety and effectiveness while also inhibiting the change of medical knowledge and proliferation of best-practices. Another serious danger happening today is the potential loss of trust and satisfaction for healthcare which results in seeking other solutions by patients and families, such as alternative medicine with little or no scientific evidence (IOM, 2000).

Science of Medicine vs. Art of Medicine

There are also problems with the dual nature of medicine in which one part is scientific and the other part is informal – that is called the "Art of Medicine". The latter causes subjective representation and dissemination of clinical knowledge and is the source of variety in healthcare practice (Malterud, 2001; Nelson, 1998).

Medical knowledge is produced either as a result of scientific studies or during clinical practice. It is unfortunately very difficult, if not impossible, to capture the latter with known methods. Years of aggregated information about a certain topic leads expert physicians to discover new medical knowledge, which is hard to describe which eventually becomes 'insight'. This is the 'Art of Medicine' (van Bemmel and Musen, 1997; Malterud, 2001; Nelson, 1998).

Unfortunately, with current paper-based documentation in healthcare, either the data are recorded in a way which makes it impossible to extract any useful knowledge from it or not documented at all. Medical record is an essential part of healthcare. It holds key information about the health status of the patient and also is a medium of communication among clinicians when delivering services. Apart from healthcare related uses, the information contained in the record is also utilized for research, education and planning purposes. However serious shortcomings of paper-based record keeping do exist: a patient file can only be at one place at a time, often may not be available or even lost. Use of free text may lead to unstructured content, illegibility, incompleteness and ambiguity (van Bemmel and Musen, 1997). In addition transcription is needed, which may lead to errors, to perform analysis for in research or planning. These problems with paper-based records ultimately degrade the safety and quality of services (IOM, 1997, 2000, 2001; Chaudry et al., 2006).

2.2.2 Trends in Healthcare

There had been important movements in healthcare caused in part by HIS or influenced HIS. All are inclined towards alleviating the fundamental problems of healthcare. In order to provide state-of-the-art healthcare services to a patient, a team of health professionals has to act in a coordinated manner. In some cases, patient's family physician or another healthcare provider might have to intervene with the process. It can be said that medicine is moving from a single provider/physician to a team of physicians from multiple providers. This is called "**Shared Care**" (CEN EN 13940-1, 2006; Haux, 2006). It is obvious that as number of parties and locations increase, the need for effective and high quality HIS is of utmost importance.

People are still in the scope of healthcare even if they are not ill. The ultimate goal of medicine is to keep people as healthy as possible by first preventing illnesses and also providing necessary treatment during times of illness. Current healthcare systems globally are more focused on the latter strategy which is more costly, ineffective and at times not ethical when better alternatives exist. In fact, preventive measures are cheaper, safer, more effective and definitely more ethical but they require long-term follow-up of people, consistent record keeping, good communications infrastructure and surveillance of health threats (environmental, biologic or chemical). So keeping people healthy requires both preventive and treatment aspects of healthcare in harmony for a long time over a number of providers. This concept is called the "**Continuity of Care**" and enables the longitudinal provision of healthcare services (CEN EN 13940-1, 2006). Continuity of care depends on the effective exchange of data and information about the clinical situation, context and services provided to an individual, between different providers involved in the process, within the framework of ethical, professional and legal, rules as defined in CEN EN 13940-1 (2006) and Continuity of Care Record (CCR) (2007). Unlike a healthcare team providing services to an individual cooperatively in shared care, in continuity of care the services provided by different providers are independent and not synchronous. The important aspect is the time-related links between those different health care services.

"Seamless Care" is a quality issue, which focuses on the timely and appropriate transfer of activity and information, when responsibility for the delivery of health care services is wholly or partly transferred from a health care provider to another (CEN EN 13940-1, 2006).

When all three concepts are present at the same time (continuity of care, shared care, and seamless care) this is called "**Integrated Care**" (CEN EN 13940-1, 2006).

Medical Education: Problem Based Learning, e-Learning, Online Resources

There is no doubt that all aspects of medical services are dependent on the quality and effectiveness of medical education. Recently a new paradigm called **Problem Based Learning (PBL)** has come into play where the courses are not just focused on medical subjects but they are based on real-world healthcare problems. PBL is an active, adult-oriented, problem-centered, student-centered, collaborative, integrated and interdisciplinary system (Camp, 1996). Teaching is conducted mainly in small groups with active discussions and frequent assignments within a particular clinical context. Since courses are highly variable, it difficult to prepare courses and assess afterwards. ICT is believed to have a large potential to enhance its quality and effectiveness.

E-learning is the facilitated and supported education by the use of ICT and aims at provision of the best and most appropriate ways of effective learning (Klein and Ware, 2003). It now offers new vistas in all aspects of medical education especially for continuing medical education (CME) (Harden, 2005).

Most of the medical knowledge that was contained only in textbooks is now accessible on the Internet in various formats. Content management is easy and rapid with electronic means of publishing. This also brings along new possibilities that were not possible with conventional methods. It has been shown that the use of online resources significantly improved the quality of answers to typical clinical problems and that they are considered as effective tools for helping decision making tasks for clinicians (Westbrook, Coiera and Gosling, 2005).

Evidence Based Medicine, Clinical Guidelines, and Care Pathways

Provision of healthcare services based on the integration of individual clinical expertise with the best available external medical knowledge from systematic research that has proven evidence for safety, effectiveness and quality is called **Evidence Based Medicine** (Sackett, Rosenberg, Gray and Richardson, 1996). It necessitates first to identify clinical situation and context, and then proper access to relevant

knowledge. Because it takes quite a long time for new knowledge to appear in classical textbooks or scientific journals, the use of electronic EBM (knowledge) repositories are preferred.

For more structured and well-defined medical problems, such as diabetes or cardiovascular disease, panel of experts from that domain design **Clinical Practice Guidelines** to be used by other colleagues safely to deliver high quality and effective services. They can be either in paper or in electronic form ready for processing in HIS to provide decision support. de Clercq, Blom, Korsten and Hasman (2004) and Grimshaw and Russel (1993, 1994a, 1994b) have shown the benefits of using clinical guidelines in practice by reducing the variability of practice and decreasing costs, while improving patient care.

Clinical/care pathways are methodologies for effective decision making and organization of services for patients falling into well-defined group of issues for a defined time period. They may be distributed as paper or in numerous electronic formats including some computer-processable formats. The aim of a clinical pathway is to enhance the quality of services by improving patient outcomes and safety, increasing patient satisfaction and enabling rational use of resources (The European Pathway Association, 2007).

Accreditation and Quality Assessment in Healthcare

In order to assess the quality, effectiveness and safety of healthcare services performed by different institutions and improve them, accreditation, licensing and certification processes started at the beginning of 1990s (The Joint Comission, 2007). An organization seeking to get accreditation or quality certification or licensing needs to effectively measure its outcomes, a set of quality measures and performance criteria. It is overtly difficult, if not impossible, to perform these tasks without the use of HIS.

2.2.3 Benefits and Promises of ICT in Healthcare

Uses of ICT in healthcare or e-Health solutions primarily aim to alleviate healthcare problems for better health. Although there have been unsuccessful implementations in

the past, the evidence shows that health ICT systems help to establish a better healthcare (IOM, 1997, 2001; President's Advisory Commission of Consumer Protection and Quality in the Healthcare Industry, 1998; WHO, 2006). IOM (1997) underlines the fact that it is difficult to assess the benefits of healthcare ICT as they are mostly indirect (non-quantifiable) benefits rather than direct (quantifiable) which make it difficult to convert into monetary benefits.

It has been reported that use of ICT greatly enhances the quality and efficiency in healthcare. This mainly results from the ability of providing guideline based services especially preventive measures, enhanced monitoring of health status and surveillance of outcomes, reduction of medication errors and lowering redundancy or irrational use of resources and services (Chaudhry et al., 2006). Organizations working for improvement of healthcare like IOM (2007) and Leapfrog Group for Patient Safety (2007) are strongly endorsing use of ICT in healthcare as a key factor for improvement.

By offering possibility of redesign of services and streamlining healthcare processes, ICT enables faster transactions which result in direct monetary gains in terms of faster billing and reimbursement in healthcare (Moorman and Bernstein, 1999; van Bemmel and Musen, 1997). Quality and safety of care is also enhanced because of decreased waiting times, better planning and automatic monitoring of certain steps such as medication orders. Fast, reliable and timely provision of health information has been demonstrated not only to improve patient care directly but also result in better management and planning activities (Godlee et al., 2004).

By the use of computerized order entry systems, errors can be significantly reduced while at the same they help to decrease costs, shorten length of hospital stays and improve compliance with treatment guidelines. They enable standardization of practice and thus decrease variability of healthcare services by adhering to clinical guidelines and pathways. Decision support is also seamlessly introduced into daily practice which improves efficiency and appropriateness of medical decisions. By means of using electronic communication tools, timeliness and quality of information is enhanced which result in better cooperation of healthcare workers and managers (Kuperman and Gibson, 2003; Sittig and Stead, 1994).

Other key areas for improvement resulting from the use of ICT in healthcare are related with supporting of scientific research and education. Transforming all data,

information and knowledge into digital form in good quality serves as an invaluable tool for medical research and education (Tang, 2003). Enhanced ability to capture and record clinical information also leads to discovery and sharing of medical knowledge which leaves a smaller gray area from the Art of Medicine. eLearning in healthcare in context of knowledge dissemination is a useful tool for education and training of healthcare workers. When effectively used, it can improve the quality of education, increase accessibility to geographically isolated people or those who have poor local learning facilities, and bring about new innovative forms of learning (WHO, 2006).

2.2.4 Types and Evolution of Healthcare ICT

Computers in Healthcare, Healthcare IT/ICT, HIS

Computers have long been used in healthcare for various purposes (van Bemmel and Musen, 1997; Haux, 2006). By the rapid advance of IT in recent years, they became somewhat more functional and usable with innovative GUIs and alternative human-computer interfaces. The increasing effect of communications technology and its seamless integration with IT resulted in the birth of ICT which is very close to the definition of IS (Wikipedia definition: ICT, 2007). Today healthcare ICT and HIS are used interchangeably for systems consisting of healthcare actors, procedures/processes, software and hardware to perform tasks necessary for providing or supporting healthcare services and products.

Hospital Automation/Information System, CIS, LIS, RIS, PIS……

Before the advent of HIS concept which embraces all types of IS in healthcare, there were many concepts somewhat confusing. Hospital automation is one example; in fact it is truly an IS. Now it is called as a hospital wide HIS (Haux, 2006). A complete and integrated Hospital IS consists of Administrative and Medical parts. The administrative part contains functions such as stock and materials management, procurement and so on. The medical part is further classified as CIS or departmental IS and Clinical Support Systems such as LIS, RIS, PIS, NMIS (van Bemmel and Musen, 1997).

Computational Systems and Embedded Software in Healthcare

There are other uses of healthcare ICT, mainly as computational systems (CT, MRI, DSA, laboratory auto-analyzers and so on) or as embedded systems in pacemakers, interpreted ECG and in new modalities such as capsule endoscopy. They have increasing level of data exchange with HIS by emergence of medical device data communication and representation standards (van Bemmel and Musen, 1997).

CPR, EMR/EPR, EHCR/EHR

As the institution centric HIS started to collect structured clinical data from patients and that this data need to be shared with other providers and organizations, a different approach was needed. The patient-centered computer-based patient record (CPR) came into play. It was merely an electronic form of classical paper-based patient record that collected data within a specific provider (IOM, 1997; Charette, 2006). The next concept was the electronic medical record (EMR) that had extended functionalities such as the ability to be shared by other providers and possibly used by patients themselves (Hammond, 2003).

Currently, by taking into consideration the importance of preventive medicine, trends like continuity of care and shared care brought about the need for electronic health records (EHR). It supports the common understanding of record keeping for individuals not only at times of illnesses but also during healthy periods. It is truly patient-centric which records all relevant care events from different sources, supports integrated care (accessible and writable by all parties in a secure zone), longitudinal (covers time-based history of health events and indelible for medico-legal purposes (openEHR Foundation, 2007; ISO TR 20514:2005 Health Informatics - Electronic Health Record Definition, Scope and Context Standard, 2005).

Primary, Secondary and Tertiary Care HIS

HIS can also be classified according to the healthcare levels that it serves for. For example a HIS for a specialized tertiary care or a university hospital can be called as a Tertiary HIS. Similarly secondary and primary HIS also exist which are all enterprise

based HIS. No matter what level they operate, main purpose is to support administrative and medical aspects of healthcare services for the better (van Bemmel and Musen, 1997).

GP Systems, Public HIS

GP (Information) Systems are in fact a member of primary HIS designed for GP only. Considerable part of it consists of EHR because a GP serves a group of individuals within a local territory over a long time – possibly a lifetime. Other functionalities include administrative tasks such as referral and medical tasks like laboratory orders and results checking. Physician office-laboratory links or practice management software resembles GP systems in functionality but serves for other physicians and specialists usually working in common practice settings together.

By the increased penetrance of ICT infrastructure and culture into the society, a different type of HIS is emerging: Public HIS. It is a large scale health information system at the citizen level which serves as the communication system between the citizens and the health professionals. The main difference from other HIS is that it aims to align two sliding reference frames – the citizen and providers. The citizen controls the access of responsible health team to parts of the electronic health record and is able to federate service providers. The main component is the EHR but it has other functionalities like decision support (personal communications with Philippe Ameline; O'Carroll, 2002).

Billing/Reimbursement Systems and other Administrative Systems

Healthcare services are usually provided according to the fee-for-service or package based model such as DRG. Independent of the payment model, the billing system requires a mix of administrative and medical information to generate invoices. Because of the high cost of medical services and rules for payment from health insurance or social security organizations, it is important first to make an eligibility check to see if patient is covered and then get provision for services before they are carried out. All these actions are carried out by the non-medical part of the HIS (van Bemmel and Musen, 1997).

Health statistics & Epidemiology

High quality data is a must in order to produce reliable health statistics. It serves for both interpreting underlying data and production of certain quality measures and performance criteria. Epidemiology studies to identify patterns, causes, and control of diseases in groups of people may easily be conducted with large datasets and in shorter time by the use of HIS (Haux, 2006; Dorr, et al., 2007).

2.2.5 Other key concepts

Coding & Classification Systems, Terminologies and Ontologies

Coding and classifying of medical records is a very old phenomenon and roots from the need to analyze vast amount of medical data by structuring free text for administrative (i.e. billing) and medical purposes (Coiera, 2003). World Health Organization (WHO) International Classification of Diseases (ICD) (2007) is an important disease classification system to collect and analyze morbidity and mortality data from different areas around the World and conduct health statistics and epidemiologic studies.

Standardized medical terminologies and controlled vocabularies, such as SNOMED and MEDCIN, are important for encoding all aspects of the health record (Systematized Nomenclature for Medicine (SNOMED), 2007; MEDCIN®, 2007). Semantic relationships might be also be defined in these terminologies. Smaller terminologies, such as addressing only a specialty or even a health issue, also exist. The Minimal Standard Terminology for Digestive Endoscopy (MST) is an example (Crespi, Delvaux, Schapiro, Venable and Zwiebel, 1996; Delvaux, 2000b). Terminologies not only provide a common language and understanding among health professionals but also enable records to be processed by computers. It is also possible to link to other knowledge sources by mapping terminologies onto each other.

Medical ontologies further define and represent concepts existing in a particular domain, their attributes and the relationships between them. In a sense they reorganize and represent medical terminologies which are optimized for human processing that contain significant amount of implicit knowledge (The Institute for Formal Ontology

and Medical Information Science (IFOMIS), 2007). HIS need explicit knowledge to be able to capture medical concepts and processes them unambiguously. Ontologies have better chance to enhance or knowledge-enable HIS because they establish a common understanding of the structure of information, enable reuse of domain knowledge, separate domain knowledge from the operational knowledge, provide basis for interoperability and provide methods to analyze domain knowledge (Natalya and McGuiness, 2002; Beale and Heard, 2007; Fernández-Breis et al., 2006; The Institute for Formal Ontology and Medical Information Science (IFOMIS), 2007). While usually medical ontologies focus into a particular subject, at times they can be very broad such as the National Library of Medicine's Unified Medical Language System (UMLS), (2007) that includes and maps numerous other knowledge sources some of which are ontologies themselves.

Structured Data Entry (SDE)

High quality data collection is still a big challenge in health informatics (van Bemmel and Musen, 1997; Moorman, 1995a). The added benefit of switching from paper-based records into electronic records without structured data is not even comparable to the chasm between collecting free text and structured text. Unfortunately medical records are still composed of free text and little, if any, structured data collected via forms. SDE is the electronic counterpart of such forms and aims to increase completeness and reduce ambiguity by collecting data according to previously defined structured and coded data layout (Moorman, 1995a). Structured data are needed in order to process and transform raw data into information by computers. Also it has been shown that SDE is superior free text data entry for high quality and rapid data collection (Los, van Ginneken, de Wildeand van der Lei 2004; Los, van Ginneken, and van der Lei, 2005).

The ability to select from a set of enumerated values in pick lists and validity checks during data entry helps to eliminate errors. Moorman, Ginneken, van der Lei and van Bemmel (1994a) identify a number of key requirements to be met for building usable SDE applications: first it has to provide maximum possible expressive power and flexibility to the physicians, second it must follow routine clinical thinking, third the presentation of data has to be consistent and fourth the time needed for filling out SDE should not be greater than free text entry. The data collected by SDE should also be

complete and unambiguous which have sufficient contextual information. These requirements for successfully realizing SDE over long time necessitate a few key steps to be taken: using terminology that users accept, good separation of information and knowledge layers and effective means of IS development (Moorman, et al., 1994a). These are the core motivations of this study for tackling the problem of structured data collection in medicine and the reason why openSDE was a strong candidate for the modeling work in the study.

NLP/NLU, Speech Recognition

Extracting coded data from free text by various computational methods is called natural language processing (NLP) or understanding (NLU) (van Bemmel and Musen, 1997; Moorman, 1995a). However Moorman (1995a) recommends that the technique has to improve considerably before being used in daily practice by referring to Baud, Rassinoux and Scherrer (1992). In addition all the ambiguity, incompleteness and errors in original documents will be transferred to the structured form because NLP/NLU does not alter data capturing process.

2.2.6 Key enabling technology for next generation medicine: EHR

Definition and History

The ISO TC215 Health Informatics Standard (2005) definition of EHR is:

> "A longitudinal collection of personal health information concerning a single individual, entered or accepted by health care providers, and stored electronically. The information is organized primarily to support continuing, efficient and quality health care and is stored and transmitted securely. The EHR contains information which is:
>
> 1.retrospective: an historical view of health status and interventions;
>
> 2.concurrent: a “now” view of health status and active interventions; and
>
> 3.prospective: a future view of planned health activities and interventions."

So EHR is used primarily to support individual patient care and also has many secondary uses in medico-legal events by providing reliable evidence data, quality

studies, education and research. By also providing access to quality health information, public health services are improved. Policy development, health service management and financial services also benefit from EHR (ISO, 2007).

Hammond (2002) recommends that data in EHR should have the features given in Table 2.1.

Table 2.1 Recommendations for EHR Data (Hammond, 2002)

• Patient-centered • Comprehensive • Aggregated • Organized • High data integrity	• Timely • Structured, semantically understandable • Sharable • Accountable • Secure and private

EHR has become more important because the healthcare today is truly a multi-contact system, where patients receive services from different healthcare providers (HCP) and important data are generated at each location. Patient centered healthcare records have proven to be a superior approach than organization-centered records (Beale, 2001). Another driver is the added complexity of managing highly mobile patients and healthcare professionals. As a result the patient information is all over the place in various formats. The consequences of this fragmentation are repeated entries of same information, repeated tests, and errors due to not being able to access key information, and evidently high cost and low quality of healthcare (Tang, 2003; IOM, 1997; Menachemi and Brooks, 2006; Grimson, 2001; openEHR Foundation, 2007).

Most governments have identified EHR as the major aid in controlling cost of healthcare while increasing quality, effectiveness and safety. U.S., UK and Australia have started national EHR projects (Charette, 2006). Within the ambitious common ICT objectives of the EU in i2010 (2007), a European Information Society for growth and employment, there is a prominent e-Health Action Plan (2007) And this action plan states clearly that in order to achieve a European e-Health structure, the road passes through electronic health records and that it calls each member state to establish regional and national frameworks.

Progress in adopting EHR systems has been slow. This little progress has mostly been in domains of standards development, open source developments and building of vast amounts of knowledge sources to be used in EHR systems. EHR standards now provide coherent definitions and technical methods hence are precursors for interoperability. Open source collaboration and development has broken the commercial single vendor type closed development. In terms of building knowledge sources, not only the content has increased and refined in computable medical terminology, ontology and guidelines but also new methods for medical knowledge representation like Archetypes had been discovered.

As depicted by Beale (2005) many open source implementations now exist in addition to numerous commercial EHR products. The most prominent one is U.S. Veteran's Health Administration's VistA system. Other projects with operational implementations include openEMed, TORCH, gnumed, and a number of EU-funded projects such as PICNIC and HARP.

Benefits of EHR

It is expected that EHR systems will reduce the cost and improve the quality of care by providing better-informed health care providers and patients, the elimination of duplicate tests, and better coordination of treatment by more than one health care provider (IOM, 1997). Some of these benefits have been tested and validated in studies conducted by Tang (2003), Bird, Goodchild and Tun (2003) and Beale (2005).

Van Ginneken (2002) identifies and explains the key benefits of EHR as: accessibility**,** readability, easy reporting, completeness of data, decision support, supporting preventive medicine, access to external knowledge sources and data analysis. Possibly the most striking effect of widespread usage of EHR will be on biomedical research. Difficulties in identifying, finding, interpreting and applying knowledge from clinical trials from published literature is well known. This results in inefficient transfer of evidence from research into clinical medicine. As a result the level of healthcare given to patients today is way behind what is known by biomedical science. Clinical guidelines and protocols are being used as effective and rapid methods to disseminate this knowledge. Linking of such knowledge sources with the EHR will provide not only proper and timely distribution but also ensure their proper use during care (Grimson, 2001). This aspect will constitute a major contribution of this study.

EHR also has implications in all aspects of security for patient records. Current legislations for conducting healthcare have strict rules especially for privacy and confidentiality of medical information. However with paper-based records, it is physically difficult to apply those rules because people in the wards where file cabinets are (usually anyone with a white coat) can access patient files. However EHR systems can provide effective and reliable security mechanisms which allow a very fine grained access to the health data (Tang, 2003).

One last but not least anticipated benefit of EHR in the context of bio-medical research is the integration with the genomic world. Over the years, we have witnessed great advances in genetics and medical sciences but the merging of these interrelated domains has been so far not satisfactory. Lack of common medical terminology and variation in practice might have played a role in this. Genomics had achieved to form a common terminology and established electronic means of data representation and sharing decades ago. EHR might be the key enabling paradigm for linking of clinical medicine to the genomic world. The merging of these two domains will enable access to relevant and reliable information for conducting integrated clinical, genetic and environmental research which will eventually lead to more effective and curative treatments for better care (Grimson, 2001).

EHR Requirements

There had been international efforts to determine the functional, technical and content related requirements for EHR. According to ISO TR 20514 (2005) and Beale (2005) some refined and key functional requirements are as follows:

- Have information and efficient user interfaces reflecting multiple levels of hierarchical biological and social organization,
- Support mobile patients,
- Keep information for a lifetime (longevity of around 100 years)
- Support multilinguality,
- Enable sharing and authoring of data by multiple users simultaneously,
- Integration with knowledge bases such as terminology, ontology and clinical guidelines,

- Support wide geographical availability of records to multiple providers and applications,
- Enable consent-based, potentially fine grained privacy rules on information use (with exceptions for emergency access),
- Withstand multiple sources of constantly changing requirements including medical technology, clinical procedures and guidelines, genomic/proteomic biomedicine,
- Provide reliable medico-legal support for all users.

Another key functionality for an EHR system, which was not possible technically with paper-based records, is the ability to provide multiple views of data. As Hammond (2002) mentions, an organization/provider view must exist to serve the need of the institution in patient care, service management, workflow management, and billing. This view provides the source of data for other views, since it is in this setting that the patient/provider encounter takes place. A second view is a composite view that represents a complete summary view (the patient-centric view) of the person's health. This view also serves needs of health and bioterrorism surveillance and epidemiology. Third view is a personal health view that is customized to each individual and their health needs.

Technical requirements of EHR are outside the scope of this study hence will not be mentioned here. The requirements on the content of EHR are within the scope of major health informatics standardization bodies: CEN TC251, HL7, ASTM E.31 and openEHR (Beale, 2001 and 2003). Like the separation of data from structure in XML, the clear separation of medical information from knowledge is truly a paradigm shift. Here only the fundamental content model is mandated by standards such as the organization of a patient dossier or use of certain demographic entities. Rest of the content is dynamically modeled by domain experts using common information components (Beale, 2000 and 2002). More detail shall be given in the section on health informatics standards and will materialize in the following section where archetype formalism and multilevel modeling strategy for modeling a medical domain is demonstrated.

EHR Architecture

Different strategies exist for deploying shareable EHR systems. According to Shabo, Vortman and Robson (2001), there are two approaches. First approach is provider-centered, creating a "Virtual EHR" without physically establishing a distinct EHR system of its own. It is also called as a "Federated EHR" in which a logical view or physical assembly of partial EHR information (EHR extracts) happens "on the fly" from distributed EHR systems (Kalra, 2003). The federated approach seems to be appealing but has many implementation and performance problems in practice when many records from many different federated EHR systems are involved. In the second approach which is more consumer-centered, a consolidated EHR repository is established and all pertinent patient data is recorded from multiple providers at the time of healthcare event and served when the record is needed at any point of care. It has important advantages over the federated EHR by a simpler access control and security which promises a much better price/performance ratio (ISO TR 20514, 2005).

The approach to realize EHR by integrating disparate organization-centered HIS has proven inefficient and not feasible in many studies. Without agreeing upon common record structures and semantics, in order to be able to exchange information between systems, a different interface has to be implemented between every single system pair (Bird, Goodchild and Tun, 2003).

From an architecture point of view, two alternative approaches also exist to realize EHR systems. In the first approach a global agreed upon architecture is planned to be used by all applications hence provide a high level of interoperability such as openEHR and CEN EHR standards. The second approach is based on messaging paradigm, extracts of EHR, between disparate HIS such as HL7. However a consensus is emerging that realization of EHR by messaging paradigm is not a viable solution and that agreeing upon common models for content and structure are required (Grimson, 2001; Beale and Heard, 2006). The strategy taken in this study to model a medical domain, gastrointestinal endoscopy, can be considered as a bottom-up approach from an architectural point of view for realizing interoperable and maintainable EHR. Standardized terminology, reference models and archetypes form the building blocks of a "mini" EHR and we will demonstrate the feasibility of the former approach.

There are discussions as to who shall host this central service. From patients' privacy point of view, neither government nor providers are suitable. Hosting this EHR service by HMO is also problematic because the possibility to analyze and use data in determining health insurance plans. Possibly an appropriate way to host it will be through an independent consortium based organization (Shabo, Vortman and Robson, 2001).

Problems & Challenges in Realizing EHR

The EHR has to accommodate individuals' biologic, sociologic and psychological complexities within its structure (Grimson, 2001). The data in EHR come from a number of different sources such as nurse's notes, progress notes, treatment plans, medications, laboratory results and imaging reports in countless different terminology, layout and formats. This necessitates that the EHR be very flexible. Not only there are difficulties with data capture and representation, but care provision requires access to and utilization of best available knowledge. So the EHR not only has to be able to add or modify knowledge but also make it usable by incorporating into its functionality and workflows. The study by Bird, Goodchild and Tun (2003) seeks to handle this complexity by adding an additional layer of modeling for domain knowledge, on top of information modeling.

The success of EHR projects does not seem encouraging at all; according to van Ginneken (2002), 9 out of 255 projects had failed. It is believed that this caused by unjustified balance between effort and benefit of EHR systems. It has also been observed that the clinicians play an important role in the content, quality, and usability of EHR systems. The main bottleneck is the reluctance of physicians for data entry because of anticipated time-loss while typing, limited coverage and concerns with physician-patient encounter. This low success rate has also been attributed to non-disciplined development process of EHR, so it is recommended to employ rigid software engineering principles. One critical success factor is properly designing business processes and informing people before introduction of EHR. It should also be noted that HIS and specifically EHR systems are usually built from scratch rather than using tested and proven software components. This not only makes design and implementation hard but also causes maintenance to be even harder (Grimson, 2001).

The modeling approach taken in this study uses standardized reference models to assemble logical medical concept constructs that can be shared and reused.

Cost of EHR systems is a limiting factor for their realization. According to van Ginneken (2005), 7.5-13.5% of an institution's budget has to be allocated for EHR. In a more recent publication, for nationwide EHR projects, the figures are frightening: in Australia while estimated cost was 500 million Australian Dollars, it has risen to 2 billion Australian Dollars. In UK the implementation costs in 2002 was 2.6 billion pounds and as of 2006 it is estimated to be at least 15 billion pounds. In U.S., since EHR projects are not publicly funded, the estimated cost is between 100-150 billion Dollars with 50 billion Dollars per year for operations (Charette, 2006). Controlling this high cost is a major motivation in this study and will be discussed in detail in Chapter 4.

The need for many different ICT systems to interact with each other is a big challenge. This brings about hard problems of interoperability from a sense starting from physical levels and reaching up to data, information and knowledge levels. Often the requirements for EHR conflict with each other which make realization of EHR problematic. For example the requirements of privacy of medical data and accessibility among different HCP across national borders by using different vendor solutions are difficult to achieve at the same time (Grimson, 2001). The use of a standardized multilingual terminology (MST), RM and Archetypes used in this study is proposed to be the enabling paradigm for interoperability and we will try to demonstrate this in gastrointestinal endoscopy domain.

The current situation with HIS and EHR is still organization-centered and episodic – meaning they contain information from a patient's visit for a health issue rather than a full history of the patient. However Grimson (2001) depicts that the maxim for EHR is community-based and patient-centered used in shared-care.

The experience with many different approaches to model and implement EHR has proven to be problematic. As the nature handles complexity by dividing into sub-components and layers, in development of EHRs, separation of concerns and also separation of domain knowledge from underlying information and data is a new promising approach and a brand new challenge (openEHR Foundation, 2007). The problems mentioned in this section constitute major part of the research problems which is discussed in detail in next chapters.

2.2.7 Standards in Health Informatics

Watching a video produced in U.S. converted to DivX format by local DVD store played on a player manufactured in China which is plugged in mains where electricity is imported from neighboring country requires seamless standards at work. According to ISO/IEC Guide 2 (1996), formal definition of standard is a document, established by consensus and approved by a recognized body that provides, for common and repeated use, rules, guidelines or characteristics for activities or their results, aimed at the achievement of the optimum degree of order in a given context. Taking into consideration many of the aforementioned problems and challenges for HIS; especially that of common terminology and coding systems, integration and interoperability, security and reliability, it is obvious that solid and well agreed standards are needed (ISO TR 20514, 2005; openEHR Foundation, 2007; Beale, 2001).

International health informatics standardization bodies include ISO (2007), CEN (2007) and WHO (2007) on which national bodies depend. There are also ad hoc, independent organizations producing health informatics standards such as DICOM (2007) or de facto standards from openEHR Foundation (2007) and Object Management Group (OMG) (2007). Relevant standards, especially for the purpose of this study will be described in detail.

The standards are divided into groups according to their area and functions:

1) General standards: They are mostly healthcare or ICT related standards.

CEN/TC251 EN13940 ContSys: System of Concepts for Continuity of Care (2006) is a European standard providing the definition and concepts to support continuity of care.

CEN/TC251 EN12967 HISA: Health Information Services Architecture (2006) underlies the information infrastructure and depicts a shared HIS architecture. It addresses to explicitly describe various services and application interfaces of systems.

2) Identifier standards: They enable proper identification of healthcare entities for use in HIS. These include unique citizen/patient identifiers, healthcare worker and

organization identifiers, medical device and material labeling (such as drugs, disposables, and blood products) (van Bemmel and Musen, 1997; Beale, 2001).

3) EHR content, structure and communication standards: They aim to assist in the interoperability and integration of distributed HIS. There are a number of different, some incompatible (and competing) standards in this category. Key formal bodies include ISO/TC215 (2007), CEN/TC251 (2007), HL7 (2007), ASTM (2007), OMG HDTF (2007) and DICOM (2007). Non-profit organizations such as openEHR foundation (2007) and Open Source Healthcare Alliance (OSHCA) (2007) are also active contributors in this arena.

ISO TS 18308 Health informatics -Requirements for an electronic health record architecture (2004) defines very fine grained EHR Requirements which are utilized by nearly all other EHR standards.

ISO TR 20514 Health Informatics - Electronic Health Record Definition, Scope and Context (2005) defines EHR definition and scope unambiguously.

CEN/ISO 13606 EHRcom (EHR Communications) has been in revision since 2001 to incorporate new models and methods provided by openEHR. It is a multipart standard and Part 1 has been released in the first quarter of 2007. It defines the way in which local HIS and EHR systems can communicate and exchange patient records. This standard helps to support shared patient care between healthcare organizations, and enables life-long provision of care by the ability to access full medical history of patients at all times. This is the first formal standard for EHR communications. Once it has been fully implemented and deployed across a regional or national healthcare network, it is believed to create a virtual life-long health record for every individual. EHRcom also specifies representation of health information, sharing of privacy policies, how specific parts of an EHR can be requested, and how the resulting EHR extract is to be returned (CEN EN 13606-1, 2007).

openEHR Specifications are in fact not formal standards. While the foundation aims to create implementable engineering specifications, they are becoming de facto standards among health informaticians. It follows the separation of concerns as depicted in the ISO Reference Model for Open Distributed Processing (RM-ODP) and separation of information and knowledge to tackle the complexity and ill structured nature of healthcare.

The specifications mainly go into the areas of EHR, demographics and access control. Each area is represented by three components: a reference (information) model, flexible knowledge model (archetype model) and a service model for computational aspects. Unlike other standards, it defines a whole set of concepts and methodologies to design and implement full fledge HIS. The demographic specifications are generic and fully linked to the knowledge model so that non-conforming demographic entities such as names, addresses and relationships can be defined without altering software (Beale, 2000, 2002).

openEHR compliant EHR systems heavily rely on external knowledge resources such as vocabularies, terminologies and ontologies, which define the semantics of terms and concepts referenced in the health record. Archetypes enable multiple terminologies, even in different languages, to be used and linked to other knowledge sources. The Archetype Definition Language (ADL) has been created to capture and represent domain knowledge in a computable way. openEHR also develops and distributes free and open source tools to author Archetypes and validate them. These tools offer even the non-technical clinicians to perform domain knowledge modeling (Beale and Heard, 2007).

openEHR specifications have a major contribution in the radical revision of the former EHRcom standard by adopting two level modeling approach. It has been published in the first quarter of 2007 as a definitive EU standard (Garde, Knaup, Hovenga and Heard, 2007). This study had resulted in considerable contribution to openEHR specifications; especially to the fundamental ADL and AOM specifications which will be discussed in Chapter 4.

Health Level 7 (HL7) (2007) is an U.S.-based, ANSI-accredited health information standards development organization. Its specifications are mostly for application-level messaging (7^{th} of OSI layers) among HIS. Other areas of interest include the structure and content of clinical documents and decision support recently. There are two working versions of HL7 standards, Version 2 and Version 3. The former (v2.4) was approved in 2004 by ANSI (2004) and also became an ISO standard by version 2.5. It is by far the most widely implemented standard in health informatics Worldwide. However the main goal of HL7 before version 3 was to standardize messaging between HIS and achieve data exchange, not to develop EHR standards. Thus there is

no guarantee for interoperability when using HL7 v2.x because there is no well defined underlying information model (Beale, 2003).

Third version of HL7 is still aimed primarily at defining application messages, but now uses a well defined information model, the Reference Information Model (RIM). While assembling messages, the content schemas are derived by a restriction process starting from the RIM, further constrained by domain information models (DIM), restricted message information models (RMIM) and common message element types (CMET). The process ends with forming hierarchical message definitions (HMD) and then generated message schemas are represented as XML documents (Eichelberg, Aden, Riesmeier, Doğac and Laleci, 2005).

There are specifications which are addressing EHR:

- ***HL7 EHR Functional Specification*** defines key functions of EHR Systems (EHR-S) to enable common and unambiguous expression of system functionality for developers. The functions are organized into two groups: the ones that provide support for direct patient care such as clinical reviews, assessments, plans, and documentation within the context of workflow and decision support and the second group includes administrative and infrastructure functions for secondary uses of data to support healthcare operations, research, and public health and quality assessment.
- ***HL7 Templates specification*** An HL7 template is a data structure, based on the HL7 RIM which expresses the data content needed in a specific clinical or administrative context. It expresses a further set of constraints on the RIM. Templates are used to further define and refine these existing HL7 models within a narrower and more focused scope. They use terminology and ontologies and describe domain concepts in a computable way. In a way, they resemble Archetypes.
- ***HL7 Clinical Document Architecture (CDA)*** is a document markup standard. Although CDA is not an EHR standard as such, it forms an important component of an EHR. Its semantics is derived from the HL7 RIM and uses the HL7 Version 3 Data Types which are also part of the RIM. It is an XML-based markup standard intended to specify the encoding, structure and semantics of clinical documents for exchange. CDA document content is intended to be human-readable and

supporting narrative text, yet still having some structure and allow for medical coding to represent concepts in a computable manner (Dolin et al., 2006; Eichelberg et al., 2005).

All these specifications brought about by HL7 Version 3 is a shift from messaging paradigm towards establishing a sharable generic EHR model that CEN and openEHR have done. The modeling work done by using openEHR methodology and components shall be presented in comparison with HL7 in Chapter 4.

American Standards for Testing and Materials (ASTM) Committee E31 on Healthcare Informatics

It develops standards related to the architecture, content, storage, security, confidentiality, functionality and communication of information used within healthcare and healthcare decision making, including patient-specific information and knowledge (ASTM, 2007).

E1384 (Standard Guide for Content and Structure of the Electronic Health Record): ASTM Healthcare DTD medical document structure is a standard specification for XML DTD in healthcare.

OMG HDTF

The OMG HDTF (2007) specifications contain a set of interface definitions for key services in HIS development. HDTF, formerly CorbaMed, was one of the pioneers to define separation of concerns. The individual specifications can be summarized as:

- ***Person Identification Service (PIDS):*** provides interfaces for identifying and correlating personal information within and among various healthcare enterprises.
- ***Lexicon Query Service (LQS) or Terminology Query Service (TQS)*** specifies a set of common read-only methods for interrogating clinical terminologies.
- ***Clinical Observations Access Service (COAS)*** provides a read-only interface to "observations" including clinical context and state of patient.
- ***Resource Access Decision Service (RAD)*** provides read and write methods for authorization and access to health information for HIS.
- There are also other specifications like ***Health Information Locator Service (HILS)***, and ***Order/Entry Tracking Service (OETS).***

Integrating the Healthcare Enterprise (IHE) is an industry initiative which specified the Cross-Enterprise Document Sharing (XDS) integration profile for this purpose. The healthcare documents are stored in ebXML registry/repository architecture in order to enable their identification and exchange. IHE XDS is not concerned with document content; it only specifies metadata to facilitate the locating of documents (Eichelberg et al., 2005; IHE, 2007).

4) Terminology/Content standards

Medicine is one of the few domains where extensive domain knowledge is defined through controlled vocabulary or terminology standards. Some of these, such as SNOMED (2007) or GALEN (Rector, Glowinski, Nowlan and Rossi-Mori, 1995; openGALEN Foundation, 2007) are rich semantic networks or in a way ontologies defined through a formal ontology language.

ICD: It was originally published by WHO for classifying and coding of mortality cases. Other uses include establishing a common naming and description of diseases and collection of comparable data for epidemiologic and healthcare management studies (WHO ICD, 2007).

CPT: It is maintained by the American Medical Association (AMA) (2007) and widely used in the U.S. for reimbursement and utilization review purposes. The codes are derived from medical specialty nomenclatures and are updated annually.

SNOMED: It is developed by SNOMED International - a division of the College of American Pathologists (CAP). SNOMED is a comprehensive, multi-axial, controlled terminology created for indexing of the entire medical record. The latest version is SNOMED-CT (Clinical Terms) which is a dynamic, scientifically validated clinical reference terminology that aims to make health care knowledge more usable and accessible. It provides a common language that enables a consistent way of capturing, sharing and aggregating health data across specialties and sites of care. Among the applications for SNOMED-CT are EHR systems, clinical monitoring, clinical decision support, medical research, clinical trials, computerized physician order entry, disease surveillance, image indexing (SNOMED, 2007; Coiera, 2003). The core terminology contains over 364,000 health care concepts with unique meanings and formal logic-based definitions organized into hierarchies. As of January 2005, the fully populated table with unique descriptions for each concept contains more than 984,000 descriptions. Approximately 1.45 million semantic relationships exist to enable

reliability and consistency of data retrieval. It is available in English, Spanish and German language editions. SNOMED will be used in the study as an example for mapping of MST to external terminology systems.

GALEN: It aims for the development and dissemination of terminology models, methods, architectures and tools for HIS. Various methodologies were developed to allow clinical information to be captured, represented, manipulated, and displayed in a radically more powerful way. The GALEN ontology and GALEN Representation and Integration Language (GRAIL) were developed to manage the content. GALEN-IN-USE project then developed the Common Reference Model (CRM) for Medical Procedures which is used to support EHR, decision support, information retrieval and natural language processing systems in healthcare. This is an important feature for EHR and modeling (openGALEN Foundation, 2007).

UMLS: It is developed by the NLM. The UMLS project develops and distributes multi-purpose, electronic "Knowledge Sources" and associated lexical programs (Humphreys and Lindberg, 1993). The UMLS Meta-thesaurus is one of three knowledge sources developed and distributed by the NLM as part of the UMLS project. The Meta-thesaurus contains information about biomedical concepts and terms from many controlled vocabularies and classifications used in patient records, administrative health data, bibliographic and full-text databases and expert systems. It preserves the names, meanings, hierarchical contexts, attributes, and inter-term relationships present in its source vocabularies; adds certain basic information to each concept; and establishes new relationships between terms from different source vocabularies (UMLS, 2007). The content comes from the electronic versions of many different thesauri, classifications, code sets, and lists of controlled terms used in patient care, health services billing, public health statistics, indexing and cataloging biomedical literature, and/or basic, clinical, and health services research which are uniformly called as "source vocabularies" (Bodenreider,Willis and Hole, 2004). UMLS has big importance for realization of shared and distributed EHRs because it has the potential to make semantic transformation from one EHR node to another in case that they use different terminology. Because UMLS maps most of the widely used coding, classification and terminology systems onto each other, the task is possible. It has also of utmost importance in linking HIS and especially EHR systems to knowledge sources in a manner that is relevant and timely at the point of care (van Mulligan, 1999). This is possible because the concepts and hence the terms are linked

to the Medical Subject Headings (MESH) which is used for indexing in biomedical publishing known as MEDLINE databases and presented online by the PubMed service (Coiera, 2003; Humphreys and Lindberg, 1993). UMLS will be used in the study for providing link of local MST terminology to other external terminologies by utilization of integrated version of MST in the UMLS Meta-thesaurus.

Eichelberg et al. (2005) list other terminology/content standards as follows:

LOINC includes identifying individual laboratory results (e.g. hemoglobin), clinical observations (e.g. discharge diagnosis), diagnostic study observations (e.g. chest x-ray impression).

UMDNS is developed by the Emergency Care Research Institute (ECRI), a non-profit organization affiliated with WHO.

GMDN is developed by EU for providing nomenclature for medical devices to be used in HIS.

ATC is developed by WHO and is a classification system for drugs.

INN is developed by WHO and provides unique naming of all existing drugs. It is important for safe prescribing and exchange of data.

5) Data Exchange/Messaging Standards

HL7 is mainly a messaging standard for HIS to enable healthcare related transactions within and among HCP, and other players, such as HMO, insurance and social security organizations. Mainly the administrative part of healthcare information is taken into account. These include patient ADT, orders for drugs, procedures or tests and their results, messages relating to finance and billing information and clinical observations focusing primarily on measurements. The HL7 specifies the precise messaging syntax to be used, including definitions of segments and internal code strings. Because many of these messages have been developed to support the administration of patient care rather than supporting the work of individual clinicians, the clinical content of the messages is often quite limited (HL7, 2007; Beale, 2003; Eichelberg et al., 2005).

DICOM defines the message formats and communications standards for diagnostic and therapeutic medical images that are widely used by the industry. The importance

of representing clinical and descriptive data within many types of medical images and interoperability issues has led to standardize imaging procedure descriptions and DICOM image interpretation reports. This standard information model for the representation of medical image structured reports is called DICOM-SR. It also uses terminology for content to permit semantic analysis of reports. This work has resulted in a joint terminology system which is published as the SNOMED DICOM Microglossary. The DICOM-SR compliant reports can directly be incorporated in the EHR (DICOM, 2007).

ASTM E31 committee produces a number of messaging standards in the healthcare domain (ASTM, 2007):

- ***ASTM E1238*** (Standard Specification for Transferring Clinical Observations between Independent Computer Systems) is used by most of the largest commercial laboratory vendors in the U.S. to transmit laboratory results.
- ***ASTM 1394*** (Clinical Laboratory Instruments to Computers) has been developed by a consortium consisting of most U.S. manufacturers of clinical laboratory instruments and is being implemented in the current laboratory instruments generation.
- ***ASTM E1467*** (Standard Specification for Transferring Digital Neurophysiological Data between Independent Computer Systems) defines codes and structures needed to transmit electrophysiologic signals and results produced by electroencephalograms and electromyograms. The standard is similar in structure to ASTM 1238 and HL7, and is being adopted by all of the EEG systems manufacturers.

ASC X12: This committee is developing message format standards for transactions between payers and providers (ASC X12, 2007).

IEEE: It also develops health informatics standards (IEEE, 2007).

- ***(IEEE) P1157 Medical Data Interchange Standard (MEDIX):*** IEEE Engineering in Medicine and Biology Society (EMB) is developing the MEDIX standards for the exchange of data between hospital computer systems. It is based on all seven layers of the OSI reference model (Beale and Heard, 2006).

- ***IEEE P1073 Medical Information Bus (MIB):*** This standard defines the linkages of medical instrumentation (e.g., critical care instruments) to point-of-care information systems.

6) Confidentiality, data security and authentication standards

Standards by ASTM Committee E31, HL7, ISO TC 215, CEN TC 251 and openEHR provide a number of such standards. Also numerous organizations active in healthcare informatics area have produced many useful artifacts for quality indicators, data sets and guidelines (Eichelberg et al., 2005).

2.3 Domain Modeling: Reference Models and Archetypes

2.3.1 The Need for Modeling in Healthcare

A model is simplified description of a complex entity or process. Modeling is used to aid in understanding problems by abstraction and simplification (Webster's Online Dictionary, 2007; Wikipedia, the free encyclopedia, 2007). Information systems development is a highly complex and abstract effort; the nature of the medical domain brings another layer of complexity when developing HIS. Thus, proper modeling is vital in order to make sufficient levels of abstraction to successfully handle the challenges. The use of conceptual structures, description languages and visual understanding when modeling helps to make the development process more efficient, productive, effective and easily repeatable. Apart from the direct benefits of modeling during software development it also acts as an efficient communication tool among technical people, managers and also users (Kontrac Whitepaper, 2003).

There are a variety of modeling methods used within the context of IS development. They are divided into two broad categories as general-purpose and domain-specific modeling (DSM) techniques. Modeling languages/tools like UML, EXPRESS, IDEF and XML belong to the former category and Object Constraint Language (OCL), Queries/Views/Transformations (QVT), macro languages and even operating system shells belong to the latter category. Modeling of Archetypes by using ADL and tools used in this study are good examples of DSM which is used throughout the study (Langlois, Exertier and Devda, 2006; Fowler, 2007).

Traditionally, software development is a number of mappings from conception of domain, to analysis and design models, and then onto source code. These mappings tend to be slow error prone and difficult to turn into coding. DSM addresses these problems by removing the resource-intensive and error-prone mappings, aiming to solve the problem only once at the same level of abstraction with the domain itself (Kontrac Whitepaper, 2003). It can be said that DSM is a computationally usable knowledge model of a particular domain. Main modeling approaches relevant for this study include: Conceptual (data) modeling, information modeling, object modeling and knowledge modeling.

2.3.2 Domain Modeling in HIS

Information systems development process, after requirements elicitation phase, usually starts with data modeling. That is an analysis of what kind of data are to be handled, how they will map to database or programming data types and their relationships. Rest of the system may be modeled with conventional software engineering formalisms such as OO by using UML. Due to the nature of healthcare computing environment, classical approach usually ends up with systems that are extremely difficult to develop (hence expensive) and hard to maintain when requirements change (mostly functional requirements related with domain) as system gets older. It has been repeatedly shown that, in healthcare, requirements are on constant change. These changes root not only from user requirements but also rapid changes of domain knowledge. Thus both the software and database have to be modified (Beale, 2000, 2002; Garde et al., 2007; Garde, Knaup, and Hovenga, 2005; Munoz et al., 2007).

The first step in domain modeling is to create domain specific and generic information models instead of complex and rigid data models. These information models are the blueprints for the real world collection, processing and persistence of data. A classical example is the structure of medical records and their features and relationships: an individual patient record consists of some folders which may have one or more sections for different purposes. Under each section, there may be different compositions such as physical examination, patient encounter form or laboratory results. Clinically relevant structures, such as folders and sections, go into the information model. Another example involves an information model for describing the details of a medical observation: A medical observation involves not only recording of

pertinent findings but also recording of patient status and context of the observation. So in the case of a blood pressure measurement, the information model must have capability to capture blood pressure values, clinical context such as sitting or lying and cuff size. In this case the overall structure and semantics of a medical finding goes into information model (Beale, 2000, 2002, 2005; Munoz et al., 2007).

It is obvious that assigning separate database fields in numerous tables to each and every medical entity or relationship is not a good idea when there are many data types and structures with highly complex content. Therefore the information model and underlying data model must be generic; meaning that it has to make a good level of abstraction of underlying data so as to effectively embody complex medical data by a simple and flexible model. The key benefit of using a generic and reusable information model is to achieve interoperability; not only at data level but also at semantic level. Because an information model coupled with a standard terminology enables the meaning of data to be understood same at each site. Only non-volatile and basic knowledge is represented in a reference model and the domain knowledge still have to be stored in the software model. Therefore changes in this kind of requirements still necessitate revisions in the software and database (Beale, 2000, 2002, 2003; ISO TR 20514, 2005; Garde et al., 2007).

Knowledge level modeling includes first separation of explicit domain knowledge and processes (i.e. workflows) from information and data models and then assembling them in a computable way. In the runtime domain specific and generic reference models are further organized and constrained by the knowledge model. Therefore when knowledge and workflow related requirements change, only the knowledge model is altered, not the application and database (Beale, 2000, 2002).

Bird (2003) classifies EHR systems development approaches as:

Unstructured approach: it is a simple data warehouse containing unstructured free-text which is very easy to build and use. However detailed queries and reporting on data is cumbersome and decision-support is impossible.

Big model approach: system is built by assigning a separate table and class for each medical concept resulting in a very big database schema and software model. It is prone to errors and becomes brittle over time because of constant addition of new concepts or changes in existing ones.

Generic model approach: a generic model is designed to accommodate a variety of data into a general purpose set of data structures. For example instead of designing a separate table and fields for different set of biochemistry tests, a single general purpose table/class is created to allow all tests in the set. However it is still problematic because virtually anything may be recorded into the generic model which degrades the data quality and results in difficulties in querying and decision support.

Archetype based two-level modeling: This adds one more layer for knowledge modeling, archetypes, to separate highly volatile domain knowledge from underlying information and data levels. It further constraints the RM and provides data validation during entry thus ensuring a better data quality. Detailed information about this modeling approach shall be given in Chapter 4.

2.3.3 Common Domain Modeling Formalisms for HIS/EHR

Object Oriented (OO) or Object Relational (OR): This is the mainstream modeling and development approach today. After capturing requirements, all the domain entities and processes are mapped to object models by possibly using UML and CASE tools. So the domain knowledge is hard coded directly into the software code, database schema and the user interfaces. There is no clear separation of information and knowledge. However this methodology is well understood by developers and has large industry support with proven success in domains with well defined boundaries and set of rules such as banking or hotel reservation systems (Beale, 2000, 2002). Our research prototype, GASTROS, has been developed with this methodology and was invaluable to experience its shortcomings and show us why to avoid such approach.

HL7 v3 RIM, Data Types and CDA Release 2

Third version of HL7 standards are based on strict models for data and information representation relevant for all aspects of healthcare communications via structured electronic messages. All messages are instances of RIM and the Version 3 Data Types (V3DT). The RIM provides an object-oriented model of clinical data (HL7 RIM, 2007). In other words it is a generic information model, which consists of six core classes: *entity*, the *role* the entity can play, *participation*, *act*, *role relationship* mediating interaction between entities in the appropriate roles and *act relationship* for

chaining different activities. The RIM and appropriate terminologies provide infrastructure for domain knowledge modeling. The content and semantics of HL7 Version 3 messages are derived by a number of restriction processes. First unused classes and attributes are removed from the RIM, then remaining classes are replicated (cloning) which is followed by another iteration of restriction to create DIM, RMIM, CMET, and HMD and finally message schemas are created and represented in XML (Beale, 2003). Within this context that the RIM describing classes (entities), attributes and their specializations for developing messages, it is argued that HL7 providing interoperability for HIS beyond data level interoperability is questionable (Blobel, 2006).

The purpose of CDA is to provide semantic level interoperability in HL7 systems. It is a document markup standard that specifies the structure and semantics of a clinical document (such as a discharge summary or progress note) for the purpose of exchange. A CDA document is a defined and complete information object that can include text, images, sounds, and other multimedia content. It can be transferred within a message and can exist independently, outside the transferring message. CDA documents are also represented in XML and they derive their machine processable semantics from the RIM, coupled with terminology. The CDA model has high expressive power enabling the formal representation of clinical statements which can be processed and understood electronically same at the receiving location in addition to humans (Dolin et al., 2006; Ferranti et al., 2006).

In overall HL7 is considered extremely complex and designed for implementation by technical people; not appropriate for conceptual modeling by domain specialists (Fernandez and Sorgente, 2005; Smith and Ceusters, 2006). For this reason domain modeling by HL7 was not preferred in the study.

ASTM International Continuity of Care Record (CCR)

CCR is a clinical framework that was originally developed by health care practitioners to meet the information exchange needs of primary care providers. The aim is to provide consulting physicians with relevant and timely information necessary to participate in a patient's care. From an EHR point of view, the CCR can be described as a health record data extract. The CCR supports the use of terminologies such as SNOMED-CT. It uses XML to facilitate the exchange of structured medical data and

has an object-oriented data model. CCR documents are specified by an XML schema and accompanying implementation guide. CCR specification appears to overlap with the CDA in both complexity and scope (Ferranti et al., 2006; ASTM, 2007).

ORCA and openSDE

The Open Record for Care (ORCA) is a powerful EPR system developed by Erasmus University, Rotterdam (openSDE Website, 2007). The most prominent feature of ORCA is the support of knowledge-driven data entry. Based on the descriptive knowledge base containing medical concepts and their semantics, ORCA provides other HIS to support menu-driven SDE. Concepts from the knowledgebase can be added to a GUI form which provides direct shortcuts to the SDE interface so that it is not necessary to navigate to these concepts. ORCA makes the distinction between domain dependent and domain independent data (Yamazaki and Satomura, 2000; van Ginneken, de Wilde, van Mulligan and Stam, 1997).

The OpenSDE project is the successor of ORCA in which the SDE component is removed and continued as an ongoing research project. openSDE is an application for SDE in clinical medicine. It is not a complete HIS but focuses on groups of patient data that is normally stored as free text (findings, reports, patient history, physical examination, etc). The approach is quite novel and generic so that it has been applied in a number of different clinical settings. It has a domain-specific modeling tool (Domain Model Editor) and a propriety language for modeling a particular medical domain. This model is used to specify both structure and content of data user can enter. It consists of a framework and contains no predefined data-model that limits the set of data that can be collected. Based on the domain model, the OpenSDE application has a strong automatic GUI generator with some basic consistency checks on the data (Los, van Ginneken and van der Lei, 2005; Los et al., 2004). For these reasons openSDE was the second strong candidate in this study.

Protégé

Protégé is an ontology editor and a knowledge-base framework tool developed by the Stanford Medical Informatics group (Noy, Fergerson and Musen, 2000; Protégé Website, 2007). It is used to construct domain models and knowledge-based

applications with ontologies. Protégé embodies a set of knowledge-modeling formalisms and functions for creation, visualization, and manipulation of ontologies in various representation formats. It is very generic and can be used to model ontologies from any domain. Other than being an ontology tool, it also has an automatic GUI generator which is driven by both the structure and the semantics of underlying domain knowledge. It has been used successfully in healthcare, mainly for clinical study data capture purposes. However due to the frame-based type of knowledge representation, constraints can not be applied on the instances of objects.

Episodus and Ligne de Vie

The system is based on two components: Episodus, the smart client and the Ligne de Vie or Life Line. It is aimed at supporting of continuity of care. Apart from being an EHR system, it is also an effective communications system. For achieving data and semantic interoperability, the system uses a very generic data model (Extended Unified Model) and an ontology for providing structure and semantics of contents. Episodus is the knowledge management component and Ligne de Vie is the front-end application to graphically present one's whole health related life events and also link to source documents such as laboratory reports, discharge letters and so on. The system is also considered as a personal health project management platform because it allows defining health issues and health goals interactively. Then the system alerts both the patient and carer if goals are not met (personal communications with Philippe Ameline).

Propriety Modeling by HIS vendors

Of course the shortcoming of classical development methods in building large scale HIS became evident for the industry in recent years. While retaining the core components of their systems with no volatile knowledge requirements, they have established generic data models and also information models for use in clinical domains. While gaining considerable advantage in terms of cost savings and client satisfaction, mostly these developments have been made in a propriety fashion - no compliance with standards. We had not observed true separation of content and structure or information from knowledge in such systems. They either use existing

modeling languages and tools such as UML or Protégé for domain models or implement their own. Since most of the products are closed-source, it is difficult to assess them (personal experience of the author).

2.4 Archetype based Multi-level Domain Modeling for HIS/EHR

Multi-level modeling of HIS helps to separate the tasks of application developers from tasks of the domain experts. In the technical environment, developers use small and generic information models (RM) and specifications to be implemented in systems. At runtime, information systems are driven by the knowledge models called "Archetypes" authored by domain experts using high-level knowledge modeling tools. Such systems can evolve smoothly, by mainly the tool-enabled modeling by domain experts, rather than by classical software maintenance methods which is costly. Considerable part of HIS which is related with volatile domain terms, business rules and processes are defined by external terminologies and ontologies, information and domain models including workflow and GUI definitions to create runtime features and behavior of software on the fly. Thus software development can proceed separately from domain modeling and if new concept models are introduced or altered the software does not need to be redesigned, coded, tested and deployed (Atalağ and Bilgen, 2007a; Beale, 2000, 2002).

2.4.1 What's in the Levels?

Most IS today are constructed according to single level modeling (Beale, 2002). The database, software and graphical user interface are developed based on OO or OR model. In typical relational developments, concepts are encoded in the relational schema and informally into program code or stored procedures. In object-oriented systems, they are expressed as an object model in formalism such as UML (Rumbaugh, Jacobson and Booch, 1998). By single-level models we mean creating a big and complex model and then writing software and creating databases according to this. When requirements change, most probably due to alteration of domain knowledge, this model has to be modified and the development process enters into a new iteration. In two or dual-level modeling, there is clear separation of information

(RM) and knowledge (Archetypes) levels. Multi-level modeling is characterized by further levels of modeling beyond Archetypes which includes templates, service models and well defined external ontology mapping services (Atalağ and Bilgen, 2007a; openEHR Foundation, 2007).

2.4.2 Reference (Information) Models (RM)

They represent the global characteristics of health record entries, how they are aggregated, and the context information required to meet ethical, legal and provenance requirements. It contains not only domain related generic information like data structures and types, demographic models but also defines non-volatile domain concepts like the general structure of EHR. An example is definition of an electronic patient dossier where a *Folder* can contain *Compositions*, which can contain *Headed sections* containing *Data items* or *Clusters* of data items. The RM also includes ability for versioning of medical transactions and change-sets to handle input errors, simultaneous multi-user read/write access to the record, provides necessary information infrastructure to meet medico-legal needs and historical process analysis (Kalra, Austin, O'Connor, Patterson, Lloyd and Ingram, 2001)

2.4.3 Archetypes and Templates

Archetypes are constraint-based models of domain knowledge which heavily use terminologies for common language and allow assembling of elements from RM and further constrain them. Each Archetype describes configurations of data instances whose classes are defined in RM. Practically they specify particular record entry names, data structures, data types, prescribed value ranges and values for some of the context attributes. When structured knowledge models of medical concepts such as “laboratory result”, “physical examination” or “medication order” are modeled by Archetypes, they can be shared and reused for establishing a high level of semantic interoperability.

Another important feature of Archetypes is that they are language neutral by using internal and external terminologies and ontologies. A good analogy to understand how RM, Archetypes and Templates relate to each other is using small number of standard LEGO parts to assemble some well-known structures such as a car or a house. Likewise Archetypes use RM elements as conceptual building blocks to assemble well-known medical entities. As one may mix different LEGO sets or even take them

to friends and mix them all, using a common RM and sharing Archetypes among different sites, the HIS will be able to handle data without loss of contextual information which provides semantic interoperability.

Templates assemble Archetypes into larger structures like a screen form, document, report or message and further constrain them for local use. They may add further local constraints on Archetypes, including removing or mandating optional sections, and may define default values. At runtime, templates are used with Archetypes to create data and to control its modification. Template design is usually strongly linked to the design of corresponding screen forms (Beale, 2000, 2002; Atalağ and Bilgen, 2007a; openEHR Archetype Definitions and Principles Revision 1.0, 2007).

2.4.4 Service Models (SM)

They provide computational viewpoint of the architecture which consists of service definitions for the EHR and mostly derived from existing work in OMG HDTF, CEN HISA and individual implementation experiences.

2.4.5 Terminology and Ontologies:

They are knowledge resources for EHR such as vocabularies, terminologies and ontologies which define the semantics of terms and concepts referenced in the health record. Archetypes enable multiple terminologies to be used in any natural language in which they are available.

2.4.6 ADL/AOM & TDL/TOM

Archetypes are defined by a formal language called Archetype Definition Language (ADL). The purpose of ADL is to provide an abstract syntax for textually expressing Archetypes and Templates. It utilizes three other syntaxes: Object Constraint Language (OCL) represented as cADL, Data Definition Language (DDL) represented as dADL, and a version of First-Order Predicate Logic (FOPL) to describe constraints. An ADL compliant archetype consists of the following sections:

- Archetype (ADL version and archetype name)
- Specialize (optional: parent archetype name)
- Concept (name of the domain concept modeled)
- Language (dADL: language details)

- Description (dADL: archetype meta-data)
- Declarations (optional: FOPL declaration statements)
- Definition (cADL: formal constraints)
- Invariant (optional: FOPL assertion statements)
- Ontology (dADL: terminology and language definitions)
- Revision_history (optional: dADL history of change audits)

The Archetype Object Model (AOM) describes an object model equivalent of the ADL syntax in terms of a UML model. It is a generic model, meaning that it can be used to express Archetypes for any reference model in a standard way. ADL and the AOM are brought together in an ADL parser: a tool which can read ADL archetypes and then create a parse-tree (resulting in-memory object representation) as instances of the AOM (openEHR Archetype Object Model (AOM) Revision 2.0.1, 2007).

Templates are expressed in the dADL syntax from ADL; so they do not have a separate language. TOM defines the object model of templates, which are themselves used to put Archetypes together into local information structures, usually corresponding to screen forms (openEHR Template Object Model Revision 0.5, 2007).

2.4.7 Developing Archetypes and Templates: Tools

The open source openEHR Archetype Editor is a tool for domain experts to create and edit Archetypes. It has a GUI which is quite user friendly for non-technical users. When creating new Archetypes openEHR, CEN or HL7 RM can be specified and it supports the ADL.

The open source Technical Workbench is another tool that allows for comprehensive review of an Archetype and all its elements while in design process. It parses the Archetype and displays it in the chosen language as a tree. It is an excellent debugging tool as error messages are displayed and line numbers are given to the modeler. The syntax and ontology section of Archetypes can also be validated by using this tool (openEHR Foundation, 2007).

Ocean Informatics Template Designer is a commercial data entry tool that allows composition of Archetypes to meet the needs of different users. Components of Templates can be constrained further including or excluding parts of the underlying

Archetypes as long as the rules of those Archetypes are not violated (Ocean Informatics, 2007).

2.4.8 Relationship of Modeling with Terminologies and Ontologies

Archetypes are not competing with terminologies; instead they are complementary to each other. They form the interface between the information model and terminology. Terms within an Archetype can be linked to multiple external terminologies. Archetypes allow linking of not only the domain terms (mapping) to the external terminologies but also can be constrained by them, such as limiting a list of allowed values for diagnosis for a specialty. This greatly enhances the semantics of concepts in Archetypes.

Archetypes can be constructed without using any external terminologies like SNOMED or ICD. The individual nodes are identified and named by using a custom built internal terminology and then this is linked to external terminologies. This feature of Archetypes enables construction of a micro-terminology for use in the particular concept, hence eliminating the need for acquiring and maintaining large and complex terminologies (Garde, Knaup, Schuler and Hovenga, 2005a). Term_binding subsection in Ontology section of an Archetype is used to denote the equivalences between Archetype local terms and terms found in external terminologies. Therefore a query engine searching for an instance of some external term can determine the equivalent local term used in the Archetype. One other important point is the ease of translation/localization of Archetypes because only the local terminology has to be translated. This helps in preserving the semantics of an Archetype across different cultures and languages (Garde et al., 2007).

Medical terminologies are important for establishing a common language among healthcare workers and also serve for data standardization. By the use of Archetypes and terminologies together, mapping of clinical terms in HIS/EHR to terminologies will be possible and form an important step in data standardization. Data interoperability will also help for achieving system interoperability (Quamar and Rector, 2007).

The mapping of internal/local Archetype terms onto terminologies and ontologies not only provides proper terms and semantics but also enables access to external knowledge sources such as UMLS, SNOMED, MESH and PubMed. This is an

important step for enabling aforementioned context-sensitive provision of knowledge at point of care (i.e. provide an obstetrician relevant knowledge about risk of abortion when examining and recording findings into EHR). It has been shown that providing physicians relevant knowledge at point of care improves decision making and outcomes (Garde, Knaup and Hovenga, 2005b; van Ginneken 2002).

2.4.9 Interoperability & Governance Issues

Perhaps the most important function of EHR is the ability to share health information among different authorized users. This requires interoperability of information in the EHR and interoperability of EHR systems which exchange and share this information. According to ISO, there are two main levels of interoperability for information:

1) Functional (data) level interoperability: the ability of two or more systems to exchange information (so that it is human readable by the receiver).

2) Semantic level interoperability: the ability for information shared by systems to be understood at the level of formally defined domain concepts (so that information is computer processable by the receiving system). However the degree of semantic interoperability may change depending on the level of agreement on terminology and the content of Archetypes and Templates used by different parties.

These two levels of interoperability of information and systems are possible by the use of RM, Archetypes, Templates and terminologies/ontologies consistently (Garde, Knaup and Hovenga, 2005b; Beale, 2000, 2002).

Archetype Repositories and Domain Knowledge Governance

Sharing Archetypes designed for a particular purpose is a useful method for standardized data collection and interpretation provided that all parties are using the same RM. This enables the reuse of not only validated and approved clinical concepts but also allows for similar analysis and design patterns in developing similar Archetypes. And therefore the reuse of Archetypes is expected to decrease the cost and enhance the knowledge development environment while increasing patient safety.

Archetype repositories are needed when a diversity of health information is present in a region or enterprise. Then a wide range of Archetypes are required to establish a federation and ideally all parties should agree on common definitions for exchanging

this information. By conforming to a common RM and Archetype model (AM), the relevant libraries of Archetypes at each repository can be exchanged. In the longer term, it is expected that by the involvement of national health agencies, academic organizations and professional bodies in the development process of Archetypes, will contribute to achieve quality evidence-based clinical practice. In the future, regional or national public domain libraries of Archetype definitions might be accessed via the Internet, and downloaded for local use within EHR systems.

Once common use of Archetypes is agreed upon by healthcare professionals for delivering best-practices, the issue emerges as how and who to manage, maintain, update and disseminate them which hold precious domain knowledge. Hovenga, Garde and Heard (2005) recommend either using already established organizations such as the Cochrane Collaboration or to establish new organizations to take responsibility of domain knowledge governance.

Because the core set of Archetypes, such as blood pressure or blood chemistry, is the result of a small number of clinicians, there will be little, if any, argument on that issue. Also the specialization of the core set of Archetypes by individual users for local use is considered safe because it does not break the original semantics. However Garde et al. (2007) depicts a long list of reasons to establish this kind of governance. It can be summarized as the need for: managing overlaps among various domains, standardizing a set of Archetypes for interoperability, being easily accessible and locatable, ensuring best quality of knowledge contained in Archetypes, maintaining and updating Archetypes when domain knowledge changes. In short, domain knowledge governance is proposed to provide that the Archetypes in use will meet the knowledge requirements of various areas while minimizing redundancy and establishing semantic interoperability.

2.4.10 Advantages and Promises of Archetypes

Archetypes lead to knowledge-enabled systems where information and knowledge aspects in systems are separated, thus allowing cheap and future-proof systems. They also provide functional and a high level of semantic interoperability by the use of standardized Archetypes and terminologies so that systems can reliably communicate at the level of knowledge concepts. The separation of information and knowledge results in separation of tasks in software development which leads to domain empowerment. Another pertinent advantage is the ability to perform intelligent and

efficient querying on data by making use of structure and logical paths of Archetypes from which the data was created. So instead of doing a brute-force database query, there is now the possibility to locate where a certain data item is likely to be stored (Beale, 2000, 2002; openEHR Archetype Definitions and Principles Revision 1.0, 2007).

Archetypes promise a great deal of improvement in healthcare by solving some of the pertinent aforementioned problems of healthcare ICT. These include reducing software maintenance (Archetypes and terminologies change only), increasing data validity (Archetypes are used to validate all data input), establishing interoperability, complying with standards for integration with other software and systems, and enabling integration with legacy systems. As a summary, in multi-level modeled systems, clinical data are more likely to be correct, are more sharable, and software is not subject to single vendor lock-in - all leading to better quality and more cost-effective clinical care (Beale, 2000,2002; Blobel, 2006; Garde, Knaup and Hovenga, 2005b).

Archetypes and Multi-level modeling are also considered as the key enabling factor realization of EBM. The widespread adoption of standard Archetypes is expected to form a basis for capturing best practices and developing clinical guidelines. Furthermore they offer the possibility to conduct large scale randomized clinical trials by using existing clinical databases, real world demographics and clinical information including severity of illness which makes possible the evaluation of healthcare outcomes (Hovenga, Garde and Heard, 2005).

2.4.11 Challenges with Archetypes and Multi-level Modeling

An assumption about this approach is that the RM on which Archetypes are built upon will remain fairly stable over time. This will of course enable systems to evolve with minimal effort as medical knowledge and processes change. However if the RM is modified, the same consequences in traditional single-level systems, maybe more, are valid for Archetype based systems: modification of software and related components. Although the Archetypes promise more cost-effective and future-proof software systems, it takes a great amount of investment and effort to create a correct, viable and agreed-upon RM (Bird, Goodchild and Tun, 2003).

2.5 IS in Gastrointestinal Endoscopy and MST

This subject will be discussed here because it is an important part of this study and is our domain of modeling and the case study as mentioned in Chapter 1. Gastrointestinal endoscopy is a relatively new field in medicine which depends on visualization of the gastrointestinal tract for both diagnostic and therapeutic purposes. It is also important that, while endoscopy is a small and manageable domain, gastrointestinal endoscopy is a highly specialized, technology oriented, and a critical niche area in medicine so the results need to be reliable, complete and unambiguous (Moorman, van Ginneken, van der Lei, Siersema, van Blankestein and Wilson, 1994b). Therefore it is logical to assume why terminology standardization and the quest for data sharing have started quite early (Moorman, 1995a; Atalağ et al., 2007b). Due to the small size and manageability of gastrointestinal endoscopy with the availability of an appropriate domain terminology already translated into Turkish had strongly influenced our decision in designing this case study.

2.5.1 Terminology Standardization

Motivation

The importance of precise language in medicine cannot be overestimated. Correct terminology is part of a correct diagnosis (Korman, Delvaux and Bidgood, 1998). Conduct of medical practice arises from the ability to observe and communicate intelligibly. Endoscopists view the gastrointestinal tract and create text and images that reflect their observations and transmit this information to other parties in patient's care team. However the traditional approach in reporting is heavily based on free-text and both the diagnostic and descriptive terms for observation are highly variable. Moorman et al. (1994b) has reported in a Delphi study that 19 from a selected set of 28 findings were not properly described in half of the endoscopy reports. Even worse, on average 14 topics were missing from each report which brings about suspicion for degraded quality of care. In another study Moorman, Siersema, van Ginneken, van Blankenstein and Wilson (1994c) has shown that referring physicians considered half of the endoscopy reports unsatisfactory. Earlier studies (Moorman, 1995a; Moorman, Siersema, de Ridder and van Ginneken, 1995b) sought the consistency and reliability for the use of non-numerical expressions (i.e. small, medium, large) for describing the

size a common endoscopic finding – gastric ulcer. The overlap among terms was large and in 31.1% of cases, the term “large” did not exceed “small”.

The discrepancies in description and interpretation of endoscopic reports also lead to problems in processing, analysis and sharing of data in HIS. Thus due to lack of common terminology and also use of unstructured free-text in reporting, most endoscopy reports are difficult to retrieve, transmit or link to other parts of the patient record (Korman, Delvaux and Crespi, 2001).

The first aim of terminology standardization is to directly improve patient care by allowing endoscopists to create complete and non-ambiguous reports more efficiently while decreasing errors and costs. The ability to access, process, and communicate endoscopic findings by electronic means is also an important benefit. Second, it is likely to enhance the education and training by the creation of more precise textual and visual definitions of endoscopic terms (Korman, Delvaux and Crespi, 2001). It is also important in communication and documentation as a reference source. Third, it is important for scientific research which makes use of endoscopy by structuring and classifying clinical data and allowing semantic linking of records and systems (Korman, Delvaux and Bidgood, 2001).

OMED and MST Terminologies in Gastrointestinal Endoscopy

These serious problems and promises of having a standard terminology in endoscopy, have led international gastrointestinal endoscopic associations to take necessary steps. The European Society of Gastrointestinal Endoscopy (ESGE) created a Committee of Terminology in 1976. The American Society for Gastrointestinal Endoscopy (ASGE) established an ad hoc Computer Committee to consider the role of computers in endoscopic practice in 1981. The initial research and proposals turned out to be insufficient so another attempt had started to create a terminology system capable of addressing the needs of the practitioner. ESGE started a second attempt in 1991 and continued with the participation of the ASGE and Japanese Society of Gastrointestinal Endoscopy in 1993. The initial requirements were easy implementation in report generator software, limitation to the most common findings without duplicate and redundant terms, acceptance by practicing endoscopists, easy learning and using and establishment of a minimum set of descriptors for a lesion to assure the quality of the description. The result of this effort was MST version 1.0, which was published as a

Working Party Report for the World Congresses of Gastroenterology and Digestive Endoscopy in Los Angeles in 1994. Although MST 1.0 was based on extensive review by multiple panels of practicing endoscopists, it had never been used as a component of report generating software. In Europe European Commission funded GASTER project and in the United States testing was performed with grant support from the American Digestive Health Foundation. A total of 23,658 examinations were performed including EGD, colonoscopy, ERCP, and flexible sigmoidoscopy. The coverage of the terminology in describing endoscopic examinations was extremely high (Delvaux et al., 2000a). The results of these projects were used in revision of the terminology and MST 2.0 was produced in 2000 (Korman, Delvaux and Crespi, 2001; Delvaux et al., 2000a). MST is now translated into eleven languages (English, French, Italian, German, Portuguese, Spanish, Russian, Hungarian, Czech, Turkish and Japanese).

MST is intended as the suggested minimal standard basis of terms which should be included in any software developed by the industry or by individual research (Crespi et al., 1996; Delvaux, 2000b). The list of terms in the MST represents unique concepts that are used to identify a finding. The basic principles for selection of terms required that it has to be readily understood, as unambiguous as possible, and used frequently. The terms for abnormal findings having frequency less than 1% were not included in the MST.

Other than providing a standard list of terms for endoscopic findings, MST also consists of terms and structure for fully describing a valid endoscopy report which is given in Table 2.2.

Table 2.2 MST terms and structure for a full endoscopic record

Reasons for endoscopy	**Organ**
Symptoms	**Findings**
Diseases	Normal
Assessment	Lumen
Sampling	Content
Therapeutic	Flat lesions
Examination data	Protruding lesions
Extent	Excavated lesions
Conditions	**Additional Procedures**
Maneuvers	Diagnostic
	Therapeutic
Complications	**Diagnosis**
	Main diagnoses
	Other diagnoses

MST helps for the mandatory administrative and medical coding (CPT, ICD CM) along with the endoscopic examination. A standard list of diagnostic and therapeutic procedures, a proposed list of common diagnoses, and reasons for endoscopy is provided so as to automatically link to those coding systems. The MST committee has also identified the linking of MST to other terminologies as important (Korman, Delvaux and Crespi, 2001). An important step was the integration of MST with SNOMED-DICOM Microglossary which enabled a standard way linking of endoscopic images with textual endoscopic reports. MST 2.0 has also been integrated with NLM Knowledge Sources: UMLS (Humphreys and Lindberg, 1993; Tringali, Hole and Srinivasan, 2002; Korman and Bidgood, 1997; Korman, Delvaux and Bidgood, 1998). Since links to other reference images, medical terminologies and bibliographic sources are available in UMLS, it is possible to link MST to many other knowledge sources directly without the need for mapping. This is a critical accomplishment to be able to provide context dependent guidance to the endoscopists on the fly while recording findings or to display relevant reference images as an invaluable assistance for diagnosis. The MST committee recognizes these events as a proof for the universal acceptance of this terminology. These were also important

points for the selection of the gastrointestinal endoscopy domain for HIS modeling in the study.

As implied by Grassi and Delvaux (2003), structured reporting by using MST will assist in the statistical analysis of databases for clinical research which will lead to standardization of data in endoscopy, support multicenter trials, overcome the problems of multilingual reporting in cooperative studies, and will promote evidence-based and outcomes research.

2.5.2 Use of Computers in Gastrointestinal Endoscopy

Substantial amount of work has been done for more than a decade in the design and development of endoscopic databases and information systems mainly for increased productivity due to automated structured data entry and rapid report generation as mentioned by Atalağ et al. (2007b). It has been reported repeatedly in studies that structured reports are superior to free-text reports in endoscopy as they offer a built-in quality control during reporting by the ability to specify the terms, valid attributes and their values unambiguously. As the availability of clinical terminologies and standards are essential elements for development of HIS, most (commercial) endoscopic information systems (EIS) have emerged after the introduction of OMED terminology (Maratka, 1995) and then subsequent publication of the Minimal Standard Terminology for Digestive Endoscopy (MST) (Crespi et. al, 1996; Delvaux, 2000b; Delvaux et al., 1998).

2.5.3 MST Applications and Evaluation

During initial evaluation of MST which has led to the second version of MST, parallel projects had been initiated both in Europe and the U.S. to test the validity of the terminology. This testing was funded by the European Commission through the GASTER Project and the American Digestive Health Foundation by ADHF Trial (Delvaux, 2000b).

We have evaluated the second version of MST in the study by using the research prototype GASTROS. In accordance with other previous validation studies, the overall usage of MST terms turned out to be very high: 85% for examination characteristics,

94% for endoscopic findings and 94% for endoscopic diagnoses. Good user acceptance proved that both the terms and structure of MST were consistent with usual clinical thinking (Atalağ et al., 2007b).

The Clinical Outcomes Research Initiative (CORI) is based in the Biomedical Information and Communications Center at Oregon Health Sciences University. CORI has been systematically collecting data on gastrointestinal endoscopy procedures since 1997. Using an international network of endoscopy centers and computer software tools developed at the center, detailed data is collected on these procedures for use in outcomes research. Data are primarily entered into CORI in a highly structured manner but with the option to enter free text if deemed appropriate. The "minimal" MST terminology has been extended by addition of free text capabilities and according to needs of the researchers who use it (Logan and Klopfer, 2000; Logan, McCashland and Lieberman, 2004; Cooper and Sivak, 2000).

Vendor implementations

Although there is a number of commercial MST based EIS exists, possibly ones we do not even know in local settings; only a one of them shall be discussed due to its wide usage and acceptance.

Endobase (version III as of 2006) is an EIS from Olympus Software. It allows the combination of different text blocks to compose a complete report besides the use of standard reports and MST. After selecting the different standard reports, text blocks or MST the composed report can be exported to a word processor. Endobase has relational database model which is able to storing different data produced in an endoscopy unit, including digital images and videos, and retrieve them. Structured data entry forms are also available for the MST (Groenen et al., 2006).

Alternatives and Challenges

None of the terminology systems including MST have the descriptive flexibility of natural language. The relative rigidity of a restricted terminology system in endoscopy has to be counterbalanced by the speed, ease, decreased cost, and quality of the information recorded in the endoscopic report (Korman, Delvaux and Crespi, 2001). It

has recently been reported by Groenen et al. (2006) that endoscopists can create faster reports (two minutes) with standard reports and text-blocks than using structured data entry such as MST. According to their experience, MST is more complex, takes more time and there is a risk of getting lost in the data tree. The advantages are ability to describe findings point by point and establish a structured database (Groenen et al., 2006). However this result may imply at least two things: First the effort and benefits are not balanced for structured versus free-text entry thus endoscopists may not anticipate the indirect benefits. And second the EIS might have better process and GUI design for MST based data entry. The challenge is to accomplish direct and sound benefits for MST based SDE while reducing time and effort during reporting. The research prototype developed in the study has addressed the problem of usability and it was preferred by the endoscopists to enter and generate reports for all of the cases during the case study.

2.6 Conclusion

We have started with the fundamental problems of two distinct but interrelated fields: Healthcare and IS. The obvious complexity in healthcare adds onto the intrinsic problems of IS development, thus making realization of HIS even harder. Effective use of HIS brings about many benefits to healthcare such as controlling of errors and cost and more effective and efficient services. It should also focus on the new trends in healthcare such as EBM or continuity of care. High quality structured data collection is still the most fundamental problem in health informatics and lessons learned in the past decades are hoped to shed light onto future directions of HIS. The EHR is a novel and promising approach for realization of interoperable, reliable and effective healthcare services based on best scientific knowledge, with less variability and equitable access. Health informatics standards will play a key role in this space.

Software modeling has been a vital paradigm for developing complex software and HIS is not an exception. In fact, the added complexity by the healthcare domain, makes modeling a necessity. There are a number of modeling strategies, some with proven success. However when we look at the practical implementations over time, the problem of changing requirements do affect them hence render their maintainability and cost. We have identified this as the core area of improvement during our relatively long experience with the GASTROS research prototype. Multi-level modeling of HIS by using Archetypes has been preferred to model our domain of

interest, gastrointestinal endoscopy, for a number of reasons. First it is a relatively small and controlled domain which is appropriate for the modeling task. Second it has a standardized and well accepted terminology to be utilized in HIS. Third there exist a number of studies and projects with alternative approaches in which we can compare.

It should be noted that the modeling methodology we have chosen is primarily designed for modeling a whole EHR. However the scope of this study is focused on modeling a niche medical domain for the purpose of only HIS development. Of course such a HIS will be a native "feeder" for EHR systems.

As the last remark during literature survey we have observed quite an overlap in the early phases of such research in both domains with publications in both IS and medical literature. According to the domain they are published, most of those studies lack significant contributions to the other domain. However it is interesting to note that the recent publications appear mostly in health informatics journals authored by interdisciplinary team of researchers. We have benefited mostly from the latter type of literature.

CHAPTER 3

EXTENDING MST and GASTROS CASE STUDY

In Chapter 2 we have described a specific medical domain, gastrointestinal endoscopy, and presented an overview of the structure and semantics of MST. After deciding on the gastrointestinal endoscopy domain and MST for the study, we first evaluated the Turkish translation of MST before performing further research. For this purpose we have built a research prototype which was used in a live clinical setting for about three years. After correcting some semantic errors we also modified the hierarchy MST for building an effective and user-friendly HIS.

This chapter consists of two parts. In the first part we will present our in-depth analysis of MST where we have identified significant problems related with the content and structure of MST. These problems and the reasons of modifications are explained in detail with our proposed solutions. The second part presents a case study using the research prototype GASTROS. The results of extensive evaluation and validation of the Turkish translation of MST will be presented and thoroughly discussed.

3.1 Analysis of MST

The main goal of MST is to provide a minimum list of terms needed to describe routine endoscopic procedures, while avoiding usage of synonyms and imprecise or subjective descriptions, to be included in HIS. MST consists of not only a list of

precise and well-accepted terms, but also prescribes considerable domain knowledge in the form of domain hierarchy and semantics in the form of domain terms and relations to fully describe any endoscopic examination. However probably due to the inefficiency of presentation by a simple tabular list within a printed document as appears in the original publication, many semantic problems became apparent when we started initial modeling work. But the most significant problem we had encountered was related with the hierarchy of MST which inhibited our Archetype modeling task. Not only it was inappropriate for developing HIS, but also there were inconsistencies of usage throughout MST. Therefore we have consulted domain experts and also used our own judgment to propose a new hierarchy and tested it with the prototype before proceeding with the modeling task. The details of these problems and our proposed solutions will be discussed in detail.

3.1.1 Problems Identified Related with the Content of MST

There were many serious semantic problems which might lead to ambiguity in describing endoscopic findings, reporting and also during data analysis. These are listed below with original MST publication table numbers where appropriate.

1. In many places there were attributes without values such as “Number”, “Size, “Diameter” and “Length”. It was obvious that the values were assumed to have certain values. However considering the same attributes also have many different enumerated values, it would have been more appropriate to give attribute values like “Number: Quantity (Count)”, “Size, Diameter, and Length: Quantity (Length). In Archetype modeling, these were all corrected.
2. In some parts, MST Terms or Attributes had an extra description and these have been placed as Attributes or Attribute Values which obviously created serious semantic problems. So we have merged them and introduced proper attributes or attribute values as shown in Tables 3.1a, 3.1b, 3.2a and 3.2b.

Table 3.1a Problem related with "Evidence of previous surgery" term in MST Table6-Terms for Esophagus

Terms	Attributes	Attribute Values
Evidence of previous surgery	**Anastomosis**	Esophago-jejunal Esophago-gastric Esophago-colonic

Problem: *Descriptive attribute value "Anastomosis" is placed as an attribute.*

Table 3.1.b Correction(s) related with "Evidence of previous surgery" term in MST Table6-Terms for Esophagus

Terms	Attributes	Attribute Values
Evidence of previous surgery	**Type**	Esophago-jejunal **anastomosis** Esophago-gastric **anastomosis** Esophago-colonic **anastomosis**

Solution: *"Anastomosis" was merged with attrribute values and a new attribute "Type" was introduced.*

Table 3.2a Problem related with "Barrett's Esophagus" term in MST Table6-Terms for Esophagus

Terms	Attributes	Attribute Values
Barrett's Esophagus	**Distance**	Z-line
	Distance	Upper end of Gastric Folds

Problem*: description of attributes are placed as attribute values.*

Table 3.2b Correction(s) related with "Barrett's Esophagus" term in MST Table6-Terms for Esophagus

Terms	Attributes	Attribute Values
Barrett's Esophagus	**Distance of Z-line**	**cm from incisors**
	Distance of upper end of Gastric Folds	**cm from incisors**

Solution*: attribute values were merged with attributes and new attribute values were introduced.*

3. In many places attributes defining distance from a certain anatomic location were present in place of Site(s) which is used for depicting anatomic location(s) of a particular term. The list of values for Site(s) is well defined in MST and it does not contain this value. Since this would pose difficulties in modeling, as a solution an extra attribute and value was introduced to correct this which is presented in Tables 3.3a, 3.3b, 3.4a, 3.4b, 3.5a and 3.5b.

Table 3.3a Problem related with "Stenosis" term in MST Table6-Terms for Esophagus

Terms	Attributes	Attribute Values	Site(s)
Stenosis	Appearance	Extrinsic Benign intrinsic Malignant intrinsic	**cm from incisors**
	Length (cm)		
	Traversed	Yes After dilatation No	

***Problem**: inappropriate appearance of distance attributes in place of Site(s).*

Table 3.3b Correction(s) related with "Stenosis" term in MST Table6-Terms for Esophagus

Terms	Attributes	Attribute Values	Site(s)
Stenosis	Appearance	Extrinsic Benign intrinsic Malignant intrinsic	
	Length (cm)		
	Traversed	Yes After dilatation No	
	Location	**cm. from incisors**	

***Solution**: a new attribute, "Location" and attribute value "cm. from incisors" were introduced.*

Table 3.4a Problem related with "Evidence of previous surgery" term in MST Table6-Terms for Esophagus

Terms	Attributes	Attribute Values	Site(s)
Evidence of previous surgery	Anastomosis	Esophago-jejunal Esophago-gastric Esophago-colonic	**cm from teeth**

***Problem**: inappropriate appearance of a distance attribute in place of Site(s).*

Table 3.4b Correction(s) related with "Stenosis" term in MST Table6-Terms for Esophagus

Terms	Attributes	Attribute Values	Site(s)
Evidence of previous surgery	Anastomosis	Esophago-jejunal Esophago-gastric Esophago-colonic	
	Distance from incisors	**in cm.**	

***Solution**: a new attribute, "Distance from incisors", and attribute value "in cm." were introduced.*

Table 3.5a Problem related with "Ulcer" term in MST Table6-Terms for Esophagus

Terms	Attributes	Attribute Values	Site(s)
Ulcer	Number		Site(s)
	Size **(mm.)**		**Site(cm. from incisor)**

***Problem**: inappropriate appearance of a distance attribute in place of Site(s) and also together with attribute.*

Table 3.5b Correction(s) related with "Ulcer" term in MST Table6-Terms for Esophagus

Terms	Attributes	Attribute Values	Site(s)
Ulcer	Number		Site(s)
	Size	**in mm**	
	Distance from incisors	**in cm**	

***Solution**: introduced a new attribute, "Distance from incisors" and attribiute value "in cm." Also "mm." next to "Size" attribute was placed as attribute value.*

4. In many places more than one Site(s) were assigned to a Term; associated with different attributes or even attribute values. This is due to the fact that in original MST hierarchy Site(s) is linked to Attribute Values. However this creates a big semantic problem for domain modeling and HIS development because there are many cases of terms without any attributes which still need definition of anatomic locations. In our modeling strategy, if attributes need to have different Site(s) information, then the whole term set is repeated, not only the attributes or attribute values. These extra Site(s) are given as boldface in Tables 3.6 to 3.10. As the solution they were removed.

Table 3.6 Problem related with "Mucosal sclerosis" term in MST Table6-Terms for Esophagus

Terms	**Attributes**	**Attribute Values**	**Site(s)**
Mucosal sclerosis	Type	Spontaneous Post-therapy	Site(s)
	Extent	Localised Patchy Diffuse	**Site(s)**

***Problem**: Multiple Site(s) assigned to different attributes of a term.*

Table 3.7 Problem related with "Enlarged folds" term in MST Table7-Terms for Stomach

Terms	**Attributes**	**Attribute Values**	**Site(s)**
Enlarged folds	Extent	Localised Diffuse	Site(s)
	Type	Thick Giant	**Site(s)**

***Problem**: Multiple Site(s) assigned to different attributes of a term.*

Table 3.8 Problem related with "Evidence of previous surgery" term in MST Table8-Terms for Duodenum

Terms	Attributes	Attribute Values	Site(s)
Evidence of previous surgery	Specify		Site(s)
	Suture material visible	Yes No	**Site(s)**

***Problem**: Multiple Site(s) assigned to different attributes of a term.*

Table 3.9 Problem related with "Evidence of previous surgery" term in MST Table7-Tems for Esophagus

Terms	Attributes	Attribute Values	Site(s)
Evidence of previous surgery	Suture material visible	Yes No	**Site(s)** **Site(s)**

***Problem:** Multiple Site(s) assigned to different attributes values*

Table 3.10 Problem related with "Gastrostomy" term in MST Table7-Tems for Stomach

Terms	Attributes	Attribute Values	Site(s)
Gastrostomy	Type	Surgical Endoscopic (PEG)	**Site(s)** **Site(s)**

Problem: *Multiple Site(s) assigned to different attributes values*

5. Some terms, attributes and attribute values were used inconsistently. In table 3.11 correct usage of "Type" attribute with "Evidence of previous surgery" is given. However in Table 3.12a same term is qualified with a different attribute for the same purpose at a different place in MST. As the solution the attribute "Anastomosis" was replaced with "Type" because not all of the given values were in fact anastomosis type surgeries. To correct this semantic problem we have appended "anastomosis" to appropriate attribute values of type anastomosis such as "Billroth I anastomosis" in Table 3.12b.

Table 3.11 Correct usage of "Type" attribute with "Evidence of previous surgery" term in MST Table9-Tems for Colon

Terms	Attributes	Attribute Values
Evidence of previous surgery	**Type**	Colo-colonic **anastomosis** Ileo-colonic **anastomosis** Colo-anal **anastomosis** Ileo-anal **anastomosis** Colostomy

Table 3.12a Problem related with "Evidence of previous surgery" term in MST Table7-Tems for Stomach

Evidence of previous surgery	**Anastomosis**	Billroth I Billroth II Gastroenterostomy Pyloroplasty Anti-reflux surgery Banded gastroplasty

Problem: *Inconsistent use of "Anastomosis" attribute. Also all listed surgeries in attribute value list are not of anastomosis type.*

Table 3.12b Correction(s) related with "Evidence of previous surgery" term in MST Table7-Terms for Stomach

Evidence of previous surgery	**Type**	Billroth I **anastomosis** Billroth II **anastomosis** Gastroenterostomy Pyloroplasty Anti-reflux surgery Banded gastroplasty

Solution: *a new attribute "Type" was introduced and "anastomosis" was appended to appropriate type of surgeries in attribute value list.*

6. In three places, new attributes had to be introduced either because they were either missing or mixed with attribute values or some attribute values were present without any attributes at all. This is presented in Table 3.13a where an attribute value is present without a proper attribute. So we have introduced the attribute "Diameter" (Table 3.13b).

Table 3.13a Problem related with "Tumor/Mass" term in the original Tables 7, 8 and 9; terms for Stomach, Duodenum and Colon respectively

Terms	Attributes	Attribute Values	Site
Tumor/Mass		Diameter in mm.	

Problem*: Missing attribute and inappropriate placement of the attribute value*

Table 3.13b Correction(s) related with with "Tumor/Mass" term in the original Tables 7, 8 and 9; terms for Stomach, Duodenum and Colon respectively

Terms	Attributes	Attribute Values	Site
Tumor/Mass	Diameter	in mm.	

Solution: *introduced a new attribute, "Diameter", and "in mm." as its value.*

7. Terms might repeat themselves with different sets of attributes or same attributes with different values. Consider the clinical expressions from the same endoscopic session "*A solitary ulcer is visualized in sigmoid colon having diameter of 35 mm with oozing type of bleeding and also there were multiple ulcers in the ascending and transverse colon with stigmata of bleeding*". The corresponding MST section is given in Table 3.14.

Table 3.14 Repetition of terms with different attributes and/or attribute values: an example term from the original MST Table 9-Terms for Colon

Terms	Attributes	Attribute Values	Site
Ulcer	Number	Single (solitary) Few Multiple	Site(s)
	Size	Largest diameter in mm	
	Bleeding	Yes: Spurting Yes: Oozing No	
	Stigmata of bleeding	Yes No	

The clinical expression can be represented according to the MST model as follows and in runtime the object representing this term may have multiple instances.

Ulcer➔ Site: Sigmoid colon/Number: Solitary/Size: 35mm/Bleeding: Yes: Oozing

Ulcer➔ Site(s): Ascending and transverse colon/Number:Multiple/Stigmata of bleeding:Yes

However this kind repetition might not be appropriate for some terms especially without Site(s) information. An example is given in Table 3.15.

Table 3.15 A non-repeatable term from the original MST-Table 9 Terms for Colon

Terms	**Attributes**	**Attribute Values**	**Site**
Hemorrhoids	Bleeding	Yes No	

In this example, the finding "Hemorrhoids" can not have "Yes" and "No" in different instances or multiple selection of attribute values. This is not explicitly stated in MST.

8. There were some erroneous items and our corrections are presented with boldface in Tables 3.16a, 3.16b, 3.17, 3.18a, 3.18b, 3.19, and 3.20.

Table 3.16a Problem related with "Normal" term in MST Table6-Terms for Esophagus

Terms	**Attributes**	**Attribute Values**	**Sites**
Normal	Z line	Distance in cm	**cm from incisors**

Problem*: redundancy in the attribute and attribute value.*

Table 3.16b Correction(s) related with "Normal" term in MST Table6-Terms for Esophagus

Terms	**Attributes**	**Attribute Values**	**Sites**
Normal	**Distance of Z line**	**cm from incisors**	

Solution: *descriptive attribute "Distance in cm" was merged with term "Z line" and attribute value "cm from incisors" was placed as an attribute value.*

Table 3.17 Problem related with "Varices" term in MST Table6-Terms for Esophagus

Terms	Attributes	Attribute Values	Sites
Varices	Red signs	Yes No	**Site(s)**

Problem: *The Site(s) is not associated with a term, attribute or attribute value.*

Solution: *The Site(s) was removed.*

Table 3.18a Problem related with "Food(residue)" term in MST Table7-Terms for Stomach

Terms	Attributes	Attribute Values	Sites
Food (residue)	**Type**	Specify if Bezoar present)	**Site(s)**

Problem**:** *Type attribute is not appropriate. Also Site(s) is also problematic because it is associated with an attribute not with a term.*

Table 3.18b Correction(s) related with "Food (residue)" term in MST Table7-Terms for Stomach

Terms	Attributes	Attribute Values	Sites
Food (residue)	**Bezoar present**	Specify	

Solution*: "Type" attribute is removed and "Bezoar present" from attribute value is placed as attribute for the term. "Specify" is left as attribute value and also Site(s) was removed.*

Table 3.19 Problem related with "Evidence of previous surgery" term in MST Table 9-Terms for Colon

Terms	Attributes	Attribute Values	Sites
Evidence of previous surgery	Suture material visible	Specify	
			Site(s)

Problem: *The Site(s) is not associated with a term, attribute or attribute value.*

Solution*: Site(s) was removed.*

Table 3.20 Problem related with "Normal" term in MST Table 11-Terms for Papilla Minor

Terms	Attributes	Attribute Values	Sites
Normal	**Site(s)**		

Problem: *The Site(s) is not placed in the appropriate column.*

Solution*: Site(s) was placed in the appropriate column.*

9. Mandatory attributes were not explicitly declared in MST which might be important for HIS functionality and also interoperability. Some attributes need to be specified for a term when describing a particular finding. For example the observation of bleeding or stigmata of bleeding might be necessary for valid description of a bleeding ulcer. Another example might be to specify the position of a stenosis from incisors in the esophagus. During archetype modeling we have captured this information by using clinical expertise.
10. Conditional existence of attributes and mutual exclusivity of both attributes and attribute values were also not declared in MST. By conditional existence we mean the dependence of certain attributes on some other attributes. For example in the stenosis of colon, the attribute "traversed" describing if the endoscopist was successful in passing through the narrowing and another optional attribute describing the length of the stenosis are both optional. However length alone can not be present if it was not possible to pass through the narrowing. So the latter attribute depends on the existence of the former. Mutually exclusive attributes can not coexist together; meaning certain attributes can not exist together with certain other attribute(s). An example is the impossible coexistence of "Bleeding" and "Stigmata of bleeding" in the same instance of the term "Tumour/Mass" for protruding lesions of duodenum. Same is true for attribute values when multiple selections of attribute values are allowed. For example the attribute values "Single", "Few" and "Multiple" are all mutually exclusive while the attribute values "Air bubbles", "Sludge", "Parasite" and "T-tube" for the attribute "Appearance" describing the term "Filling defect" in abnormalities of biliary system can coexist together. All these important information are incorporated into the Archetype model by using clinical expertise.

11. It was not stated at all whether a single attribute value or multiple values could be selected at a time in MST. This is extremely important especially for GUI design and consequently in data compatibility and interoperability with other HIS. We have also elicited this information by the help of domain experts and also our clinical experience.
12. There were some minor translation errors and typos in the Turkish translation of MST. They became apparent as the research prototype was used by domain users and were corrected.

3.1.2 Problems Identified Related with the Structure of MST

As shall be stated in Chapter 4, the original hierarchy of MST (Personal communication with the US MST editor Dr. Louis Korman on 13.05.2004) caused serious problems during modeling by Archetypes which is illustrated in Figure 3.1. So we have modified this and then validated in the case study by developing a research prototype. This modified structure resulted in more consistent model for HIS development especially for generating a user-friendly graphical user interface (GUI). The details of modeling of MST are mentioned in detail in Chapter 4.

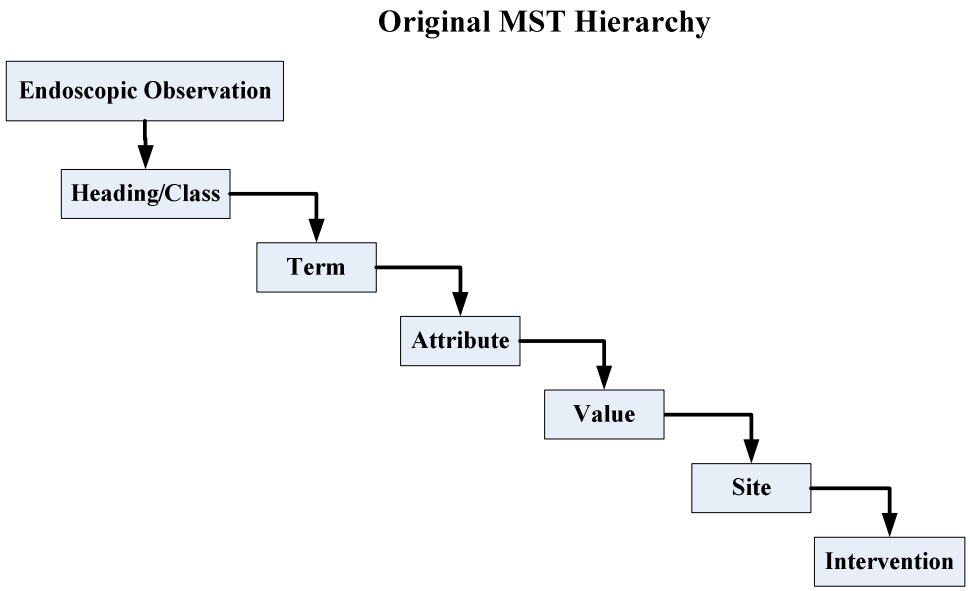

Figure 3.1 The original MST hierarchy (personal communication with Dr. Louis Korman)

3.2 Extensions to MST

In the original MST hierarchy, “Site(s)” was directly linked with the Attribute value. From an informatics point of view, we have decided that the Term was the real-world entity that could have Site(s) information; not the individual attributes which were just modifiers of that term. So we have repositioned “Site(s)” and linked with “Term”.

As we had foreseen that most of the user selection for a particular term would definitely involve multiple sites and that linking “Intervention” to “Site(s)” in the hierarchy would certainly diminish the expressive power of the model and user-friendliness of the GUI, we have extended MST and modeled endoscopic observation and intervention concepts separately.

MST content has also been extended by introducing whole new anatomic sites, sections, terms, attributes and mostly attribute values for better clinical coverage. These are:

1. A new anatomic site “***Anal canal***” was added to MST Table 2-Sites for location of findings in the lower gastrointestinal tract.
2. New diseases “***sclerosing cholangitis***” and “***biliary fistulas***” have been added to MST Table 18-Reasons for Performing ERCP.
3. A new attribute value “***heat-probe***” was added to attribute “Device” of the term “Thermal Therapy” under the heading “Therapeutic Procedures” in MST Table 14-Terms for additional diagnostic and therapeutic procedures.

Rest of the extensions are shown with bold and italic text in Tables 3.21 and 3.22 with original MST table references as they appear in the original publication (Delvaux, 2000b).

Table 3.21 Extensions to MST Terms for Esophagus, Stomach and Colon

Headings	**Terms**	**Attributes**	**Attribute values**	**Site(s)**
MST Table 6-Terms for Esophagus				
Lumen	Lower Esophageal Sphincter	Tone	***normal***	
Protruding lesions	Tumor/Mass	Type	***ulcero-vegetan***	
MST Table 7-Terms for Stomach				
	Rapid Urease Test	***Result***	***positive*** ***negative***	
MST Table 9-Terms for Colon				
Lumen	Evidence of previous surgery	Type	***ileo-anal pouch*** ***colo-rectal*** ***anastomosis***	
Flat lesions	Angioectasia			***Site(s)***
Protruding lesions	Hemorrhoids	**Type**	***internal*** ***external***	
		Grade	***grade I through IV***	

Table 3.22 Extensions to MST Diagnoses

MST Table 19-List of Esophageal Diagnoses	
Main diagnoses	**ectopic gastric mucosa**
Other diagnoses	**hypotonic lower esophageal sphincter**
MST Table 20-List of Stomach Diagnoses	
Main diagnoses	**pangastritis** **antral superficial gastropathy** **alkaline reflux gastropathy** **bulbitis**
Other diagnoses	**bulbus deformity** **stenosis**
MST Table 22-List of Colon Diagnosis	
Main diagnoses	**anal fissure** **stricture**
Other diagnoses	**suspicion of flat adenoma** **perianal abscess**

In addition the single colonoscopy examination type was split into colonoscopy and rectoscopy, because endoscopists felt the need to differentiate them even though the very same MST based data tables and forms were used for both.

3.3 GASTROS Case Study

The research question considered in this case study was to determine whether MST was valid and appropriate for use in gastrointestinal endoscopy HIS development.

The case study has started in the beginning of year 2000 and continued till August 2003. It has been conducted at Başkent University Hospital Endoscopy Unit in Ankara, Turkey. Key persons involved in the case study were Dr. Sedat Boyacıoğlu,

director of the unit and also general secretary of the Turkish Society of Gastroenterology responsible for translation of MST into Turkish and Dr. Gürden Gür who continuously monitored the usage of the system, and provided us with feedback.

Research prototype was installed in two workstations as standalone. During the case study period, 15,777 records were collected with a data file size of 21 megabytes. There was no down-time of the systems during this period.

After investigating the domain and examining the Turkish translation of MST, requirements were captured to evaluate MST for use in an information system. The most important goal at the beginning of the case study was to develop a research prototype application complying with the structure and semantics of MST. After designing the database and software by using classical Object Relational methodology, the prototype application GASTROS was implemented. It was a 32 bit Windows application programmed with Microsoft Visual Basic 6. Microsoft Access was used as low-cost relational database for data modeling and for persistence of the application data. Endoscopic images captured from video endoscope were also linked with records and stored in the file system.

The GASTROS database has a primary table (B Table) which contains fields like: automatically generated unique examination ID, patient demographics (name, surname, sex, age and etc.), clinical information (hepatitis/HIV markers and disease status), examination information (examination type, endoscopic device, premedication, dates and etc.), coded and free text main diagnoses and other information like link to referring department of the hospital, endoscopists' code, sign-out history and image status. Other related tables are linked to the main table via the unique examination number to store structured data conforming to MST. These tables are:

1. **KOLON**: Findings data for colon.
2. **ÖMD1**: Findings data for esophagus.
3. **ÖMD2**: Findings data for stomach.
4. **ÖMD3**: Findings data for duodenum.
5. **ERCP**: Findings data for pancreas, papilla major, papilla minor and biliary tree.
6. **GENEL3**: Examination characteristics, reasons and complications data for ERCP

7. **GENEL12**: Examination characteristics, reasons and complications data for upper and lower gastrointestinal examinations.
8. **EK**: Terms for additional diagnostic and therapeutic procedures.

Please refer to Figure 3.2 for relationship diagram of the database.

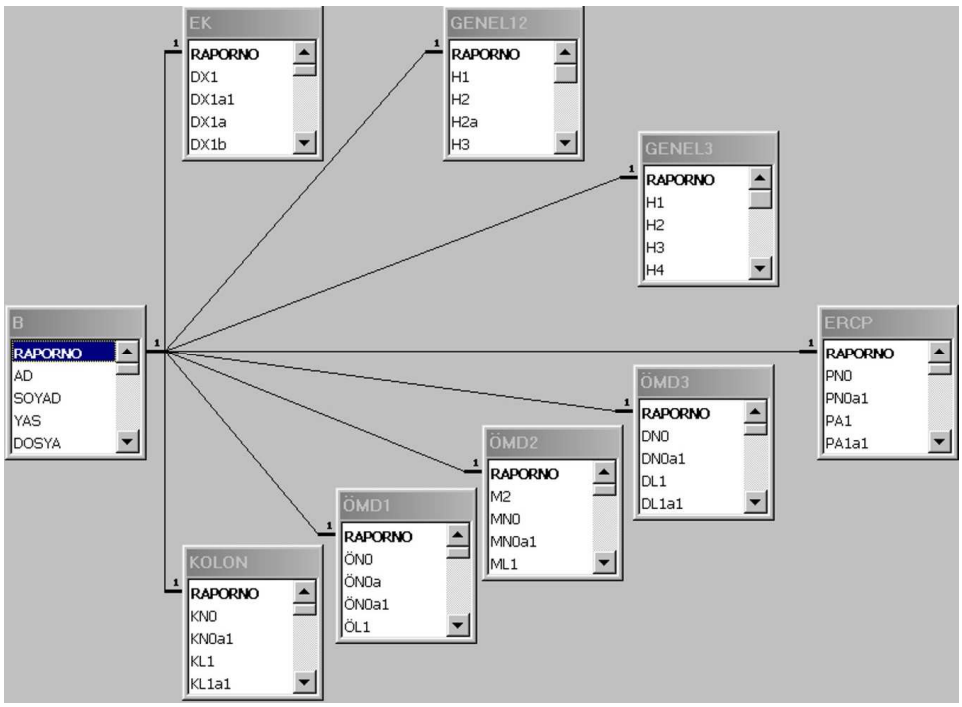

Figure 3.2 Database relationship diagram of primary table B

A system database file was also created which contains tables for storing attribute values for anatomic sites, reasons for examination, list of diagnoses for each organ as depicted in MST 2.0 as well as other tables containing operational data like list of referring departments, endoscopy devices, endoscopists' and fellows' names and titles, social security and medical insurance types and temporary tables used during automatic report generation.

MST hierarchy was represented in the database via specially encoding the table and field names with the following rules:

1) The name of the database table denoted "Examination type",
2) First character (uppercase letter) in a field name in a particular database table denoted "Organ",

3) Second character in a field name (uppercase letter) denoted MST “Heading or Class”,
4) Third character (consecutive positive integers) in a field name denoted MST “Terms” and fields with three characters were used to store the presence of a term (data type: Boolean),
5) Fourth character (consecutive lowercase letters in alphabetic order) in a field name denoted MST “Attributes” for a particular MST “Term” and fields with four characters were used to store enumerated MST “Attribute values” of this attribute (data type: Byte),
6) And the fifth character (positive integer) in a field name denoted MST “Site(s)” for a particular MST “Term” and fields with five characters were used to store bitwise computed values of multiple anatomic sites filtered for each organ (data type: Long Integer).

For example in the KOLON database table, the field “KL2” with Boolean data type (Yes/No) was mapped to the second term of the “Lumen” Heading/Class: “Stenosis” in MST. Its first attribute “Appearance” was mapped to the “KL2a” field, and the site data were mapped to the “KL2a1” field. All the attribute values with given lists were enumerated and these numeric values were stored in the database. In “Site(s)” fields, where multiple selections were possible and frequently used, each selected anatomic site was represented by a single bit and the resulting number was computed and stored in the database for achieving a smaller database size and increased performance.

Each workstation had its own database and application software installed; initial requirements did not force us to design a networked multi-user system. While designing the GUI, primary concern was user friendliness and simplicity as these are among key success factors for acceptability of computers and software by clinicians. We aimed for reduced time and effort during data entry and report generation. During automatic report generation, the structured data collected by SDE forms and data from free-text blocks were merged and formed a valid endoscopic report. In this process basic grammatical and syntactic rules of the Turkish language were employed; such as capitalization of the first letters of words following a dot or usage of appropriate suffixes after certain words.

The program has a simple main menu which allows users to select functions like Data Entry/Update, Search and Analysis, System Operations and Program Exit. Steps

needed for complete reporting of an endoscopic examination are listed below with some screenshots from application software GUIs (which are originally in Turkish but translated into English in figures showing screen captures):

1) Enter patient demographics, clinical and examination specific information (type, device, sedation, and etc.) using the general Data Entry/Update form (Figure 3.3),

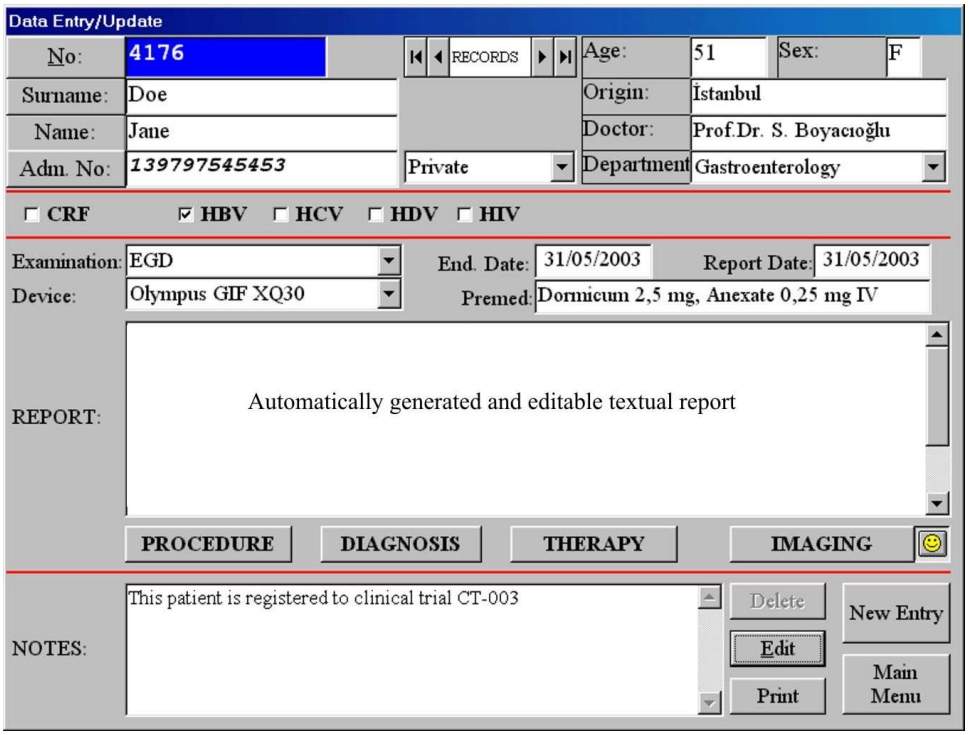

Figure 3.3 GASTROS main data entry form

2) Select examination type and use MST based SDE form for examination characteristics and reasons for examination,

3) Select organ and enter endoscopic findings using MST based SDE forms and select Site(s) (Figure 3.4),

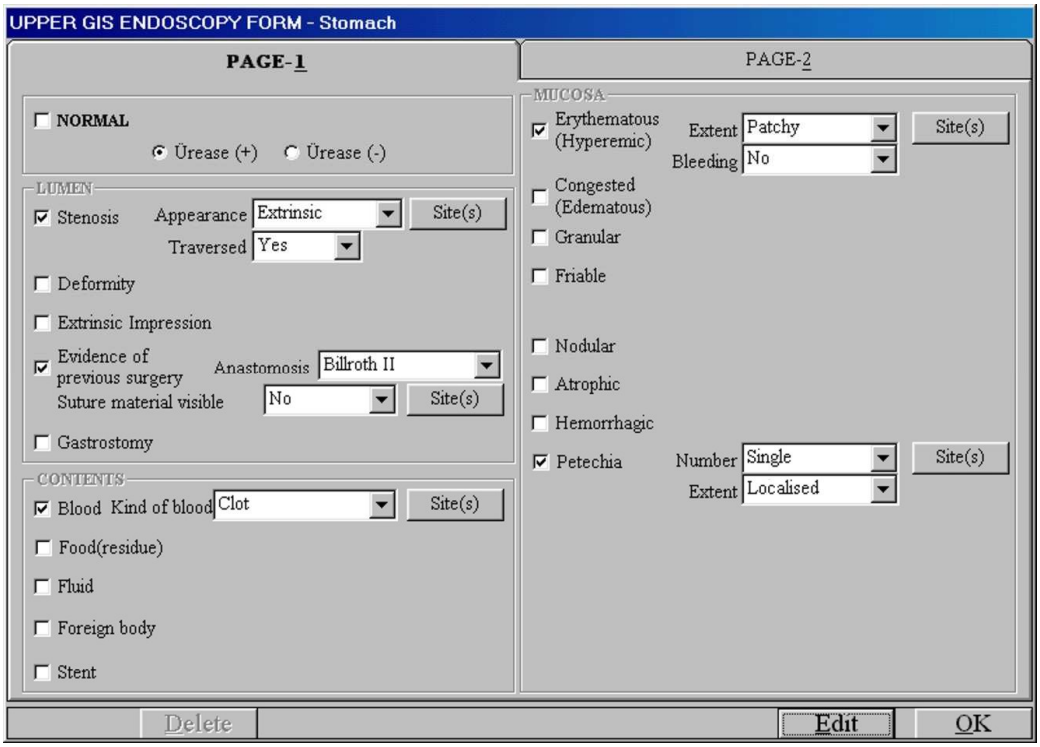

Figure 3.4 MST based SDE form for entry of findings in stomach during EGD

4) Enter additional diagnostic and therapeutic procedures,

5) Enter diagnosis and comments,

6) Generate the report and edit if necessary (the users are given freedom to edit automatically generated reports as necessary and add comments),

7) Select images to be associated with examination and further mark for printing,

8) Generate final endoscopic report, sign-out and print.

The very comprehensive yet easy to use search and analysis functions can be performed in four categories:

1) **General Search**: Mainly used to find patients' previous examinations via surname and/or name or hospital admission number. A date or a date range can also be specified in search.
2) **Diagnosis Search**: Structured queries can be designed with user friendly and easy to use interface by first selecting an organ and an associated diagnosis from a

dropdown combo-box which is automatically filled with list of MST 2.0 diagnoses. It is also possible to search multiple diagnoses of the same or a different organ by combining multiple lines with Boolean operators (AND, OR). These search criteria can also be combined with the Advanced Search criteria to be able to find records having certain diagnoses and other attributes such as sex, age or any other field present in the main table (Figure 3.5).

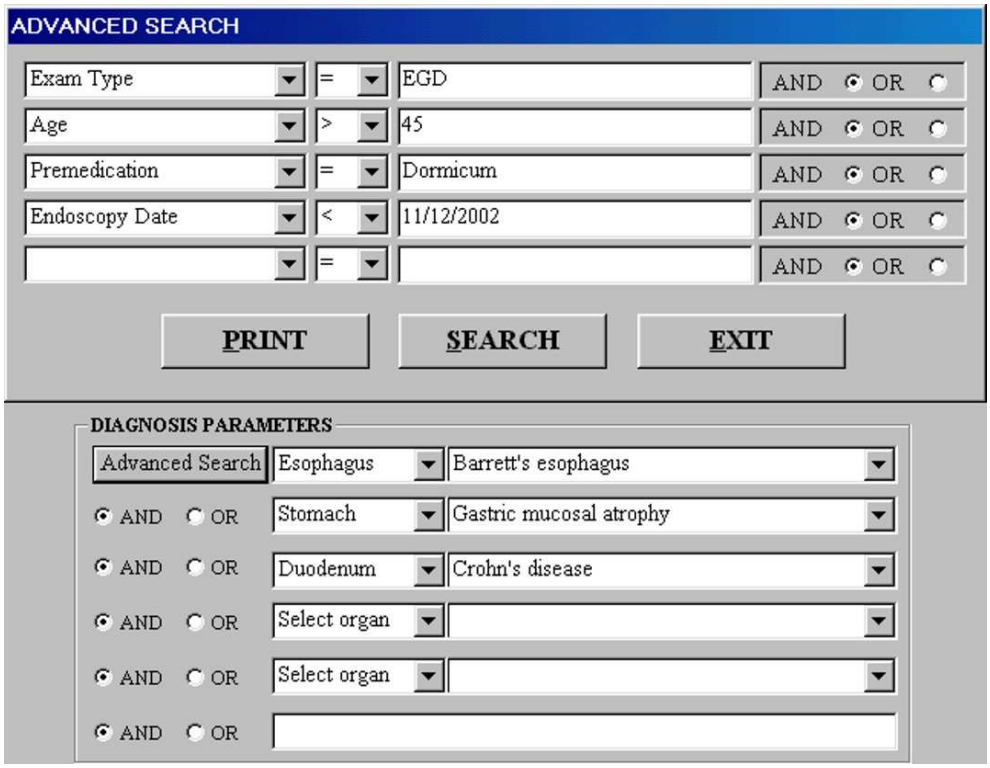

Figure 3.5 Diagnosis Search form combined with advanced search options

3) **Procedure Search**: This is a very powerful feature of the application software that it enables users to use the very same SDE forms during data entry during the search. They simply point and click as they would have entered the case and then search the database.
4) **Advanced Search**: All fields in the main table can be searched by combining with Boolean operators by a very simple interface.

These search functions enable users to query both the free text parts of the examination data and the structured MST data. Boolean operators can be used to create complex queries by using a very simple interface. A powerful feature of GASTROS is that the same but empty SDE forms are used for dynamically building the search criteria by a simple point and click operation.

Due to the fact that GASTROS was used as the only means to generate endoscopic reports at the unit by all endoscopists, including emergency cases, no field other than the unique examination number was mandatory to fill.

MST and other extra information (such as patient demographics, clinical info, infection markers and so on) were directly incorporated into application code, graphical user interface (GUI) and database schema. After initial installation in a university hospital endoscopy unit, the system was used to capture more requirements and also demonstrated ways of modeling MST useful for HIS development and aid in designing a user friendly and robust GUI. It was evident that the physicians preferred our novel modeling strategy over the original MST hierarchy. As the requirements frequently changed, it was obvious that it took very long time and serious effort to modify the software and redeploy again (Liebowitz, 1999). Another shortcoming was that nearly all of these modifications resulted changes in the database so an extra task of converting data was needed. This clearly showed that the system was not feasible to maintain - not only economically (Dairo, Giuse and Kuhn, 2003). However it was an invaluable tool to refine initial requirements. We have managed to collect more than 15,000 endoscopy records, analyze and evaluate them (Atalağ, et al., 2007b).

The bottomline is that this prototype was invaluable for experimenting with alternative modeling strategies of MST to identify best suited one to implement a usable endoscopic information system. However it should also be noted that it has shown us what shouldn’t be done when designing and implementing such systems from a technical point of view (i.e. embedding highly complex clinical model into software code and database schema). It became evident that adaptive and future-proof IS development methodologies were needed to keep up with ever changing requirements.

3.4 Validation of the Turkish Translation of MST

Considering many unsuccessful attempts for terminology standardization in clinical medicine and efforts for use in computerized systems we have decided to first test the clinical usability and acceptance of MST before proceeding with the study. This validation process with the research prototype not only provided us with invaluable experience with MST but also gave us considerable insight for domain modeling. This had set forth the principles of Archetype modeling in further stages of the study.

The validation study was performed in a live clinical setting by using the research prototype GASTROS (Atalağ, et al., 2007b). It consists of first inspection of the collected data, then data analysis and interpretation. The usage of MST terms was observed and user acceptance was measured to assess the validity of MST and the research prototype. There was no selection bias in the validation process and after data cleaning 15,638 records were analyzed which were officially signed out and printed as valid endoscopy reports.

It should also be noted that we have also identified alternative MST hierarchies and then tested if useful for building a usable HIS. Although this is not a formal usability study by any means, it was quite invaluable during both prototype development and also in further stages during MST Archetype modeling.

3.4.1 Data Preparation

Before data analysis, the data in the two separate workstations were checked for consistency and some erroneous records with duplicates or null entries were discarded (i.e. only examination numbers were assigned but remaining fields were empty). The data from the two workstations were then consolidated.

3.4.2 Data Analysis

For data analysis Structured Query Language (SQL) statements were created using Microsoft Access. For example to determine whether MST based diagnostic terms had been used, corresponding field values were checked whether they were greater than zero, the default value for a newly added record. For determination of usage of free

text in fields allowing both free-text and MST terms, their values were checked whether they were null or empty (i.e. deleted later on). The discrimination between the missing values (null) and the zero values (empty or deleted) was accomplished by this way.

General distribution of records according to examination type was as follows: 11,381 (72%) EGD, 2,616 (17%) colonoscopy, 1,079 (7%) rectoscopy, and 562 (4%) ERCP. Overall usage of MST for recording examination characteristics (extent and limitation) was 85% (13,322 of 15,638 records). When we look at each examination type, 10,277 of 11,381 (90%) EGD, 2,369 of 2,616 (91%) colonoscopy, 624 of 1,079 (58%) rectoscopy and 52 of 562 (9%) ERCP records had at least one MST based entry for examination characteristics. Reasons for endoscopy were recorded by using MST terms in total of 346 (2.21%) records. Their distribution according to examination type was: 261 of 11,381 (2.29%) EGD, 61 of 2,616 (2.33%) colonoscopy, 5 of 1,079 (0.46%) rectoscopy and 19 of 562 (3.38%) ERCP records. These usage data were determined by building SQL queries joining primary database table B with GENEL12 and GENEL3. After running queries, records having non-null entries were calculated for each examination type.

The usage of MST terms for description of endoscopic findings is given in Table 3.23. We had determined the usage by counting number of valid entries in database tables that were linked to the primary database table B. Therefore the number of records of a particular examination type in the primary database table B may be different than (equal or greater) the number of entries recorded in the related SDE database tables due to records with no MST entries.

Table 3.23 Overall and detailed usage of MST terms for recording of endoscopic findings by examination type and organ

Exam Type and Organ	Total No. of Exams	MST Usage
EGD-Total	11381	11216(98.55%)
EGD-Esophagus		11210(98.50%)
EGD-Stomach		11199(98.40%)
EGD-Duodenum		11167(98.12%)
Colonoscopy-Colon	2616	2471(94.46%)
Rectoscopy-Colon	1079	751(69.60%)
ERCP-Total	562	258(45.91%)
ERCP-Duodenum		242(43.06%)
ERCP-Other Organs		250(44.48%)
Overall usage	15638	14696(93.98%)

The comparative usage of MST terms and free text for recording of endoscopic diagnoses is given in Table 3.24. The usage was determined by analyzing both the fields which contained enumerated MST diagnostic terms and also free text fields for each record.

Table 3.24 Comparative overall usage of MST terms and free text fields for recording of endoscopic diagnoses

No. Of Exams	MST Terms(+)	MST Terms(-)	Sub Totals
Free Text(+)	4788(30.62%)	719(4.60%)	5507(35.22%)
Free Text(-)	9911(63.38%)	220(1.40%)	10131(64.78%)
Sub Totals	14699(94.00%)	939(6.00%)	15638(100.00%)

Free Text (+/-): Free text was used for diagnosis or not

MST Terms (+/-): At least one MST Term was used or not

Further data analysis on MST diagnoses is given in Table 3.25 including the frequency of normal cases and top three diagnoses by examination type and organ. Presentation of these results follows the same structure of the publication of European Union framework project GASTER (Delvaux, et al., 2000a).

Table 3.25 Frequency of the use of MST diagnostic terms by examination type and organ

Exam Type and Organ	MST Diagnosis	No. of entries	% of entries	% of exams
EGD-Esophagus (Total terms: 11905)	Normal	6840	57.45	43.74
	Reflux esophagitis	2222	18.66	14.21
	Hiatus hernia	997	8.37	6.38
	Hypotonic LES*	996	8.37	6.37
EGD-Stomach (Total terms: 14177)	Normal	525	3.70	3.36
	Antral superficial gastritis*	2871	20.25	18.36
	Erythematous (hyperemic) gastropathy	2383	16.81	15.24
	Pangastritis*	1842	12.99	11.78
EGD-Duodenum (Total terms: 11924)	Normal	6182	51.85	39.53
	Bulbitis*	2642	22.16	16.89
	Duodenal ulcer	992	8.32	6.34
	Erosive duodenopathy	634	5.32	4.05
Colonoscopy-Colon (Total terms: 2828)	Normal	539	19.06	3.45
	Hemorrhoids	838	29.63	5.36
	Polyp	496	17.54	3.17
	Diverticulosis	322	11.39	2.06
Rectoscopy-Colon (Total terms: 1146)	Normal	51	4.45	0.33
	Hemorrhoids	828	72.25	5.29
	Anal fissure*	211	18.41	1.35
	Fistula	17	1.48	0.11
ERCP-Duodenum	Normal	1	100.00	0.01
ERCP-Biliary System (Total terms: 402)	[Normal: cholangiography, Post-sphincterectomy , Post-cholecystectomy]	[69,9,2]	19.90	0.51
	Choledocholithiasis	165	41.04	1.06
	Cholelithiasis	59	14.68	0.38
	Bile leak	13	3.23	0.08
ERCP-Pancreas (Total terms: 167)	Normal	150	89.82	0.96
	Chronic pancreatitis	8	4.79	0.05
	[Pancreatic tumor, Failed pancreaticogram]	3	1.80	0.02
	Pancreas divisum	2	1.20	0.01
Total MST Diagnoses	**42550**			

Overall usage of MST terms for additional diagnostic and therapeutic procedures was 19% (2,953 of 15,638 records). For each examination type the figures were as follows: 2,489 of 11,381 (22%) EGD, 315 of 2,616 (12%) colonoscopy, 62 of 1,079 (6%) rectoscopy and 87 of 562 (15%) ERCP records. These usage data were determined by calculating the number of records with non-null entries by building SQL queries joining primary database table B and database table EK.

7,476 (48%) female subjects versus 6,163 (39%) male subjects were present and 1,999 (13%) records had null values in sex field. Numbers of records for some fields with missing values were: 1,522 (10%) age, 1,383 (9%) premedication details and 15,161 (97%) patient origin.

3.4.3 Interpretation of the Results

This is among the first evaluation studies of second version of MST to our knowledge. We believe that high coverage rate of the Turkish translation of MST for reporting endoscopic examinations in a university hospital endoscopy unit is a strong point for the validation of the terminology. It is important to note that the high usage rate of MST based SDE forms purely resulted from user acceptance as no field was obligatory to fill and it was possible to write free text in final report. Some sort of software control measures (i.e. warnings, compulsory fields) might have been applied because high rate of missing values in fields like age, sex, patient origin, and clinical information because it is responsible for diminishing data quality. However in this study these missing data were recorded in the central hospital information system. This may be an explanation for the high number of missing values. Likewise, “Reasons for endoscopy” were also recorded in extremely low rates. This might be due to organizational preferences or problems with GASTROS. Examination of free text entries for endoscopic diagnoses revealed that they were mostly used for additional notes regarding the technical aspects of the study or success of the procedure which should normally have appeared elsewhere in the report. There were also high numbers of repeating diagnoses like “hypotonic LES” which were later added to the pick list. However, it is evident that further work on MST is needed for ERCP for better coverage because as inline with previous studies, usage was quite low compared to other examination types (Delvaux, et al., 2000a).

Normal cases were relatively few especially in EGD examinations. We believe that this might have resulted from three reasons:

1) Prevalence of H. Pylori in Turkey is believed to be very high which decreases the number of “normal” reports,
2) Endoscopists prefer not to give too many normal reports (False positives are better)
3) It is not crystal clear what is “normal” and “not normal” in clinical medicine. Also as another factor, diagnosis of duodenum was not routinely included in ERCP reports at the unit, in only one out of 562 ERCP studies, a MST based diagnostic term (Normal) was selected for the duodenum.

A major weakness of GASTROS was its inability to allow selection of a MST term with different set of attributes or attribute values more than once. For example if the endoscopist had observed two different kinds of polyps in colon, each having different attributes and possibly site data, it was possible to record only one. This was one of the major reasons for free-text editing of the final report. However in MST it is not given explicitly whether a term or its attribute(s) must be mandatory (existence), how many attributes terms can contain (cardinality) and the number of times they can occur (occurrence). Therefore we have incorporated this information in our Archetype models and we strongly suggest this information to be introduced into future versions of MST.

CHAPTER 4

MULTI-LEVEL MODELING AND ARCHETYPES

In Chapter 2 we have extensively reviewed pertinent literature on health, information systems and also health informatics fields. This review was centered on the objectives of this study and delineated the position and scope of our research. Chapter 3 described in detail the gastrointestinal endoscopy domain and also discussed many aspects of its standardized terminology MST. Initial classical modeling and its implementation as a research prototype gave use valuable insight for Archetype modeling work.

In this chapter we will first discuss the rationale in the search for a modeling strategy that would enable HIS development and maintenance process to be performed more efficiently with less effort and cost, and yet result in systems with sufficient interoperability, flexibility and longevity. openEHR Archetypes and Two-level Modeling methodology has been identified as the modeling methodology for the study after comparison with potential alternatives. We will explore this new approach for use in developing information systems for niche clinical domains by extensively making use of standardized terminologies and ontologies. Within the scope of this study, our clinical domain for modeling has been selected as gastrointestinal endoscopy. This chapter is devoted to the discussion of our contributions towards Archetypes and multi-level modeling.

The knowledge represented in the standardized terminology MST has been modeled with the methodology described in this chapter. We had initially developed a research prototype by using classical Object Relational method to test validity and usability of MST structure and terms. That work is explained as a case study in Chapter 3.

The results of the current study are two-fold: first, our work made significant contributions to the MST which can be applied in other clinical domains. Second, the modeling methodology originally for use in domain concept modeling in EHR systems, was further extended to be able to effectively model a whole clinical domain and fully express the structure and semantics of MST. It is also demonstrated that this methodology provides the means to extend MST for local needs in a feasible way.

4.1 The Modeling Paradigm

Today IS development is a well established discipline with rigorous engineering methods. The requirements elicitation and design phases of development are extremely important because they determine what system is needed and how to build it. Most of the errors in development process are attributed to these phases and they are costly to correct. The modeling comes into play in these two pre-implementation phases to aid in the identification and understanding of domain issues. It also serves as an effective way of communicating these abstract entities among different groups. The UML, for example, may be used throughout the entire development process, from requirements to design and then implementation, and it is even possible to transform manually or automatically design artifacts into program code. Today we can not think about IS development without modeling.

When we look at HIS development, as we had mentioned extensively in Chapter 2, it has to deal with the complexity and changeability of the medical domain and also with the essential difficulties of IS development. So far, according to literature and the author's personal experience over more than a decade, the classical modeling and development techniques have fallen short of expectations. This is especially evident in HIS development for niche clinical domains where the depth of domain knowledge is greater and this highly volatile form of knowledge influences most of the software specification.

The design process of IS development usually starts with data modeling after requirements are elicited. It is assumed that these data requirements are fairly stable and this is the first incorrect assumption in classical development in healthcare domain. Then shortly after or in parallel, object oriented analysis yields abstraction of the domain, at a level sufficient for the system to be developed, a conceptual model of the IS itself as well as their relationships and dependencies. The second mistake is assuming the domain will stay unchanged, at least for the lifetime of the particular project, and hard-coding the basic assumptions into the software model. This occurs for two main reasons. First, the level of abstraction is not sufficient from the outset which may result in very fine details manifesting themselves in upper levels of the design, such as the trend of a tumor marker on the treatment scheme. Second, it is not uncommon that the modeled domain changes altogether or a new interrelated domain comes into play; such as genomic science, genetic treatments or new diagnostic modalities like MRI, PET or SPECT. When these happen, and they do happen, the consequences are major: alteration of data model and software model, revising documentation, retesting and redeployment.

In the above scenario, rock solid domain concepts like the details of a classical doctor-patient encounter, the steps of physical exam, audit trails or even clinical data structures and types are mixed with concepts like the Glasgow coma scale, Forrest bleeding criteria or even description of a naked-eye dermatologic examination. The situation is also the same with the administrative data derived from medical data, such as billing codes related with procedures or even worse with both diagnosis and procedures plus other parameters like length of stay. This is the third and biggest mistake during the whole development process.

One method for improvement in handling of rapidly changing domain concepts is using a generic data model, such as the Entity-Attribute-Value scheme, where regardless of their meaning or context all data are stored in small number of tables in the database. This overcomes the problems of data modeling but with a price: structured queries are difficult to build as is analyzing data and maintaining data integrity. However the problem of software modeling to reflect changes and then redevelopment still exist.

This thesis is centered on finding better methodologies to overcome the problem of changing requirements and its negative effects on HIS development. We propose

improvements in the early phases of development process and we address the issue of modeling as the target of research.

4.2 Alternative Modeling Formalisms

Based on pertinent literature and also taking into consideration similar research and development projects, we have decided to evaluate openSDE, Protégé, openEHR and HL7 v3 specifically for the purpose of this study.

4.2.1 Evaluation of openSDE

openSDE (Los, van Ginneken, and van der Lei, 2005; Los, et al., 2004; OpenSDE Website, 2007) which allows for Structured Data Entry (SDE) in clinical medicine is an excellent tool for providing flexibility to clinicians to record what and how they may want to. While doing that, it also hides the complexity and changeability of the underlying clinical model from HIS developers. openSDE functionality is based on a domain model, a knowledge base, which defines the terms for concepts, their structure and semantics to be presented in a particular context. A visual data entry application then uses this domain model and user input to dynamically generate options for data entry. The overall components are given in Figure 4.1.

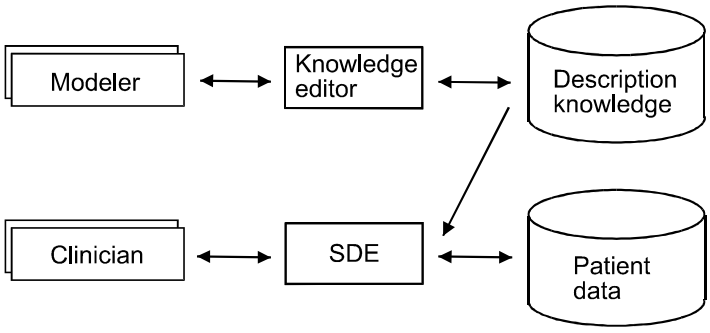

Figure 4.1 The components and relationships of openSDE

We have observed the main innovation of openSDE as the separation of knowledge from the technical layer; thus providing means for tackling classical problems of HIS.

This knowledge base mainly consists of what they call as the description knowledge which defines which terms can be used in what combinations to form medically meaningful descriptions. It should be noted that openSDE only focuses on the capture of patient data that the clinicians usually record by handwriting or dictation. These typically involve patient history, family history, physical examination, progress notes, and examination reports such as radiology, endoscopy, and pathology.

One big advantage of openSDE for modeling of gastrointestinal endoscopy was that the team had already experience with this domain. However our request for getting more information or artifacts from this study was unfortunately not met later on.

The core of the domain model consists of medical concepts, which may be described by an external terminology, used for data entry. In openSDE these concepts are organized as nodes in a tree structure. In this tree, nodes may have further sub-tree(s) for allowing a more detailed description of a finding. The tree can also accommodate certain information constraints for the presentation of data entry options. The openSDE viewer allows for traversing the tree of medical concepts and selecting those nodes which correspond to medical observations to be recorded. The trees within a particular domain model are specific for that domain.

The two main tools of openSDE, the domain model editor and the data entry application, are versatile and easy to use. However as provided from its Website and the open source portal, the documentation is very limited and not very intuitive. Due to the unavailability of sufficient documentation and also lack of formal support, we have decided first to use the sample domain model for our evaluation which is cardiology. However as we grasped the use of editor and domain modeling, small fragments of MST like the list of diagnoses and findings, were modeled by the editor. A segment of the cardiology domain model as appears in the domain model editor is given in Figure 4.2.

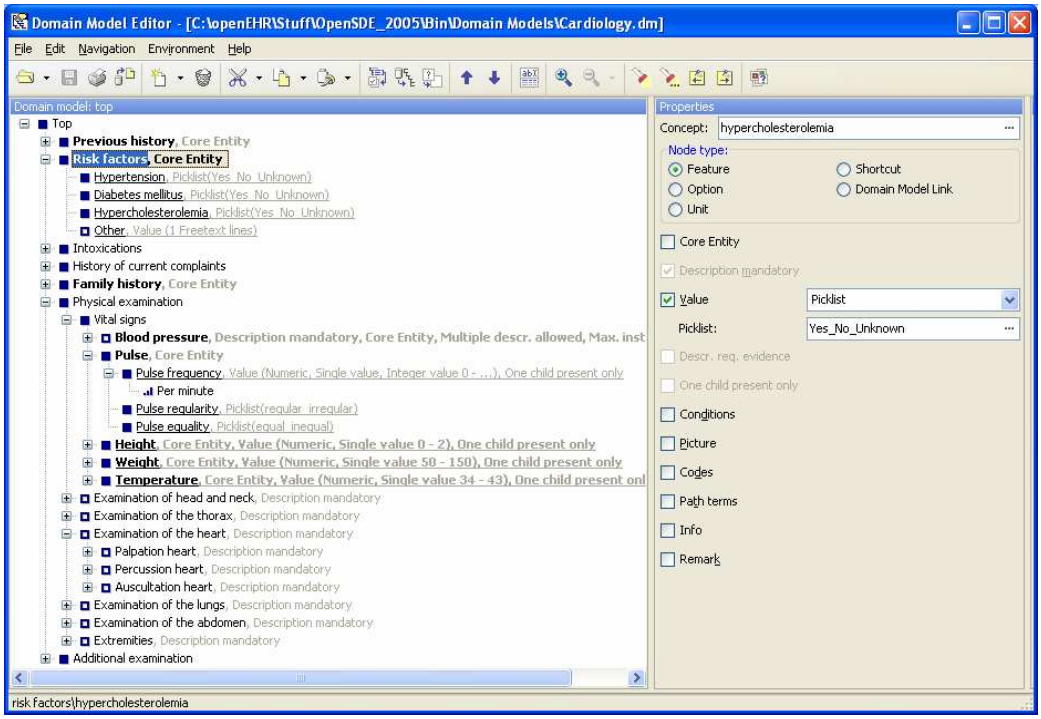

Figure 4.2 Domain Model Editor of openSDE (Screenshot from the 2005 release of openSDE by the author showing the sample Cardiology domain model)

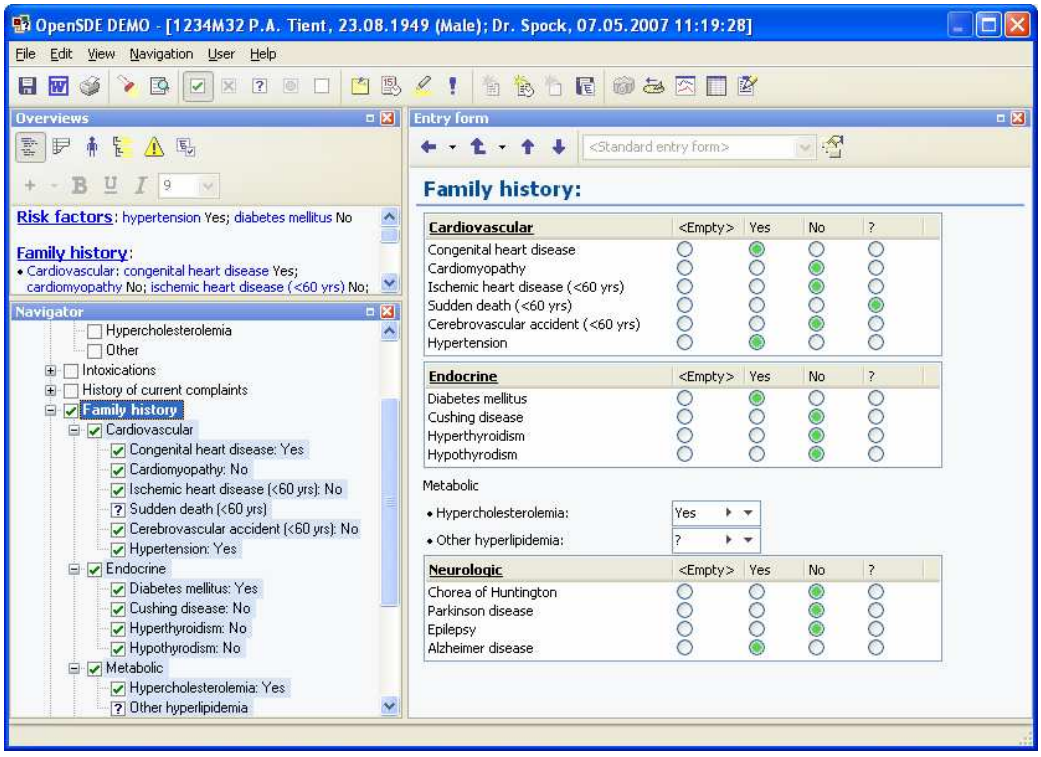

Figure 4.3 Data Entry Application of openSDE (Screenshot from the 2005 release of openSDE by the author showing GUI of the sample Cardiology domain model)

Depending on the structure of trees and node attributes, the data entry application creates the GUI automatically as shown in Figure 4.3.

A wrapper application to enter patient demographics and encounter is also provided to simulate the use in a real HIS; this was important for our evaluation with minimal development effort. One of the advantages of openSDE is the availability of a persistence component as a DLL which could be used to interface with a database management system of choice when integrating with HIS.

If we summarize the advantages of openSDE for the purpose of this study:

- It is a mature product with field testing,
- It is open source (readily accessible) and academic project,
- Separation of domain knowledge from underlying information and data models to tackle complexity and changeability of HIS is a big plus,

However we have identified a number of disadvantages which resulted in exclusion from further evaluation in the study.

- It provides just applications, not a framework for modeling HIS that can be implemented using different approaches and technologies,
- The modeling is limited only to the free text fields to be replaced with structured data fields,
- Other parts of HIS like demographics, security, EHR structure can not be modeled,
- Underlying data structures and types are propriety; they do not comply with the recent EHR standards,
- Dependency on a small team with limited support is a big shortcoming.

4.2.2 Evaluation of Protégé

Although designed for a quite different purpose, authoring ontologies and building knowledge based systems with decision support; we have considered evaluating Protégé (Noy, Fergerson, and Musen, 2000; Protégé Website, 2007). The reason is three fold: first, defining ontology very broadly as a common vocabulary for people who need to share information in a domain, then MST can be seen as an ontology given its structure and constraints for concepts. Second, it can separate domain knowledge from the operational knowledge and enables reuse of domain knowledge.

Third, it provides a basic data entry interface which is created from the underlying knowledge model.

Since we have accepted that MST is little more than just a simple terminology, we started modeling MST Colon findings. However the classical Object Oriented (OO) modeling approach turned out to be very difficult in modeling. Since MST Findings are organized in at least five levels of tree nodes, current structure of Protégé with Class, Attribute (slot) and Attribute Values and Objects (instances) created an "impedance mismatch". The MST Colon ontology is shown in Figure 4.4.

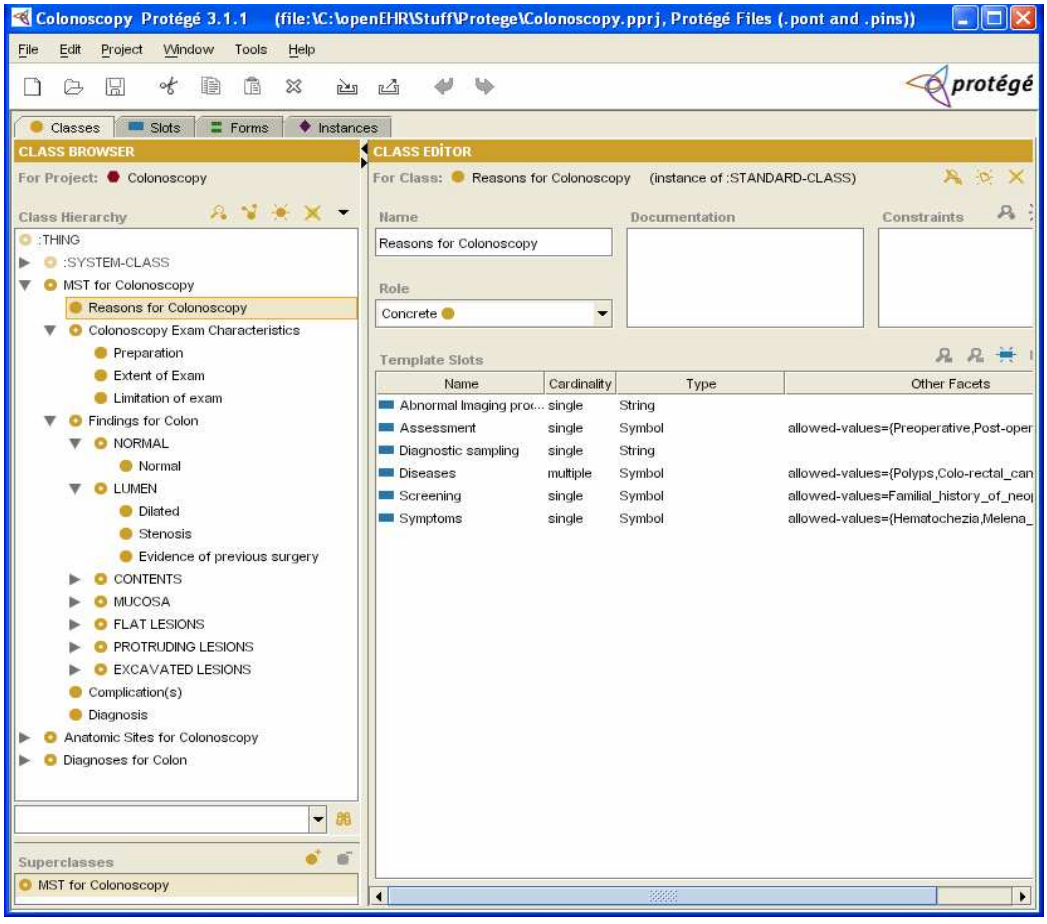

Figure 4.4 The ontology model for MST Colon Findings (Screenshot from the Protégé by the author showing draft endoscopy domain model)

At first sight, modeling an ontology may seem similar to domain modeling for designing real HIS. However, Protégé is designed to model for certain purposes, such

as knowledge base development, decision support and generic data entry. Therefore we have observed limited expressive power while modeling our domain.

The pros and cons of using Protégé are very similar to openSDE, therefore it was excluded from further evaluation in the study.

4.2.3 Health Level 7 (HL7) Version 3

Considering that our modeling work had started in year 2003, HL7 v3 was not available by then. In late 2005, the HL7 community released the first version of HL7 v3 ─ the Normative Edition 2005. Later on in mid-2006, the Normative Edition 2006 was published (HL7, 2007). Although we had selected our method for modeling as openEHR, HL7 v3 was later evaluated for the purpose of the study.

The RIM based message building and interactions enable certain level of functional and semantic coherence. However the overtly complicated process of creating messages together with formally reported shortcomings in the RIM caused it to be excluded from further evaluation. The main body of RIM is shown in Figure 4.5.

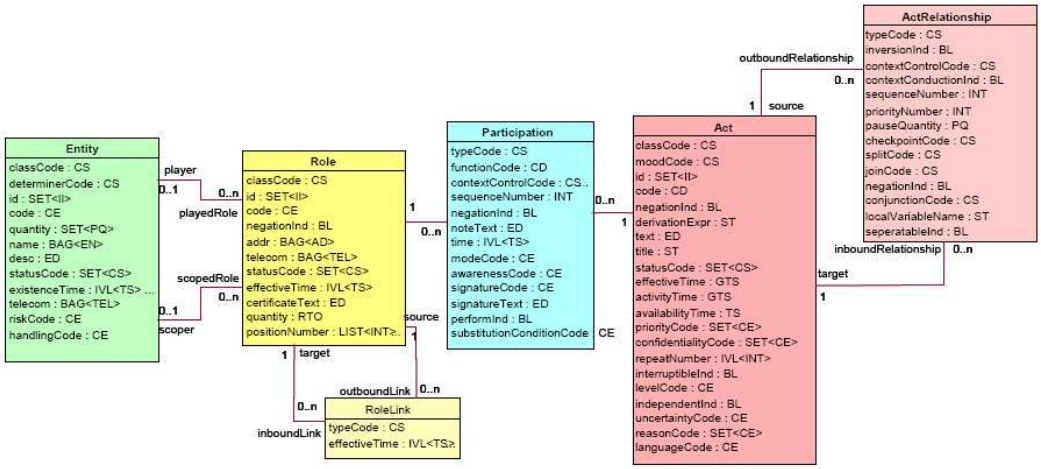

Figure 4.5 UML class diagram of context-level HL7 v3 RIM (HL7, 2007; HL7 RIM, 2007)

One big disadvantage with an unstable RIM is that it brings about a big burden during the maintenance phase of HIS because most of the software has already been specified by the old RIM. Another big disadvantage of HL7 v3 to be used as a modeling methodology for HIS is the proven insufficiency of the messaging paradigm for later establishing a federated virtual EHR.

4.2.4 openEHR Multi-level Modeling and Archetypes

In search for a robust methodology for modeling a variety of clinical domains and yet meeting aforementioned expectations for developing HIS, a strong candidate was openEHR Archetypes and Multi-level Modeling approach rapidly evolving with support from a large body of internationally recognized experts and materializing under the newly established openEHR Foundation (Beale, 2000, 2002; openEHR Foundation, 2007).

The pros and cons of this methodology are as follows:

The approach is a novel innovation and a paradigm-shift in the design of HIS. Even though theoretical work and practical applications do exist in other sectors, neither to our knowledge nor an in-depth analysis of related literature yielded a comprehensive approach. As in some other evaluated alternatives, it also tackles the problems of domain complexity and requirements changeability by separating data, information and knowledge levels in software. Like openSDE, this also results in the separation of the tasks of domain experts and technical people which results in more efficient development process and empowers non-technical people. A schematic representation of this is given in Figure 4.6.

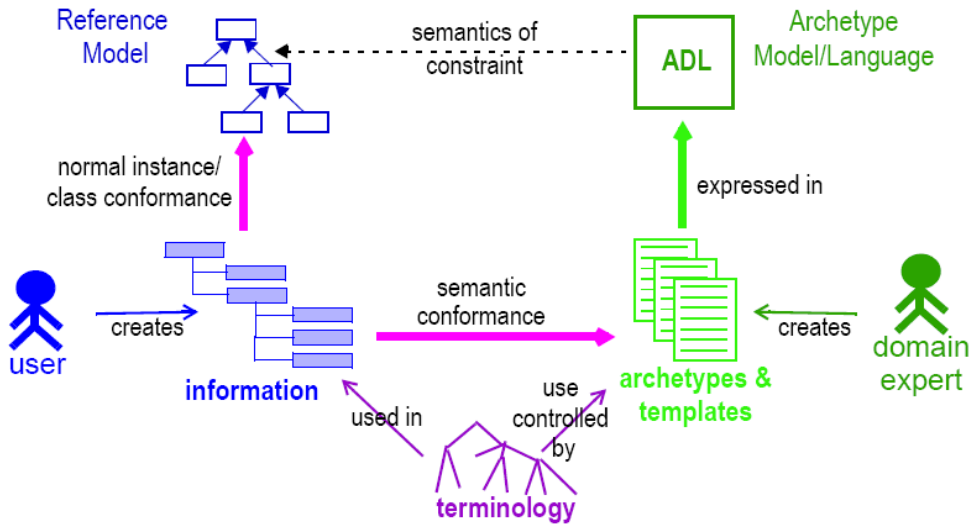

Figure 4.6 Separation of information and knowledge which also separates the tasks (Beale and Heard, 2006 - with permission from Thomas Beale)

The separation of information from knowledge is a difficult task and one needs a solid methodology to do so. The use of RM and Archetypes helps to achieve this. At the information level RM depicts all the necessary informational entities, their relationships and also semantic rules (occurrences, cardinalities) that make it relevant to the domain of interest. Interestingly, the data modeling process which is mostly the starting point of classical IS development, is no longer needed because here all data are explicitly instances of RM. The consensus-based structure and semantics of openEHR RM which is an accumulation of nearly three decades of relevant research and projects is a very strong point. As of the writing of this thesis, with release 1.0.1, it is also considered quite stable that now many researchers and developers Worldwide are implementing it.

At the knowledge level, a novel concept called Archetype is authored by the domain specific modeling and knowledge representation language, Archetype Definition Language (ADL), which allows for formally describing clinical concepts. The real power of Archetype comes from the fact that, while describing healthcare specific aspects of the concept and its meta-data, it also formally defines the structure, semantics and rules of the concept by using RM entities as building-blocks and further constrains the general healthcare related constraints expressed in relevant RM. An analog is given in Figure 4.7.

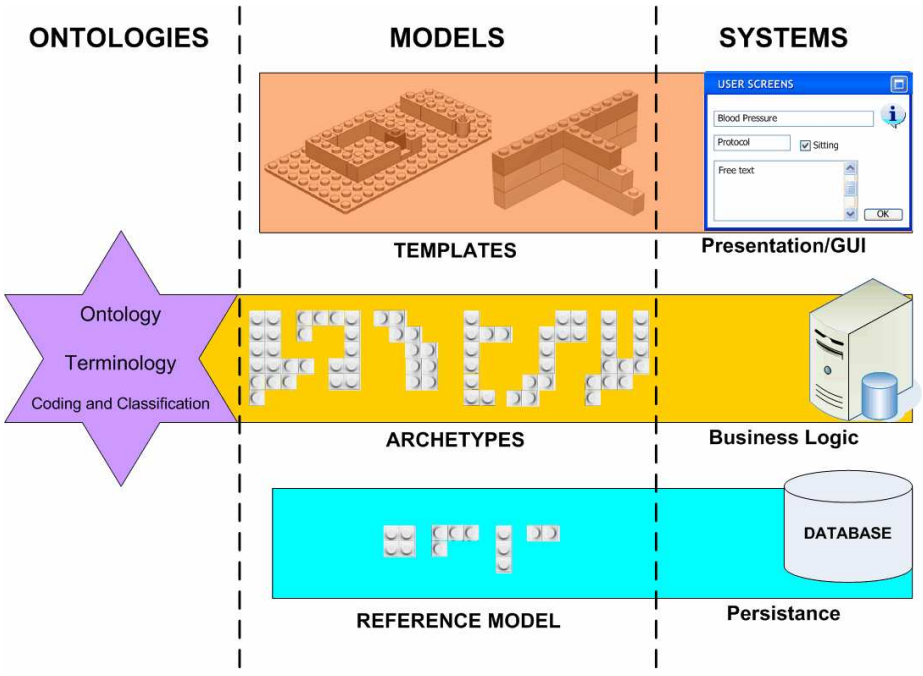

Figure 4.7 The Multi-level Modeling paradigm (Atalag and Bilgen 2007a)

Archetypes have their local terminology, that is regardless of whether external terminologies are available or not, it is still possible to express any medical concept. In the case of terminologies with hundreds of thousands of terms, such as SNOMED, when only a few terms are to be used it is a good alternative to have them as internal Archetype terms. However when there is need to utilize a common terminology system, such as bounded by a national or organizational policy, local Archetype terms can be mapped to any number of external terminologies directly or via a terminology service.

The ontology section of Archetypes allows for multilingual representation of internal terms which bring about language independence of the concept being modeled. In our domain, since MST is available in 11 languages, this is an extremely important aspect.

When well agreed medical concepts, such as blood pressure or a hematology result, are modeled by Archetypes and these Archetypes are used thoroughly by HIS it is possible to achieve a high level of semantic interoperability. However while modeling less common and specific medical concepts, it is very likely that similar but different models might already exist or the need to alter these models emerges. Archetypes allow the freedom to make changes to meet local user needs without breaking the original semantics by a method called as Archetype Specialization. Here, modifications that would result in more strict constraints than the original or additions are allowed. Since MST is likely to be extended (added) it is an extremely important point for our study.

Archetype repositories will help to establish governance rules for important medical concepts likely to be reused globally. MST, which is the official terminology for gastrointestinal endoscopy Worldwide, needs such a registry to be properly change managed. The multi-axial archetype identifications and embedded key meta-data like "author", "date", "languages", "specialization" and so on are important points to achieve proper dissemination of these models which will help in a high-level of interoperability among humans and HIS.

The recent acceptance of Archetypes and the separation of information and knowledge explicitly by other standards developing organizations, like HL7 and CEN, is a definite sign of universal acceptance of this approach (Munoz, et al. 2007). This alone is an affirmative reason for preference of openEHR in this study.

The biggest disadvantage of this methodology was its immaturity and incompleteness at the time we had started the modeling process at the end of year 2002. There were no tools and reliable documentation which really blocked our progress for a long time. Although now the specifications are stable and tools are available, the lack of reference implementations is obvious.

openEHR community and developers approach the modeling work from atomic concepts leading some small part of a real HIS. It is believed that these small portions of well modeled critical concepts, such as blood pressure or antenatal examination, will eventually lead to a federated virtual EHR. However it should be noted that our approach is at the opposite side; we model the bigger part of HIS via Archetypes and try to set forth the path to complete modeling of a clinical HIS in this study which makes it novel.

4.3 Contributions of the Study to the Modeling Methodology

4.3.1 Extending openEHR Archetype Definition Language (ADL)

Identification of a problem in ordering of internal references

The order of concepts and their features are important in a clinical model to be used in a HIS for complying with the routine clinical thinking of physicians during data entry, validation and querying which greatly affects user acceptance. This model is also important for the design of user interfaces and the application logic. For example in the case of a simple body mass index, without measuring weight and height, you can not determine manually or automatically the value of this index. Another example might be the order of physical examination which usually starts from the top (head) to the bottom (toes) or according to organ systems. In our modeling task, this ordering became very obvious during modeling of MST Findings. We have encountered highly repetitive and nested structures with many children nodes. Some of these nodes have common repeating attribute patterns like the site of the lesion or extent of a finding which is already formally described elsewhere. In this case there is no need to rewrite the whole attribute and an internal reference consisting of the full path to the original describing node. Internal references are shown in italic text in the example given in Box 4.1.

```
CLUSTER[at3300] occurrences matches {0..*} matches { -- Granular
items cardinality matches {0..*; ordered} matches {
        use_node ELEMENT /items[at3000]/items[at3100]/items[at3110]          -- Extent
        use_node ELEMENT /items[at0050]/items[at0100]/items[at0500]}}        -- Site(s)
CLUSTER[at3400] occurrences matches {0..*} matches { -- Friable
items cardinality matches {0..*; ordered} matches {
        use_node ELEMENT /items[at3000]/items[at3100]/items[at3110]          -- Extent
        ELEMENT[at3420] occurrences matches {0..1} matches {-- Bleeding
                name matches {
                        CODED_TEXT matches {
                                code matches {[ac3420]}}}                    -- Bleeding
                value matches {
                        CODED_TEXT matches {
                                code matches {
                                        [local::
                                        at3421,   -- Yes: Spontaneous
                                        at3422,   -- Yes: Contact bleeding
                                        at1122]   -- No
                                }}}}
        use_node ELEMENT /items[at0050]/items[at0100]/items[at0500]          -- Site(s)
```

Box 4.1 The use of internal references within a container attribute of type: CLUSTER

The precise order of the internal references, which occur frequently in the modeling task of MST, could not be set together with various nodes containing elements, complex types and other internal references. We have issued a problem report which resulted in a formal change request at openEHR (CR-000104)[1].

The solution to the problem was to change the syntax rules of ADL to allow a correct parser to be built (ADL Issue 1.2, Release 0.95, November 15, 2004). The changed rules of ADL are:

- Any identifier with a leading capital letter is taken to be a type identifier (i.e. a class name),
- Any identifier with a leading lower-case letter is taken to be an attribute name,

[1] This change request resulted in modification of ADL and the parser which was published in ADL 1.2, Release 0.95, 15 Nov 2004.

- The only other place where identifiers exist in ADL is as tags in invariants; here the identifier may have either form (initial upper or lower case).

Introduction of a new method for further constraining occurrences of internal references at referencing location

The problem was that in an Archetype model it was not possible to constrain the occurrences of internal references (use_node) at referenced point. Technically when two or more items exist (with {0..y} occurrences) under a container structure together with one or more internal references (at the original location having occurrences wider that needed at referencing site such as {0..x}, there was no way to constrain specific occurrences (number of instances) or conditional existence of these items in ADL (See Box 4.2).

```
CLUSTER[at3100] occurrences matches {0..*} matches { -- Erythematous (Hyperemic)
    items cardinality matches {0..*; ordered} matches {                    (Container Attribute)
        ELEMENT[at3110] occurrences matches {0..1} matches {          -- Extent
            value matches {
                CODED_TEXT matches {
                    code matches {
                        [local::
                        at3111,          -- Localised
                        at3112,          -- Patchy
                        at3113,          -- Striped
                    }}}}
    ELEMENT[at3120] occurrences matches {0..1} matches {          -- Bleeding
        value matches {
                CODED_TEXT matches {
                    code matches {
                        [local::
                        at1121,          -- Yes
                        at1122,          -- No
                        at3123]          -- Stigmata of bleeding
    }}}}
    use_node ELEMENT /data[at0003]/items[at0050]/items[at0100]/items[at0110] --Site(s) (At origin
occurrences {0..*}
```

Box 4.2 An example depicting inability to constrain referenced node

As can be seen in above example, there are three items (two elements and one reference) under container structure CLUSTER. The constraint on occurrences of the internal reference which is shown in italic text is zero to many {0..x} at original site. The CLUSTER's cardinality is zero to many {0..*} which means it can have none or infinite number of items. A typical situation as shown here is that the reference should occur at least once only if any of the elements occur. Another frequent situation might be that the occurrence of the reference item should be a certain number of times or an interval.

Our proposal for solution was to override original "occurrences" property of reference items at referencing location(s). This has enabled proper definition of the semantics for occurrences on internal references correctly which enhanced expressive power of ADL.

The solution has resulted in the following changes:

-The ADL specification is changed to allow occurrences to be specified on internal reference node, so that this overrides the occurrences of target point, which are otherwise taken as default,

-The AOM explanation of the internal reference (ARCHETYPE_INTERNAL_REF) class is improved to clarify how occurrences should be parsed and serialized.

The resulting change in ADL can be visualized as bold and underline text as shown below:

use_node ELEMENT **occurrences {0..1}**
/data[at0003]/items[at0050]/items[at0100]/items[at0110] --Site(s) (At origin occurrences {0..*}

We also have issued a problem report which resulted in a formal change request at openEHR (CR-000233)[2].

[2] This change request resulted in modification of both ADL and AOM which was published in ADL 1.4 and AOM 2.0.1, Release 1.0.1, 13 Mar 2007.

Extended Archetype Metadata to include Bibliographic Information and other Knowledge Sources' References

The importance of having bibliographic information for any artifact in science such as list of bibliography in an academic publication is beyond dispute. Since the archetypes are strong candidates for sharing of knowledge in clinical medicine, it is essential that they have references to existing publishing and other online sources. If we look at the proposed methodology for Archetype development, there are many similarities in the design and quality control of Archetypes with that of scientific articles. Peer-reviewing is done by domain experts and technical people by using online tools (Beale and Heard, 2006).

The problem we had encountered was that archetype meta-data part (description) did not include a bibliographic section.

Our solution was to extend the ADL meta-data to include an extensive bibliographic section (See Box 4.3) by incorporating information items common to all kinds of knowledge sources (such as articles, books, patents, online repositories, clinical guidelines and so on). These are:

- **Type** (scientific article, book, book chapter, conference proceedings and presentations, government reports, personal communication, scientific publications on the Web, technical reports, pamphlets, clinical guidelines and online resources)
- **Status** (published/working/preprint/in progress for articles and books, may include different items for other types)
- **UID**: Unique Identification for resource (For articles PMID, books ISBN or DOI)
- **Reference**: main description of the resource in classical scientific reference list format is recommended.
- **Organization**: the organization which published the resource of author's affiliation.
- **Access_URI**: the URI of the file or online resource Internet access address.
- **Annotation**: A free area to note any relevant information about the resource and how it relates to the archetype

```
Description
    original_author = <>
    details = <
        ["en"] = <
            language = <"en">
            purpose = <"To record some clinical concept">
            use = <"For capturing data in a standard way">
            misuse = <"Not to be used for other concepts">>
    lifecycle_state = <"initial">
    bibliography = <
        ["1"] = <
            type = <"scientific article">
            status = <"published">
            UID = <"PMID=15032077">
            reference = <"Atalag K, Bilgen S. Modeling of domains. JAMA 2001; 40(4):275-87">
            organization = <"METU Informatics Institute, Ankara, Turkey">
            access_URI = <"www.ii.metu.edu.tr/papers/1001.pdf">
            annotation = <"This article sets forth the domain modeling principles used in this archetype">>
        ["2"] = <
            type = <"personal communication">
            reference = <"on defining semantics with Semih Bilgen on 14/12/2005">
            organization = <"METU Informatics Institute, Ankara, Turkey">
            annotation = <"any information relevant for archetype such as topic,result, etc.">>
        ["3"] = <
            type = <"online publication">
            reference = <"MTHMST2001 on UMLS Knowledge Sources v 2007a. Accessed 11/03/2006">
            organization = <"National Library of Medicine, Bethesda, USA">
            access_URI = <"http://umlsks.nlm.nih.gov/kss/">
            annotation = <"Main source of terminology used in archetype. Free access ">>
```

Box 4.3 The extended meta-data of archetype with example

This solution now provides the ability to provide basic decision support by providing users with recent and relevant published papers about a particular subject in Archetype. Since the papers published in indexed/refereed/peer-reviewed journals are still by far the ultimate sources of knowledge, this might be extremely useful for context dependent and seamless linking of Archetype based clinical information systems to these knowledge sources. More advanced uses might be enabling access to

clinical practice guidelines and Evidence Based Medicine (EBM) repositories (van Bemmel and Musen, 1997).

Extended Archetype Ontology Section Term/Constraint Bindings parts so as to reference terms directly to UMLS

The large number of internationally significant coding, classification and terminology systems, medical ontologies, clinical repositories and even some software system's databases in the Unified Medical Language System (UMLS) is presented online as UMLS Knowledge Sources. Recently MST has also been integrated to UMLS which is extremely important for our Archetypes to connect to other systems (Tringali, Hole and Srinivasan, 2002). The glue among these different knowledge sources is the unique CUI – concept unique ID. However there is an important point to be taken into account when used or referenced from information systems: For each concept there may be multiple terms (i.e. SNOMED-CT, ICD10AM, ICPC and so on) each with a different Term Unique ID. So in order to specify a particular term one needs to specify both the CUI and Term Unique ID.

A specific concept in MST_Esophagus archetype term_binding section is given in Box 4.4.

Concept: Esophageal anastomosis procedure
CUI: C0940040
Semantic Type: Therapeutic or Preventive Procedure
Definition: None found.
Synonyms:
Esophageal anastomosis
Esophageal anastomosis (site)

Box 4.4 Representation of an MST Term linked to a UMLS concept

As can be seen there are two synonyms for this concept each having a different Term UI. So in order to specify a particular term in UMLS, the corresponding Term UI also need to be specified as well.

Synonym 1: Esophageal anastomosis, CUI: C0940040, Term UI: L1834016

Synonym 2: Esophageal anastomosis (site), CUI: C0940040, Term UI: L1834017

The problem was that it was not possible to reference to correct terms in ontology section, term_bindings and constraint_bindings subsections.

Therefore in this study we have extended ADL to store both CUI and Term UI in appropriate subsections under ontology section which is shown in part of an MST archetype in Box 4.5.

```
["UMLS"] = <
    items = <
        ["at0003"] = <[umls::C0014876-L0014876]>        -- Esophagus
        ["at0100"] = <[umls::C0577015-L1098274]>        -- Normal esophagus
        ["at0501"] = <[umls::C0939942-L1834903]>        -- Esophagus, crico-pharyngeus
    >
```

Box 4.5 The extended archetype reference to UMLS by using CUI and Term UI together

A possible approach to model concepts with encapsulated data

When authoring an observation archetype for an examination modality containing images or bio-signals, such as ECG or endoscopy, the formal report contains both the textual and the multimedia data. As increasing numbers of these devices comply with proper data exchange standards, structured textual data such as contextual information or image annotations and automatic interpretations are readily available. However in HIS, usually the textual and electronic version of report is kept separately and the non-textual part of the examination data is simply stored as binary objects which render the textual information inaccessible. This valuable contextual data and clinical information are also present in other modalities like EEG, EMG to name a few.

Currently any complex e-Modality data is represented by the DV_ENCAPSULATED or DV_MULTIMEDIA data type in an Archetype which is shown in Figure 4.8. But all the embedded structured data becomes inaccessible again - at least feasibly.

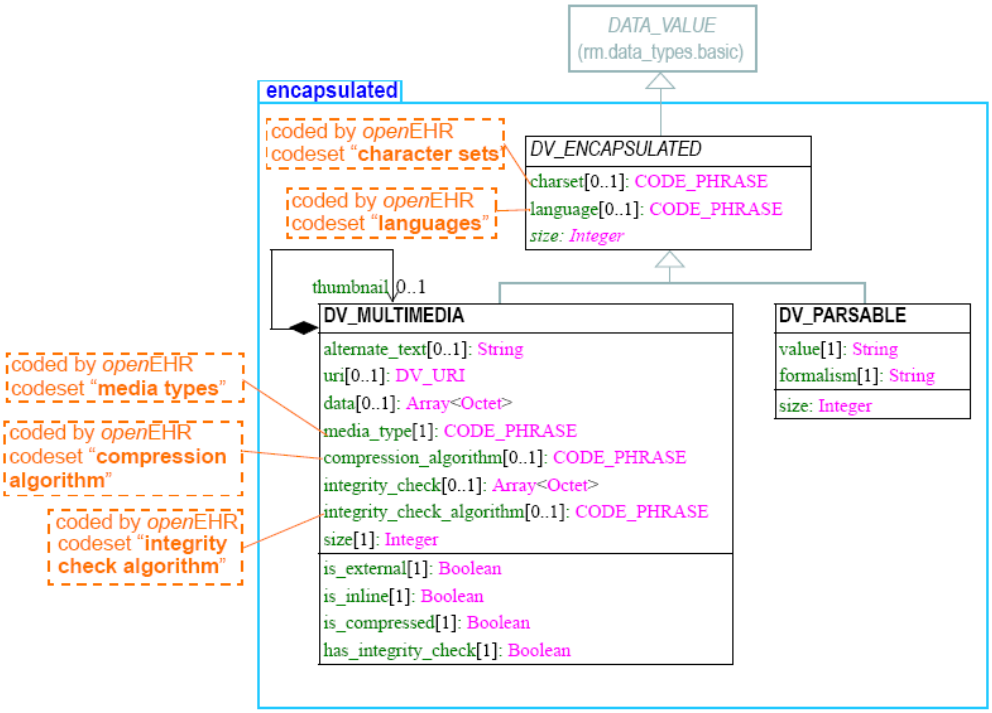

Figure 4.8 The Encapsulated Package which contains data types for non-textual data (openEHR Reference Models, 2007 with permission from openEHR Foundation)

As can be seen, the only meta-data allowed is the "size" in DV_ENCAPSULATED class, the attributes of DV_MULTIMEDIA such as "media_type". Specific attributes or a generic attribute to cover modality specific features or accompanying contextual information such as type of contrast media in a radiological study is not present.

Our novel method enables archetypes to model such situations effectively, without loosing data and without altering the current semantics of openEHR specifications too much such as creating a data value for each modality. The details of the methodology are given as follows:

a) Design an Observation Archetype with the usual data structures and elements
b) Place the multimedia data including contextual information and annotations complying with DICOM into DV_MULTIMEDIA data type such as "dicom"
c) Create in Multimedia or Encapsulated class a parametric function (such as item ID) mapping to the item of interest within the DICOM file (i.e. frequency, latency, duration etc.). This function will "get/fetch" the

corresponding data items and convert them to appropriate openEHR data types.

d) Some of the elements in the Archetype may set their values by calling this function with an appropriate parameter.

Actually there are similarities with the internal references or Archetype slots. The notion of referencing values or structures is already implemented in Archetypes. The only big difference is "when" this happens: The currently supported referencing is at "Design/Class" level, meaning it happens before object instantiation. However in our approach, a value will never be available before Archetype instantiates but the structure and type info will be available.

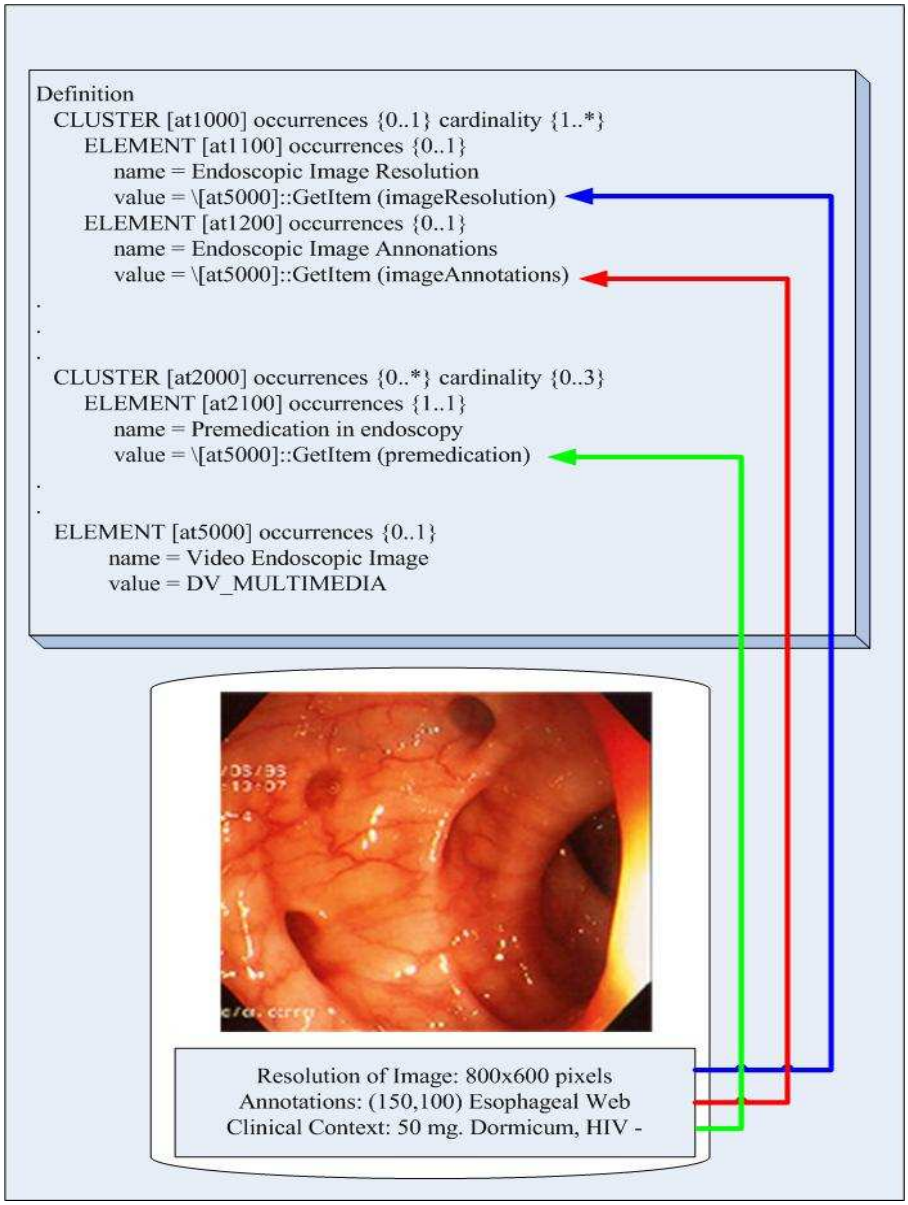

Figure 4.9 Schematic representation of runtime referencing of intrinsic data within encapsulated files. The sample image in the figure shows an endoscopic view of human gastrointestinal tract captured from a videoendoscope during the case study.

When Archetypes are instantiated, each one loaded with different multimedia data, referencing elements' values can be populated directly from the multimedia data they embrace. A huge benefit we foresee is that a query against the persisted Archetype data can efficiently search for the structured data rapidly by using Archetype structure as guides. Otherwise either the multimedia file has to be indexed or parsed at runtime (not efficient and may lead to redundancy) or these data are entered by the feeder system (unlikely). A schematic description of the methodology is given in Figure 4.9.

Since this approach is generic and simple, it can be applied to other modalities without altering/breaking the overall structure and semantics of openEHR approach. Possible candidates are electronic ECG files and we anticipate many more to come from biomedical and genomic world.

For an example, it enables an openEHR based application to retrieve the resolution and size (pixel) features of an ordinary image file which is not currently supported in the meta-data part. Also it is very likely that a radiologist might want to instantly know the slice thickness and parameters like electrical potential and current (kV/mA) of a computerized tomography scan. None of these attributes are currently supported and it is not a good idea to hardcode these into generic data types.

4.3.2 A Novel Paradigm: Defining Semantic Equivalence of Archetypes

When modeling a complex clinical domain with the guidance of a specific terminology or domain ontology, there exists a natural/original structure or hierarchy of the concepts that describe that domain which is also called as canonical form. This is usually the format of the original article, consensus report (if already defined formally) or as taken from common practice after analysis. In this case, the modeling work usually follows that hierarchy. However when modeling a clinical observation, such as a radiologic study, physical examination or endoscopy, one might encounter a different hierarchical organization of the concepts other than the original structure in routine clinical practice. Consider, for example, a radiological study of the chest where similar types of findings (i.e. solitary nodes) are dispersed all around the thorax. In this case the radiologist might prefer to report the findings by first describing the abnormal finding and then report in which locations it exists. We can call this a “finding” oriented hierarchy. In another case where many abnormal findings are observed in a single or few locations, this approach might not be practical and the radiologist may prefer a "site" oriented hierarchy for describing the results of the study and for reporting. Here the location is given first and then observed abnormal findings are described instead of describing each and every finding first and then giving the location. So we can safely conclude that different modeling hierarchies for a given domain exist.

This variation creates problems because different models for a single domain might degrade data processing capability and interoperability of HIS. However, each model in fact is a valid representation of that domain which humans can perfectly understand and interpret the same way. But the problem is that computers can not comprehend these alternative models and process them. Our solution was first to define semantic equivalence of these alternative models and then define the means to represent it and effectively manage in Archetypes.

We can describe the semantic equivalence by an analogy to usual human written communications. Unconsciously we agree on an alphabet, words and a grammar to form sentences to properly understand each other. In addition we must also know how different sentences composed of synonyms or having a different grammar have the same meaning. In domain modeling, the RM can be considered as the alphabet and the terms possibly from a known terminology can be thought of as words. In the same way that one has to put words into a logical order to be able to form understandable sentences, the terms in a clinical model need to have a certain organization in order to represent a clinical concept. Inline with the analogy, humans communicate and understand each other while computers exchange data and interoperate. When computers are able to send each other messages with an agreed data schema they are considered to have data level or functional interoperability. This is accomplished by the RM in the Archetypes and multi-level modeling methodology. Semantic interoperability simply means that each communicating party not only is able to exchange data but also understand and interpret it as the same. This is accomplished by the Archetypes coupled with external domain terminologies which assemble terms into a proper hierarchy and ensure the meaning of the terms are interpreted the same by the receiving party. There are different levels of achieving semantic interoperability and the quality of domain terminology and links to other knowledge sources determine this level. As humans need to know how different sentences with different words and grammar may have the same meaning, computers also must be provided with some means to identify models with the same purpose but with different terms or hierarchy. Semantic equivalence of domain models means the possibility of having different computationally usable domain models with different terms or hierarchy without breaking the semantics of the domain concepts which are readily transformable to other equivalent models without any loss.

There was no notion about semantic equivalence of Archetypes in the modeling methodology. First we have shown that such a paradigm exists and elaborated some notes towards the definition of semantic equivalence of Archetypes. Then we proposed a methodology to validate and effectively manage semantic equivalent Archetypes. An example showing the original MST hierarchy and the modified model is given in Figure 4.10. Then two other alternative models which are equivalents of model two are presented in Figure 4.11. Sample terms are provided for ease of understanding next to the boxes of domain hierarchy. It should be noted that due to errors in original MST hierarchy it was not considered as the canonical form in the modeling work in this study.

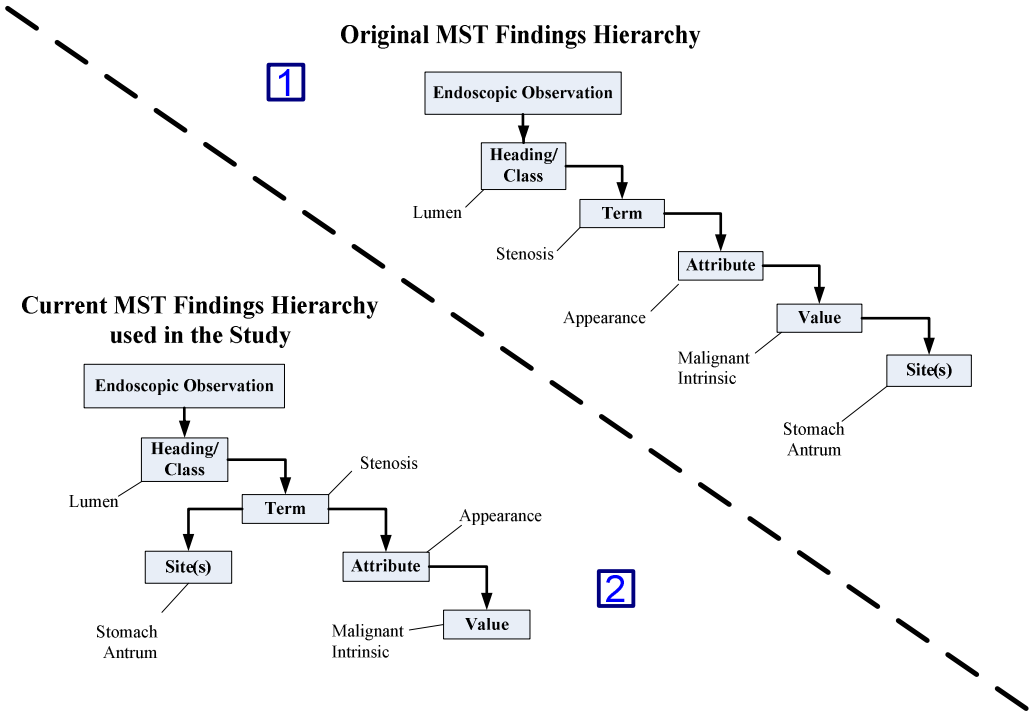

Figure 4.10 The original MST Findings hierarchy and modified model used in the study

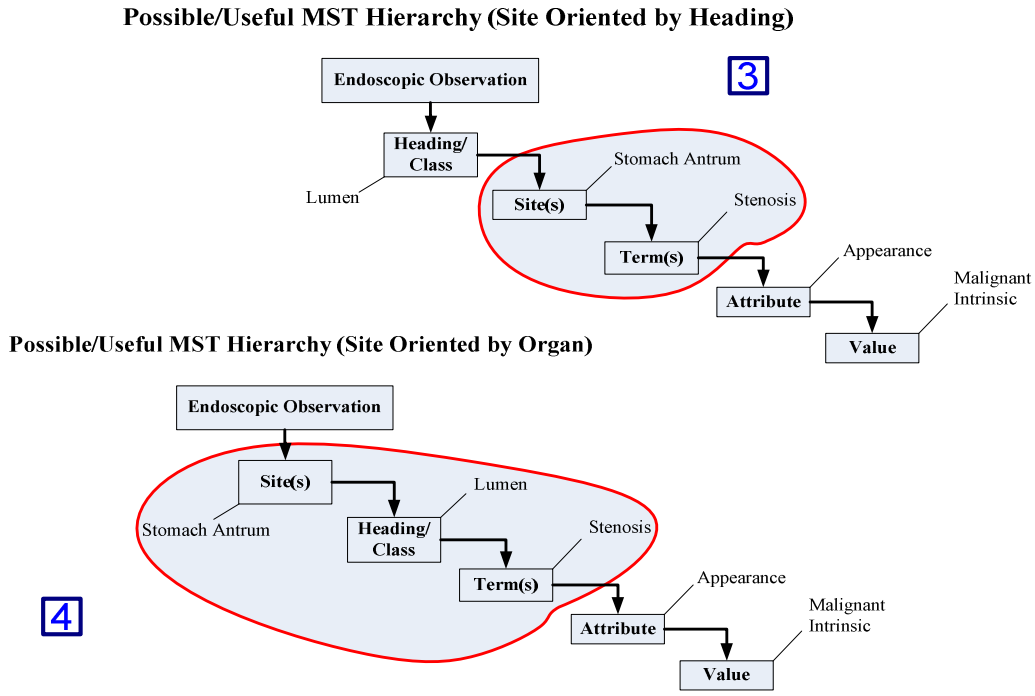

Figure 4.11 Two alternative models which have semantic equivalence to model two

As can be seen, there are four distinct models, in reality describing the same endoscopic findings by using same terms but with a different structure. Some alternative expressions of this observation according to the model in comparison with common clinical language are:

Natural language expression: "Malignant intrinsic appearing stenosis of the lumen at the stomach antrum"

- Model 1: "Lumen has stenosis having appearance of malignant intrinsic at stomach antrum "
- Model 2: "Lumen has stenosis in stomach antrum having appearance of malignant intrinsic"
- Model 3: "Lumen at stomach antrum has stenosis having appearance of malignant intrinsic"
- Model 4: "At stomach antrum lumen has stenosis having appearance of malignant intrinsic"

A clinician can perfectly understand all expressions above as semantically equivalent. However in order for computer systems to interpret these expressions as being the same and process, they must also share the models and terminology that generate these

expressions. The importance of semantic equivalence becomes apparent at this point because even after sharing models, there must be a way to explicitly define models describing the same domain. Otherwise we can not guarantee that systems are able to fully understand each other and have a high level of interoperability.

While the models one and two aim to model endoscopic findings and use the very same terms, they are not semantically equivalent because they not only differ significantly structurally but also all expressions from both models are not equivalent. So it was not possible to convert between the two models without using human reasoning and clinical expertise. However models three and four are semantic equivalents of model two. The simple reason is that these three models are in fact different views of the same clinical observation used commonly in clinical practice. And computationally it is possible to convert data from one model into the others: the same terms are used and none is left unused, and the same constraints can be applied such as existences, occurrences and cardinalities. However although the same terms are used in model one, the constraints for defining semantics are incompatible. An example is the inability to express mutual exclusivity or conditional existence of terms and attributes in model one. Another problem occurs with linking of the site to attribute values; there are many terms without attributes but definitely a need to have site information. All these points are possible to express in other models. Unfortunately there is no single rule to detect semantic equivalence but we have identified some important points for semantic equivalence in addition to expert opinion which are listed below.

Notes towards the Definition of Semantic Equivalence of Clinical Domain Models

1. Models with different hierarchies usually constitute alternative views for different purposes and can be said to have certain axes. We have identified a mono-axial semantic equivalence in the study; that is only one concept in the hierarchy can change its position at a time without breaking the semantics of the expression. Showing existence of dual or even multi-axial cases in which two or more concepts change positions at the same time might be a good future study.
2. In this study the semantic equivalence shown in MST models have the same terms but different hierarchy. We believe that semantic equivalence may also be

established by using different (but similar) terms which happen in natural human discourse.

3. Concepts that can interchange always have an Entity-Attribute relationship. In our model this is the Term-Site(s) and Heading-Site(s) relationship where Site(s) is an attribute of the Term which then this relationship reverses. Site(s) may also become an entity for Heading.
4. These transposing elements are mostly real-world entities that can exist by themselves; such as anatomic locations, time intervals or certain events like high blood glucose levels. They do interchange with other real-world entities.
5. A strong feature of these attributes is that they accompany all the entities in the model; nearly all the terms in our model have anatomic location attributes (Site). In places where this attribute is not present in MST is not because it does not apply for a particular term but because it is assumed by all to have a certain value such as Hiatus Hernia of the esophagus in which the location is implied in the name of the term. So attributes which apply for most of the entities can be considered to form an axis to the model and are good candidates for transposition.
6. Adjacent concepts in the hierarchy are more likely to change positions. In our model **Site,** which is an attribute of **Term** and adjacent to it.

Considering above notes about achieving semantic equivalence, the position of Site(s) after Attribute Value violates points three through six compared with other models. Consider the following small part of MST Findings for Esophagus in Box 4.6 and a possible expression:

Mucosal sclerosis	Type	Spontaneous Post-therapy	Site(s)
	Extent	Localised Patchy Diffuse	Site(s)

Box 4.6 MST Findings for Esophagus depicting list of terms, attributes and values for describing abnormal finding mucosal sclerosis

As we had mentioned before, linking Site(s) to attribute values results in inconsistencies as shown in above situation. While it can be inferred that the Site(s) in the first attribute can be applied to both attribute values, it is linked to the third

attribute value in the second attribute. It is not clear whether Site(s) is not needed for the first and second attribute values so it is left to interpretation of clinical experts.

While perfectly valid expressions can be built according to the model, meaningless expressions are also possible such as:

"Mucosal sclerosis having spontaneous type at middle-third of esophagus and having diffuse extent at cardia, lower-third and anastomosis".

This expression does not make sense because there is a single finding and it can not occur in different locations for each attribute value. In other words, for each observed term the attributes of that term has to define consistently a single observation. It should also be noted that it is not stipulated whether attributes are mandatory or not.

Even though the first model is not computationally useful, it is still a model that helps clinicians by establishing a common language. Due to distorted expressive power and potential inconsistencies in the data collected, we can comfortably say that it does not conform to semantic equivalence which we describe here.

In summary, different models may exist for a particular domain some of them with the same meaning. Depending on different people, purposes, consequences of observations and findings, these alternative models might be used. However in order for computers to interpret these models as same and integrate and consolidate data collected, semantic equivalence has to be defined and formally represented in the model.

For validation of semantic equivalence, the compositional knowledge within UMLS; specifically the canonical model of SNOMED might be used to test equivalence of expressions derived from alternative models. In this method the Archetype paths and links to SNOMED concepts for each node in the canonical Archetype model will produce a canonical form based on the SNOMED hierarchy. And then semantic equivalence of candidate models can be tested by checking if they comply with that form. While complete matching of models with this method guarantees semantic equivalence, non-matching models might still be semantically equivalent (SNOMED-Transforming Expressions to Normal Forms, 2006).

The author is aware that the stipulations above do not constitute a strictly formal definition of semantic equivalence of archetypes. This is because the representative power of these models has been based on natural languages, which, as known very

well, defy formal semantic processing. The problem of semantic equivalence, however, has long been a significant issue, as illustrated here, and enhancing the modeling methodology to cater for it is considered to be a major necessity.

Our solution consists of first discovering alternative domain models and manual identification of semantic equivalence then followed by Archetype modeling. The solution also provides methods to effectively manage these Archetype models. These methods will enable transformation of semantically equivalent Archetype models and also aid in data integration and consolidation. We would like to point out to the fact that semantic equivalence is not a formal equivalence of Archetype domain models where in addition to vocabulary and semantics, compositional knowledge within the model and its information models should also be taken into account. In other words the extra knowledge provided such as existence, occurrence, cardinalities and rules defined in invariant sections such as conditional existence and also information model types and classes has to be equivalent which is a difficult task and will be considered as future research.

Archetype Modeling of Semantically Equivalent Domain Models

There are two basic approaches: either to design a separate Archetype for each alternative or find a means to represent all alternatives within a single Archetype. In the former approach, each and every Archetype needs to reference others. But there is a potential shortcoming. Alteration of one Archetype may break semantic equivalence and problems with data integrity and system interoperability might follow. The latter approach is safer and, although harder to implement, was our preference in the study. Here there is only one Archetype which includes means to extract any of the equivalent Archetypes from the original one. This necessitates addition of a new section in ADL to define the methods which include transposition rules and element identifiers. The single Archetype will be designed according to the canonical hierarchy, but concepts that can have alternative positions need to be declared in that section.

This approach should also ensure keeping the semantic equivalence intact during specialization of an Archetype.

Creating the optional "Equivalence" Section in Archetypes

The proposed new section is presented in Box 4.7.

archetype (adl_version=1.4)
openEHR-EHR-ITEM_TREE.MST_Colon.v2
concept
[at0000] -- MST Findings for Colon
language: language description
description: archetype meta-data

equivalence: new section for declaring semantic equivalence of Archetypes

definition: main archetype content
ontology: terms, constraints and bindings

Box 4.7 The new optional section in Archetypes for semantic equivalence

The details of the new section is presented in Box 4.8

equivalence

<1>
different_terms = <True/False>
name = <"Site oriented MST findings model under headings">
purpose = <"To provide an anatomic site oriented view of MST findings for each Heading">
source_entity = <Archetype path of relocating entity>
target_position = <Archetype path of new position>
target_occurrences = <"{1..*}">

<2>
different_terms = <True/False>
different_hierarchy = <True/False>
name = <"Site oriented MST findings model for whole organ">
purpose = <"To provide an anatomic site oriented view of MST findings for each organ">
source_entity = <Archetype path of relocating entity>
target_position = <Archetype path of new position level>
target_occurrences = <"{1..*}">

Box 4.8 Details of the new equivalence section

First item lists the first Archetype features and declares positions for transposition. Sub-items are:

different_terms: is a Boolean value whether terms other that the original Archetype are used

name: a string consisting of a friendly name for the particular equivalent Archetype

purpose: a string consisting of a short explanation of the purpose

source_entity: the absolute Archetype internal path of the relocating entity (i.e. anatomic site)

target_position: the absolute Archetype internal path of new position(s) under which the relocating entity will be placed

target_occurrences: a string value for overriding occurrences of original entity of attribute type

The relocating entity might be positioned at either a single or a set of locations within the Archetype hierarchy. These locations can be depicted manually by writing paths separated by commas or by using a wildcard character to copy relocating entity to all positions below that level. The details are explained in Box 4.9.

target_position = </items [at1000]> -- Single position under this level

target_position = </items [at1000], /items [at2000]> -- Multiple positions under this level

target_position = </items [*]> -- All positions under this level

target_position = </> -- Top level position(s) in the hierarchy

Box 4.9 Explanation of target position definition in equivalence subsection

Handling of Different Terms in Semantic Equivalent Archetypes

If different terms are used in the equivalent archetypes, then new terms with same local Archetype terminology identifiers [atXXXX] has to be declared in the ontology section. We propose that each alternative term should be declared in the ontology

section together with the original term. The equivalent Archetype identification number has to be provided for each alternative term. In cases where the same term is used in both archetypes then there is no need for extra declaration. An example from openEHR-EHR-ITEM_TREE.MST_Colon.v2 Archetype is given in Box 4.10. Hence an Archetype parser will be able to select correct terms during transformation.

```
["at1200"] = <
        description = <"TERM: Stenosis">
        text = <"Stenosis">
        description(1) = <"TERM: Narrowing in the lumen">
        text(1) = <"Narrowing">
>
["at1210"] = <
        description = <"Attribute: Appearance">
        text = <"Appearance">
>
```

Box 4.10 An example for declaration of different terms in archetype ontology section

Here in ["at1200"] the first description without any identifier belongs to the original Archetype. The following alternative term declared with text in italics has equivalent Archetype identification number in parenthesis. In ["at1210"] there is only single declaration which means both Archetypes will use same term.

Transposition of Entities and Deriving Semantically Equivalent Archetypes

The method for transposition of entities is described as follows:

1) Check the **equivalence** section

2) If equivalent Archetypes exist then get the details of the first Archetype

3) Cut the entity from its original position declared in **source_entity**

4) Paste it to the position(s) given in **target_position** according to the definition. If it is placed in multiple positions then use internal references to first entity at other positions.

5) Clear all internal references to the original entity if any (use_node references)

6) Check CLUSTERs and delete “items” attribute if it no more contains any child nodes (after removal of internal references)

7) Check whether different terms are used in the equivalent Archetype by assessing **different_terms**. If different terms are present then fetch appropriate terms from the ontology section using the identification number of the equivalent Archetype

8) Continue with the next equivalent Archetype if any and start again.

All semantic equivalent Archetypes can be derived from the original Archetype by using this method. It is also useful to convert and consolidate data collected conforming to each equivalent Archetype.

There are certain questions to be answered in defining semantic equivalence of different Archetype domain models. Since the determination of semantic equivalence needs to have an expert opinion, it is possible to have disagreement among different experts even though computationally these models are readily transformable. It should be noted that Archetype development usually is performed by a panel of domain experts, a consensus may be reached easily. Conversely it is also possible that some models to have semantic equivalence even though domain experts reject. Currently as we propose, the definite decision of semantic equivalence is subjective. However, future research on this issue may yield formal methods of defining semantic equivalence. Currently we can only demonstrate the existence of this paradigm and propose some guidelines to discover them.

4.3.3 Extending openEHR Reference Model (RM)

Introduction of “Flavors of Null” for Missing Data into Top Level RM Structures

In clinical medicine, when describing a clinical story, the first step is to depict if certain information is present or not. As an example, during patient history taking, when a physician asks if patient is smoking, the outcome might essentially be positive, negative or no information. Flavours of Null codify reasons of missing data such as:

not asked, patient not able to answer, patient do not answer, doctor has forgotten and so on. When modeling this concept, “Smoking” item will be represented by a container structure: CLUSTER in openEHR or abstract class in Object Oriented formalism. Then further attributes such as quantity (how many cigarettes per day), duration (since when) will be represented by leaf nodes: ELEMENT in openEHR (See Box 4.11).

```
CLUSTER [at1000] occurrences matches {0..1} matches {          -- Smoking      (Container Structure)
   items cardinality matches {0..*; ordered} matches {
      ELEMENT [at1100] occurrences matches {0..1} matches {   -- Quantity of smoking   (Leaf Node)
         value matches {*}
         null_flavour matches {CODED_TEXT matches                -- Reason for missing data
            code matches {
               [openehr::
                  271,      -- No information
                  272,      -- Unknown
                  273,      -- Masked
                  274]      -- Not applicable
      ELEMENT [at1200] occurrences matches {0..1} matches {  -- Duration of smoking    (Leaf Node)
         value matches {*}
         null_flavour {0..1}
```

Box 4.11 Introduction of Flavours of Null to container structures

As can be seen it was not possible to record reasons of missing data about the top level structure such as the reason for having no data about smoking which is extremely important in clinical medicine in the above example. The only indirect solution is to introduce a new attribute (ELEMENT node in openEHR) depicting the presence of each container structure and use its “value” and “null_flavour” attributes to record this vital information.

Our solution for this problem is to introduce a **null_flavour** attribute into container structures such as CLUSTER (Figure 4.12). Other advantages of this solution might include:

- Reducing size, increasing manageability and understandability of (big) Archetypes,
- During querying of leaf-nodes in a huge repository, the search algorithm can first check parent nodes' presence and then conditionally go down to leaf-

nodes. If a whole branch is null from the top, then there is no need to search for lower levels and the performance might be enhanced.

- The control of adherence to standards will be kept within the standardization body because the decision of how to implement the problem will not be left to the third parties.

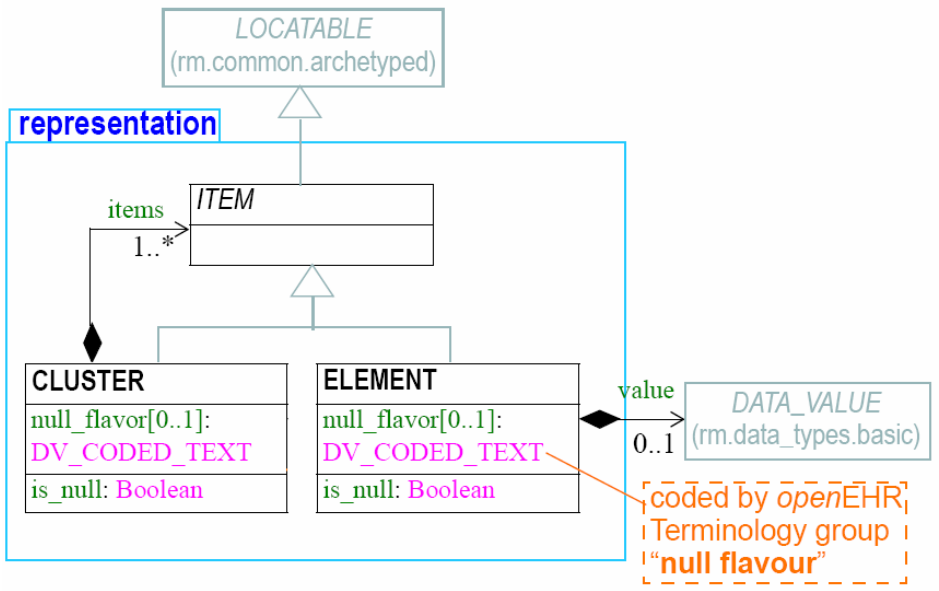

Figure 4.12 Introduction of null_flavour to CLUSTER

Relaxing the Optionality of "items" attribute of CLUSTER Data Structure

The CLUSTER data structure in openEHR has "items" attribute which contains child nodes that may be ELEMENTs or other CLUSTERs. In MST Archetype modeling, we have confronted situations where we had to use four levels of nested nodes. For example HEADING (Lumen) > TERM (Stenosis) > ATTRIBUTE (Length) and SITE(S). The leaf nodes are naturally ELEMENTs with values. But the upper level nodes must be of CLUSTER type. When CLUSTERs do not contain any child nodes by design or deleted during a semantic equivalence transformation operation the mandatory attribute should be optional. We have changed the multiplicity of this attribute from {1} to {0..1} in order to accomplish our task.

4.3.4 Structural Archetypes

There exist complex medical concepts where the structure is highly nested with similar repeating context-free components. An example from MST findings is the upper gastrointestinal examination, an OBSERVATION Archetype, which consists of findings from three different organs: esophagus, stomach and duodenum. The straightforward way of modeling this examination is first modeling findings for each organ as separate Archetypes to avoid redundancy as some findings of organs are also used in other examination types. It is possible to chain Archetypes together via allowing Archetypes to be consumed by others. The mechanism works by providing external references. Thus a parent Archetype for the upper gastrointestinal examination is constructed which reference these three organ based findings Archetypes.

An Archetype should contain sufficient contextual data for human comprehension and also for correct interpretation in computerized systems. This is accomplished by having a number of contextual information within the Archetype such as state and protocol information of a clinical observation. While modeling the findings of upper gastrointestinal examination, the findings from each of the three organs are also OBSERVATION type Archetypes each having their own contextual information and also mandatory RM classes like HISTORY, EVENT coming from EHR RM. However when an Archetype is referencing other Archetypes having same contextual information, the necessity of repeating this information and the possibility of having differences does not make sense and violate data integrity.

We propose that such contextual information should only exist in the top level archetype and the archetypes being referenced should be free of them. Our solution to this problem is introduction of structural Archetypes which are devoid of all contextual meta-data or some of the mandatory RM classes. These Archetypes contain only the necessary data structures and values for domain modeling; for example instead of starting with the general type of the Archetype (i.e. OBSERVATION or EVALUATION) and accompanying mandatory RM classes such as HISTORY, they start with top level data structures which depict whether it is a tree or list of a simple value. Valid structural archetypes can be composed of ITEM_TREE, ITEM_LIST, ITEM_SINGLE, ITEM_TABLE and also a container structure CLUSTER which are defined in openEHR Data Structures RM (openEHR Reference Models, 2007) as

shown in Figure 4.13. Then these structural Archetypes can be used as reusable components in building formal clinical Archetypes.

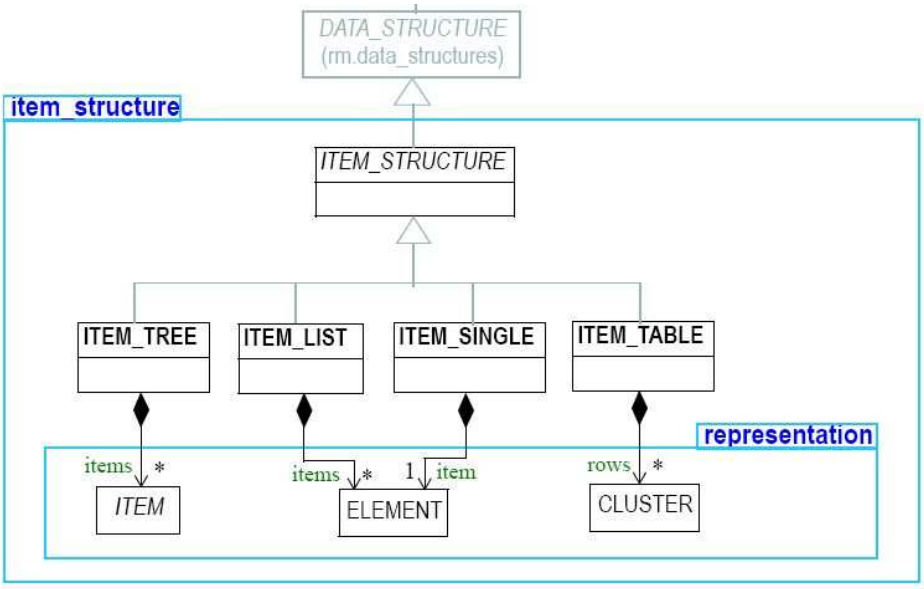

Figure 4.13 Valid data structures to be used in structural Archetypes

By using structural Archetypes of type ITEM_TREE, the modeling of MST Findings was possible which is represented in Figure 4.14. Here the single examination type (involving one or multiple organs) is modeled by an OBSERVATION Archetype and the individual organ based findings are represented as structural Archetypes.

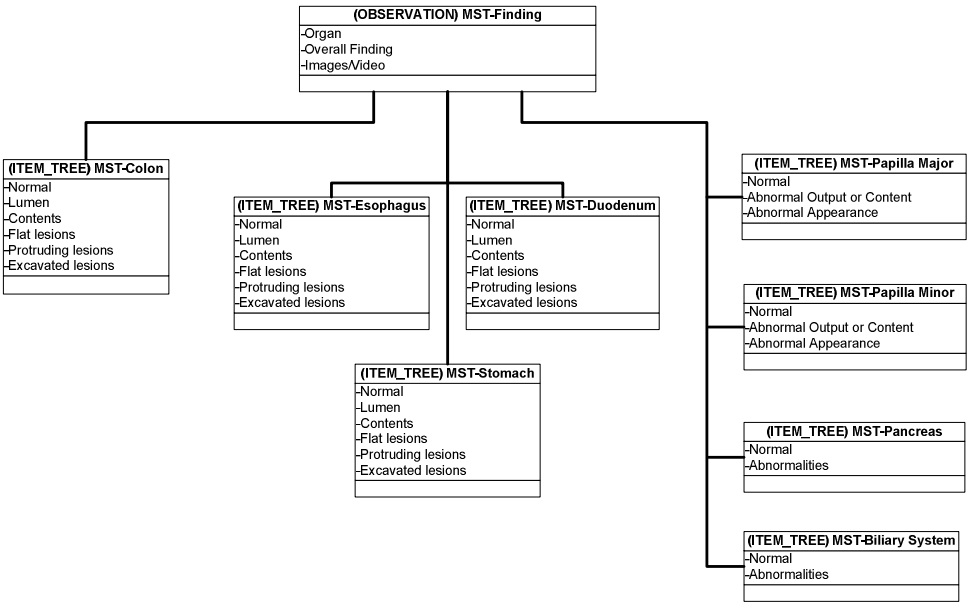

Figure 4.14 The use of structural Archetypes in modeling of MST Findings

4.3.5 Archetype Modeling of MST

For complex observations such as endoscopy there are many ways of describing findings; such as for a colonoscopy examination one might start by first describing the site of the organ, general class of lesion, type of lesion, attributes of this type of lesion and their values. However, the following approach proved to be a better one due to the fact that the organs visualized during endoscopy are mostly hollow and long and same lesions generally tend to involve more than one site. Therefore describing findings in the hierarchy of general class of lesion, type of lesion, attributes and anatomic sites this lesion was observed is obviously easier and quicker for the user. There was also another major design decision which altered not the content but structure of the MST. During an endoscopic examination in addition to observations physicians may also perform some diagnostic or therapeutic procedures, such as stone removal or polyp excision. This was linked directly to the Site so it had been quite difficult and most possibly not useful at all to implement in research prototype software model. Therefore it was separately modeled from the findings as shown in Figure 4.15

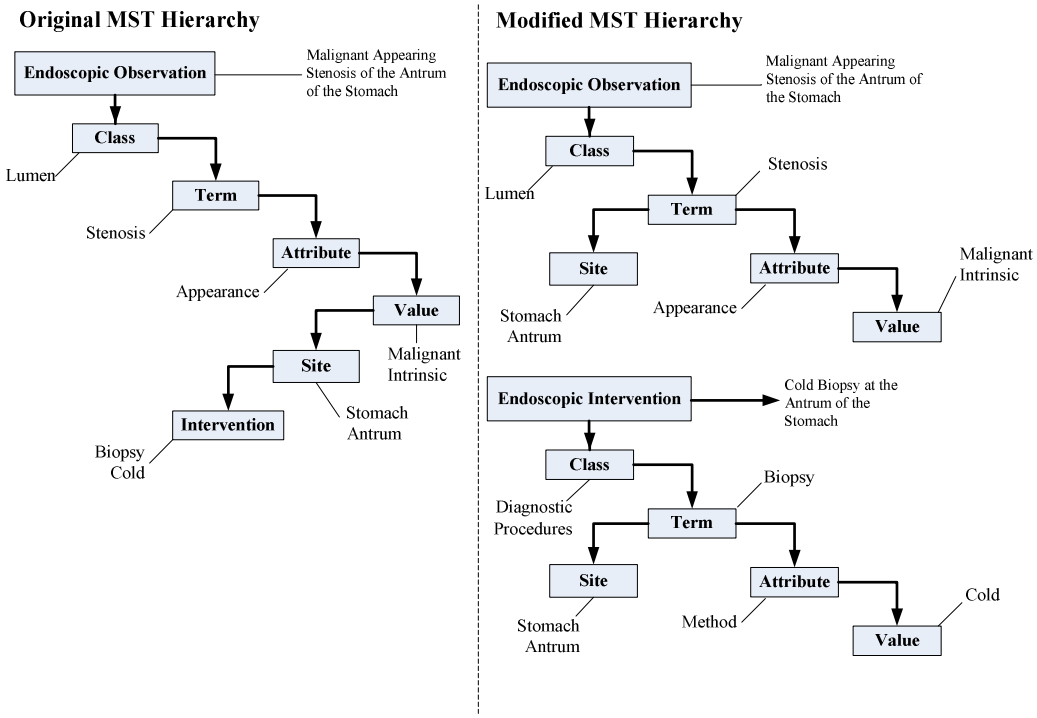

Figure 4.15 The original and modified structure of MST in the study

```
ontology
  primary_language = <"en">
  languages_available = <"en", "tr">
  terminologies_available = <"MTHMST2001", "SNOMED_CT">
  term_definitions = <
    ["en"] = <
      items = <
        ["at0000"] = <
          description = <"MST Findings for Colon in lower GIS endoscopy">
          text = <"MST Findings for Colon">
        >
        ["at0003"] = <
          description = <"the name of TREE Structure">
          text = <"Colon">
        >
        ["at0005"] = <
          description = <"HEADING: NORMAL">
          text = <"NORMAL">
        >
        ["at0100"] = <
          description = <"TERM: Normal">
          text = <"Normal">
        >
        ["at0110"] = <
          description = <"Site(s) for Colon">
          text = <"Site(s)">
        >
        ["at0111"] = <
          description = <"Colon: Site">
          text = <"Anus">
        >
```

```
["tr"] = <
  items = <
    ["at0000"] = <
      description = <"Alt GIS Endoskopisinde Kolon için MST Bulguları">
      text = <"Kolon için MST Bulguları">
    >
    ["at0003"] = <
      description = <"TREE Yapısının İsmi">
      text = <"Kolon">
    >
    ["at0005"] = <
      description = <"BAŞLIK: NORMAL">
      text = <"NORMAL">
    >
    ["at0100"] = <
      description = <"TERİM: Normal">
      text = <"Normal">
    >
    ["at0110"] = <
      description = <"Kolon için Bölge(ler)">
      text = <"Bölge(ler)">
    >
    ["at0111"] = <
      description = <"Kolon: Bölge">
      text = <"Anüs">
    >
```

Box 4.12 The English and Turkish definitions of MST Colon Archetype

After modeling the main organs which are colon, esophagus, stomach and duodenum according to MST structure and using English terms, the Turkish terms were also added as a second language in the ontology section. This enabled the dual-language representation of the model shown in Box 4.12 which will bring about multilinguality in HIS and also establish language independence of collected data across borders.

In the study term bindings to UMLS Metathesaurus version of MST (MTHMST2001) as described by Tringali, Hole, and Srinivasan (2002), SNOMED (2007) and UMLS (2007) were performed. This will be extremely important for linking of collected data with medical knowledge sources.

As we had stressed throughout the thesis, one of our goals was to develop HIS that is easy and feasible to maintain. Since MST Archetype model provides the means to extend the terms and the structure of the domain model without breaking the original semantics by specialization, it will be possible to extend MST for future needs by domain experts using high level tools. This extension is more than likely because as the name implies it is a minimal list of terms to describe endoscopic findings and that many users will extend for their local use. Archetype modeling will not only enable extension of MST but also preserve the original MST thus seamlessly provide data

consolidation and interoperability. This will clearly reduce the time and effort during maintenance of MST based HIS. A graphical overview of resulting MST Archetype Model is given in Figure 4.15.

4.3.6 Validation of the MST Archetype Model

All MST terms and structure has been modeled by using openEHR Archetypes and Multi-level Modeling methodology. The extended MST hierarchy has been preferred during modeling and it has been demonstrated to completely cover the original MST structure and semantics.

More technically the resulting Archetypes that are components of the top level endoscopic record composition were tested by using the openEHR Archetype Workbench which complies with ADL and also utilizes appropriate Reference Model classes to formally validate the syntax of Archetypes. The Archetype Workbench also validates the ontology section of Archetypes which contain MST terms and links to external knowledge sources. All of MST Archetypes have been successfully validated by using this tool.

Further validation studies may be conducted after implementation of a gastrointestinal endoscopy HIS by using MST Archetypes and multi-level modeling and development methods which are mentioned in Chapter 5 as future-work.

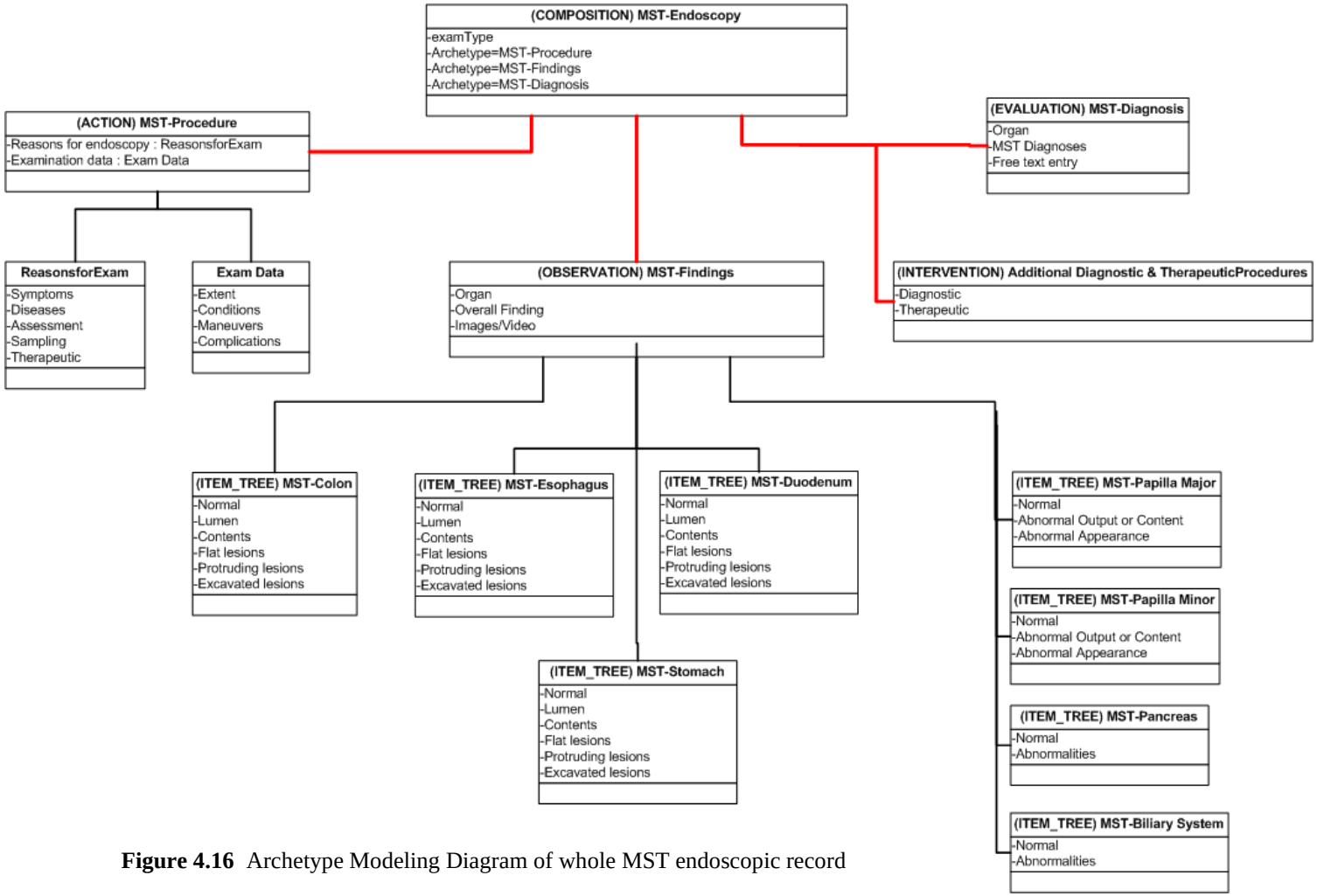

Figure 4.16 Archetype Modeling Diagram of whole MST endoscopic record

CHAPTER 5

CONCLUSIONS

In this study we have identified the separation of domain knowledge from software code as the primary approach for alleviating difficulties in HIS development related with affordability/feasibility, usability, maintainability, longevity and interoperability. In current practice, the domain knowledge for a particular medical field is directly hard-coded into software code, user interfaces and database schema. When functional requirements change due to changing or emerging domain concepts, the full development cycle starts again which leads to problems of high cost of development and maintenance, late delivery and degraded usability due to latency in the incorporation of user feedback. Consequently the lifespan of the HIS is also shortened as maintenance costs override cost of new HIS development. Classical methods which prove to be successful in other areas unfortunately fall short in the healthcare area due to the size, complexity and high variability of medical practice. Interoperability also becomes extremely difficult to achieve with such development methods because as the requirements change it is hard to keep the structure and semantics of HIS in control. Therefore it was evident that we needed efficient methods to handle these problems.

Gastrointestinal endoscopy has been selected as the medical domain in the study. It is a small and manageable domain where very good level of common medical language has already been reached. MST which is the official terminology Worldwide is available in many languages including Turkish. MST not only contains domain terms

but also depicts a sufficient level of domain structure which would help us for conducting our research.

We have developed a research prototype application, GASTROS, based on MST to capture initial requirements and also to get user feedback for better usability. It turned out to be extremely useful for this purpose and helped for detection of problems with MST and also extension. Before proceeding to the next phase of the research, we have conducted a validation study on the coverage and acceptance of MST by endoscopists. Large set of data were collected during three years and analyzed. The positive results encouraged us to proceed.

After an extensive literature survey, we have evaluated a number of modeling methodologies. openEHR Archetypes and Dual-Level or Multi-Level Modeling methodology has been selected as the fundamental focus of this study. The true separation of data, information and knowledge while modeling HIS is a paradigm shift. Common information items and some domain knowledge unlikely to change constitute the reference model (RM). All the data collected and processed by the resulting HIS are instances of this RM so a detailed data model is not needed; just a simple generic schema is sufficient. One big novel aspect in openEHR methodology is the modeling of highly volatile domain knowledge by a formalism called Archetypes. Domain experts may use high level tools (i.e. Archetype editors) to glue together RM items to assemble valid clinical concepts to be used in the HIS, either in the application logic or presentation layers. Archetype modeling of volatile domain knowledge by domain experts which drives the function and appearance of HIS is a paradigm shift and it is proposed to decrease the cost, effort and time to build HIS. It may also significantly improve maintainability and thus provide a longer life to the HIS. Because all Archetype domain models use the same or a compatible RM and all data are instances of this RM, a high level of interoperability can be reached. We can now safely conclude that openEHR approach provided efficient methods to reach our research goals.

We then started with modeling of MST, we have detected and reported a number of problems and corrected them. We also extended the modeling methodology by introducing novel concepts and methods. This study is novel because we have used a methodology which was developed for a different area, EHR, and applied the same methodology in another area - modeling and development of a specialized HIS.

Finally we have modeled the whole endoscopic examination defined by MST and also our extensions. While building the model we have consulted domain experts and collected key knowledge about MST concepts such as existence, cardinality and occurrences, and also their interrelations (i.e. mutual exclusivity or conditional existence). This should enable full computer processing capability for both HIS development and also in data processing. By this process, MST which is a medical terminology system has now become a formal medical ontology.

5.1 Contributions of the Study

It should be noted that this study is an interdisciplinary one covering both IS and Medical domains so contributions have been attempted in both domains.

5.1.1 Development of a Research Prototype Application (GASTROS)

For the purpose of validating MST and having a hands on experience with the research domain we have decided to develop a prototype application based on MST terms and structure for reporting endoscopy examinations to be used in a real clinical setting. It was an invaluable tool for further refining initial requirements to build a usable gastrointestinal endoscopy information system and provided us with insight for conducting further stages of the study.

We had to modify the original hierarchy prescribed in MST so as to design more intuitive and user-friendly data entry forms and persistence layer. It was evident that the physicians preferred our novel modeling strategy over original MST one as GASTROS has been used exclusively at the unit during the research period.

MST model and other extra information were directly modeled into application code, graphical user interface (GUI) and database schema. This approach turned out to be extremely problematic during development and especially in maintenance phase because the functional requirements arising from domain concepts were on constant change. Therefore it has shown us what shouldn’t be done when designing and implementing such systems from a technical point of view (i.e. embedding highly complex clinical model into software code and database schema).

5.1.2 MST Related Contributions

During analysis of MST, substantial amount of problems related with content were detected which were mostly semantic errors. For example some attribute values were not appropriate for the attributes or at some places attributes defining the position of a term were wrongly placed in the MST hierarchy. These errors were corrected by consulting to domain experts.

The most prominent problem with MST was related with its structure. The depicted hierarchy for describing findings was neither appropriate nor capable of correctly expressing all findings. Also the content of MST did not follow this hierarchy at all times. So we had to modify the hierarchy by first splitting the endoscopic observation and interventions and then linked the special attribute for anatomic sites directly to terms describing findings and interventions.

After the installation of the research prototype at the endoscopy unit, some extensions to MST content were made. These include addition of new anatomic sites, new terms, attributes and attribute values for describing findings and interventions and also a whole new section which was not present in MST before.

We have decided to first test the practical usability and acceptance of MST before proceeding further with the study. The experience gained during initial period of prototype usage encouraged us to collect more data and perform a large scale validation study. After three years of uninterrupted live usage, the data collected by the research prototype were analyzed in accordance with previous validation studies and has been validated formally. This validation study is the first one to evaluate second version of MST in literature and the results indicate a clear improvement over previous version.

5.1.3 Modeling Methodology Related Contributions

The modeling of a clinical domain with openEHR broadly involves selection of an appropriate reference model and then modeling the domain knowledge by a novel knowledge representation formalism called Archetypes. The Archetypes are defined by a formal language called Archetype Definition Language (ADL). RM classes are

used as building blocks to assemble domain terms and structure into valid clinical concepts. Links to external knowledge sources can also be declared.

It should be noted that this modeling methodology was quite new at the beginning of our modeling work and the studies based on it were mostly experimental. During the MST Archetype modeling process, we have encountered several problems and also did some extensions. After communication with the openEHR community, some of our contributions have been incorporated into the formal specifications.

Extensions to openEHR Archetype Definition Language (ADL)

We have identified a problem in ordering of clinical concepts in the ADL parsing rules when internal references (use_node) are used in container attributes. This problem became apparent while modeling MST Findings with many children nodes including internal references. The precise order of various nodes containing elements, complex types and other internal references could not be set.

It was not possible to constrain the occurrences of internal references (use_node) at referenced point. So we have introduced a new method for further constraining occurrences of internal references at referencing location. The method overrides original "*occurrences*" property of reference items at referencing location(s). This has enabled defining the semantics for occurrences on internal references correctly.

Another contribution was to extend Archetype metadata to include bibliographic information and other knowledge sources' references. Since the Archetypes are strong candidates for sharing of information and knowledge in clinical medicine, it is essential that they have references to existing published and online sources. We have extended the ADL metadata to include an extensive bibliographic section consisting of information items common to all kinds of knowledge sources (such as articles, books, patents, online repositories, clinical guidelines and so on). By this extension it is now possible to provide users with recent and relevant published papers about a particular subject in the Archetype.

Archetype Ontology Section has been extended so as to reference terms directly to the National Library of Medicine Unified Medical Language System (UMLS) which is a gateway to internationally significant knowledge sources; such as coding, classification and terminology systems, medical ontologies, clinical repositories and

even some software system's databases (Lindberg, 1990). The problem was it was not possible to reference to correct UMLS terms in Archetype ontology section. Our extension includes storing both CUI and Term UI in MST Archetypes in appropriate subsections in the ontology section.

Another contribution includes a methodological approach to model domain concepts with encapsulated non-textual data such as X-ray or simple photographic images. In current HIS implementations usually the textual and multimedia part of report is handled separately and the non-textual part of examination data is simply stored as binary objects which render the structured textual information within them inaccessible. In openEHR this kind of encapsulated data are represented by the generic DV_ENCAPSULATED or DV_MULTIMEDIA data types in An archetype which makes access to embedded structured data very difficult if not impossible. Our methodology consists of introducing dynamic runtime references from the definition section of Archetypes to the encapsulated data which contains parsable structured textual data. Our novel approach enables Archetypes to model such situations effectively, without loosing data and without altering the current semantics of openEHR specifications too much.

Extensions to openEHR Reference Model (RM)

In clinical medicine, recording reasons of missing data such as not asked, patient not able to answer, patient do not answer, doctor has forgotten and so on are also important for clinical decision making. openEHR RM contains "Flavors of Null" to represent this in leaf nodes but it was still not possible to record reasons of missing data about top level structures. We have introduced "Flavors of Null" for missing data into top level RM structures. The advantages of this method include reducing size, improving manageability and increasing understandability of Archetypes, efficient querying of leaf-nodes in a huge repository, and better adherence to standards by confining implementation strategy within the standardization body.

We have also relaxed the optionality of "items" attribute of CLUSTER in openEHR Data Structures RM. It is now possible to have clusters without any child nodes but yet still able to have Null_Flavor value to check during runtime for presence of a value without further specifying an ELEMENT Class solely for this purpose. This is

also important in semantic equivalence during transformation of equivalent Archetypes when no child nodes are left due to deletion of internal references.

Creation of Structural Archetypes

It is possible to chain Archetypes together via allowing Archetypes to be consumed by others. The mechanism works by providing external references. However when an Archetype is referencing other Archetypes having same contextual information, the necessity of repeating this information and the possibility of having differences does not make any sense and it may even violate data integrity. In our approach such contextual information only exists in the top level Archetype and the so called structural Archetypes being referenced only provide the needed subparts. Structural Archetypes are devoid of all contextual information and some of the mandatory RM classes. These Archetypes contain only the necessary data structures and values for domain modeling.

5.1.4 Defined Semantic Equivalence of Archetype Models

When dealing with complex domains, modelers usually follow the natural/original structure or hierarchy of domain concepts. This is usually the format of original article, consensus report or as taken from common practice. However alternative models describing the same domain might be possible and at times necessary. We have discovered that different models might represent the same domain and thus can be considered to have “semantic equivalence”.

In openEHR there was no notion about this paradigm. Our solution consists of first discovering alternative domain models, manual identification of semantic equivalence and then followed by a novel Archetype modeling of these domains. The solution also provides methods to effectively manage these Archetype models. These methods will enable transformation of semantically equivalent Archetype models and also aid in data integration and consolidation

5.1.5 MST Archetype Modeling

This is the last and most important part of the study. We are convinced that our approach, after making significant contributions in both MST and modeling methodology, is capable of reaching the objective and goals of the study. As the proof of acceptance of our modeling strategy the initial MST-Colon archetype model has been published electronically by the openEHR society.

The modeling task consisted of first mapping MST hierarchy onto existing openEHR EHR RM and then deciding on appropriate data structures, types and also some auxiliary models like support and common RM. The major part of MST is composed of endoscopic findings which were modeled as a single Observation Archetype. Findings for each organ according to the examination type were modeled by Structural Archetypes all of which are embraced by the top level Observation Archetype. MST diagnoses were modeled by Evaluation Archetypes, MST interventions by Intervention Archetypes and all the remaining MST parts by Action Archetypes. Finally all these Archetypes constituting parts of MST were collected under a single Composition Archetype which defined a whole valid endoscopic examination that can be used to represent a formal clinical report.

After modeling of the findings for different organs using English terms, the Turkish terms were also added as a second language in the ontology section. This is important for enabling multilinguality in HIS and also establishes language independence of collected data. Term bindings to UMLS and SNOMED-CT were also performed which is important for linking of collected data with medical knowledge sources.

Our goals included better maintainability and interoperability of endoscopic HIS and we strongly believe these goals can be reached by the MST Archetype model. It provides the means to extend the terms and the structure of the domain model so it is possible to extend MST for future needs by domain experts. Archetype modeling not only enables extension of MST but also preserves the original semantics by a method called Archetype specialization, thus it will enable easy data consolidation and interoperability. Now that the software will be based on more stable requirements, it would be expected to require less effort and resources to develop and maintain HIS.

One important point related with modeling we want to emphasize is that our novel MST Archetype model differs from existing studies due to the fact that a whole clinical domain is modeled here in contrast with other studies where only particular

concepts are modeled. In other words the modeling methodology had been designed for the purpose of establishing a standard lifetime EHR for a single individual but we have extended it to be able to model a complex clinical domain in which the developed HIS can operate by itself and may act as a feeder system to the EHR if available. It was extremely important to demonstrate that the RM and Archetype modeling methodology was capable of meeting the needs of our research.

5.2 Auto-critique and Future Work

To reach the objective of the study we have set a number of goals and conducted our research accordingly. All specific goals of the study listed in Section 1.2 have been reached. Validation of the MST and also refining of initial requirements for development of an endoscopic HIS were performed by the research prototype which was an important step. However it took much more time and effort to reach this stage than we have anticipated. In addition the modeling methodology we have selected, openEHR Archetype domain modeling, turned out to be very immature to meet the needs of our modeling task at the time we had started in year 2003. Therefore it was not possible to develop the next-generation endoscopic HIS by using Multi-level Modeling technique and MST archetypes. However we intend to accomplish this as further research. We came to the conclusion that first our objective was over-ambitious and the goals were difficult to reach in the given time and with available resources. Secondly the inevitable delays due to problems in MST and serious limitations in the modeling methodology also played a role in this outcome.

It might be criticized that comparative modeling with alternative modeling formalisms has not been performed. However it should be noted that Archetypes and multi-level modeling formalism is quite new and alternative methodologies were either too immature to evaluate or did not meet the needs of our goals from the start. For example the closest approach, HL7 v3, has been released quite recently and many problems still exist. The classical OO or OR methods do not permit separation of knowledge from information explicitly.

As the immediate future work, development of an endoscopic HIS based on the methodology described in this thesis might be done. We must stress that reference implementations have just recently been started and that development of a full-fledge

HIS via openEHR formalism has not been performed. Therefore we foresee that this might be a high impact research in the field.

It can be observed that in this study a terminology system authored by domain experts coming from the medical society has been transformed into computationally usable domain ontology by using the methodology underlined in this thesis. Apart from developing HIS, the methodology also offers solutions for the governance of the terminology content. These include the possibility of local extensions to the terminology without breaking original semantics by Archetype specialization and also establishment of regional or global Archetype repositories. Another short term future work is realization of this by collaborating with the associated medical society.

Now that HL7 v3 CDA and a more mature RIM are available, comparative modeling might be performed as another future work. Although HL7 has originated to establish messaging among disparate HIS, this version includes several methodologies to model clinical and administrative domains and thus establish interoperability. It might be interesting to investigate the proposed capability of this message based paradigm in HIS modeling. The results of this study might shed light onto evolution of this novel approach.

Other future work may include performing various studies on HIS modeling strategies other than multi-level modeling such as refining Object Oriented methodology specifically for HIS. Automatic code generation from UML models might be explored for providing the benefits mentioned in this study.

Currently we are able to model the static aspects of an HIS; that is the knowledge and operational rules. The dynamic aspects of an HIS including workflows and other program flow due to user input or external events need to be considered for fully realizing a workable HIS from the model. This would highly decrease the cost, effort and time when developing HIS. The foundations of workflow modeling work have been set forth by openEHR society and various academic studies (Barretto, 2005). This is a very broad area and has potential for a wide range of future work in the long term.

Multi-level modeling and Archetypes cover a considerable part of an HIS during design. UML is also being used widely for HIS development. Both approaches overlap with each other and both are useful. So a hybrid modeling notation is needed for

providing developers with a framework in HIS development. This is also expected to be a high impact research work for future.

Narrowing the gap between genetics and medicine requires establishment of common understanding among both researchers and information systems. Currently vast amount of genomic data, such as Human Genome Project (2007) and also disease or population specific databases are available. However these neither relate to individuals and their clinical outcomes nor to environmental factors. Integration of individual clinical data captured by HIS with genomic databases is essential for finding real cure for many diseases and to design personalized drugs and treatments for individuals. The modeling formalism might be extended so as to model genetic domain and hence enable this integration. This future research is of paramount importance.

BIBLIOGRAPHY

Accredited Standards Committee (ASC X12). Virginia, USA. Retrieved April 24, 2007, from http://www.x12.org.

American Medical Association (AMA). Chicago, USA. Retrieved April 24, 2007, from http://www.ama-assn.org.

American National Standards Institute (ANSI). Washington DC, USA. Retrieved April 24, 2007, from www.ansi.org.

American Society for Testing and Materials (ASTM) Standards. ASTM International, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA, 19428-2959 USA. Retrieved April 24, 2007, from http://www.astm.org.

Armour, P.G. (2000). The Five Orders of Ignorance. *Communications of the ACM, 43(10),* 17-20.

Atalağ, K. & Bilgen, S. (2007a). Multi-level Modeling and the Role of Archetypes in the Design of Health Information Systems: A Modeling Example in Endoscopy. In L.E. Akman, N. Baykal, T. Can, M. Iscan & U.E. Mumcuoglu (Eds.), *HIBIT '07 Proceedings of the International Symposium on Health Informatics and Bioinformatics, April 30 - May 2, 2007*. Ankara, Turkey: Middle East Technical University (METU) Press.

Atalağ, K., Bilgen, S., Gür, G. and Boyacıoğlu S. (submitted). Evaluation of the Turkish translation of the Minimal Standard Terminology for Digestive Endoscopy by development of an endoscopic information system. *The Turkish Journal of Gastroenterology.*

Ball, M.J. & Lillis, J. (2001). E-health: transforming the physician/patient relationship. *International Journal of Medical Informatics, 61*, 1-10.

Barretto, S.A. (2005). *Designing Guideline-based Workflow-integrated Electronic Health Records*. Ph.D. Thesis, University of South Australia.

Baud, R.H., Rassinoux, A.M. & Scherrer J-R. (1992). Natural language processing and medical records. In K.C. Lun et al. (Eds.), *Medinfo 92 (*pp. 1362-1367). Amsterdam: Elsevier Science Publishers.

Beale, T & Heard, S. (2006). *openEHR Architecture - Architecture Overview*. Retrieved March 3,2007, from http://svn.openehr.org/specification/BRANCHES /Release-1.0.1-candidate/publishing/architecture/overview.pdf.

Beale, T. (2000). Archetypes - *An Interoperable Knowledge Methodology for Future-proof Information Systems*. Whitepaper, openEHR Foundation. Retrieved March 15, 2007, from http://openehr.org/publications/archetypes /archetypes_beale_web_2000.pdf.

Beale, T. (2001). *Health Information Standards Manifesto*. Whitepaper, V2.5, Dec 2001. Retrieved April 1, 2007, from http://www.deepthought.com.au/health/ HIS_manifesto/his_manifesto.pdf.

Beale, T. (2002). Archetypes: Constraint-based domain models for future-proof information systems. In K. Baclawski & H. Kilov. (Eds.), *Eleventh OOPSLA Workshop on Behavioral Semantics: Serving the Customer, Seattle, Washington, USA, November 4, 2002,* (pp. 16-32). Boston: Northeastern University Press.

Beale, T. (2003). *A Model Universe for Health Information Standards*. Retrieved Feb, 12, 2007, from http://www.deepthought.com.au.

Beale, T. (2003). *A Comparison of Health Information Specifications using a Model Universe Approach*. Retrieved Feb 12, 2007, from http://www.deepthought.com.au.

Beale, T. (2005). EHR section synopsis: The Health Record - Why is it so Hard? In R. Haux & C. Kulikowski (Eds), *IMIA Yearbook of Medical Informatics 2005: Ubiquitous Health Care Systems* (pp. 301-304), Stuttgart: Schattauer.

Beale, T., Heard, S. (2007). An Ontology-based Model of Clinical Information. Submitted to *MedInfo Conference 2007*.

Beale, T. (2007). *The openEHR Modelling Guide*. Retrieved April 1, 2007, from http://svn.openehr.org/specification/BRANCHES/Release-1.0.1-candidate/publishing/architecture/modelling_guide.pdf.

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

Blobel, B. (2006). Advanced and secure architectural EHR approaches. *International Journal of Medical Informatics, 75,* 185-190.

Bodenreider, O. Willis, J. & Hole, W. (2004). The Unified Medical Language System: What is it and how to use it? *Presentation at the 11th World Congress on Medical Informatics, MEDINFO 2004, September 8, 2004*; San Francisco, CA, USA.

Boehm, B. W. (1988). A Spiral Model of Software Development and Enhancement. *IEEE Computer, 21(5),* 61-72.

Brooks, F. P. (1987). No Silver Bullet: Essence and Accidents of Software Engineering. *IEEE Computer, 20,* 10-19.

Camp, M.G. (1996). Problem-Based Learning: A Paradigm Shift or a Passing Fad? *Medical Education Online, 1(2),* 1-6.

CEN EN 12967-1 (2006) Health Information Services Architecture (HISA) Standard. The European Committee for Standardization, Brussels, Belgium.

CEN EN 13606-1 (2007) EHR Communications (EHRcom) Standard. The European Committee for Standardization, Brussels, Belgium.

CEN EN 13940-1 (2006) System of concepts to support continuity of care (ContSys) Standard. The European Committee for Standardization, Brussels, Belgium.

CEN/TC251 (2007). European Committee for Standardization Technical Committee 251: Technical Committee of Health Informatics. Retrieved July 6, 2007, from http://www.centc251.org.

Charette, R. (2006). EHRs: Electronic Health Records or Exceptional Hidden Risks? *Communications of the ACM, 49(6),* 120.

Chaudhry, B., Wang, J., Wu, S., Maglione, M., Mojica, W., Roth, et. al. (2006). Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Annals of Internal Medicine, 144(10),* 742-52.

Coiera, E. (2003). *Guide to Health Informatics, 2nd Edition*. England: Hodder & Stoughton.

Continuity of Care Record (CCR). The concept paper of the CCR. Retrieved April 5, 2007, from www.astm.org/COMMIT/E31_ConceptPaper.doc.

Cooper, G. & Sivak, M. (2000). More on endoscopy databases. *Gastrointestinal Endoscopy, 51(6),* 759-761.

Crespi, M., Delvaux, M., Schapiro, M., Venable, C. & Zwiebel, F. (1996). Working party report by the Committee for Minimal Standards of Terminology and Documentation in Digestive Endoscopy of the European Society of Gastrointestinal Endoscopy. Minimal standard terminology for a computerized endoscopic database. *The American Journal of Gastroenterology, 91,* 191-216.

Dairo, A., Giuse, A. & Kuhn, K.A. (2003). Health information systems challenges: the Heidelberg conference and the future. *International Journal of Medical Informatics, 69*, 105-114.

de Clercq, P.A., Blom, J.A., Korsten, H.H.M. & Hasman, A. (2004). Approaches for creating computer-interpretable guidelines that facilitate decision support. *Artificial Intelligence in Medicine, 31,* 1-27.

de Lange, T., Moum, B.A., Tholfsen, J.K., Larsen, S. & Aabakken, L. (2003). Standardization and quality of endoscopy text reports in ulcerative colitis. *Endoscopy, 35(10),* 835-840.

DeLone, W. H. & McLean, E. R. (1992). Information systems success: The quest for the dependent variable. *Information Systems Research, 3(1),* 60-95.

Delvaux, M. (2000b). Minimal standard terminology in digestive endoscopy. *Endoscopy , 32(2),* 162-88.

Delvaux, M., Crespi, M., Armengol-Miro, J.R., Hagenmüller, F., Teuffel, W., Spencer, K.B., et al. (2000a). Minimal standard terminology for digestive endoscopy: Results of prospective testing and validation in the GASTER project. *Endoscopy, 32(4),* 345-355.

Delvaux, M., Korman, L.Y., Armengol-Miro, J.R., Crespi, M., Cass, O., Hagenmüller, F, et al. (1998). The minimal standard terminology for digestive endoscopy: Introduction to structured reporting. *International Journal of Medical Informatics, 48,* 217-225.

Digital Imaging and Communications in Medicine Standard (DICOM). National Electrical Manufacturers Association (NEMA). Virginia, USA. Retrieved April 24, 2007, from http://dicom.nema.org.

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

Dorr, D., Bonner, L.M., Cohen, A.N., Shoai, R.S., Perrin, R., Chaney, E., et al. (2007). Informatics Systems to Promote Improved Care for Chronic Illness: A Literature Review. *Journal of the American Medical Informatics Association, 14(2),*156-163.

eHealth Action Plan. Retrieved April 9, 2007 from http://europa.eu.int/information_society/activities/health/policy_action_plan/index_en.htm.

Eichelberg, M., Aden, T., Riesmeier, J., Dogac, A. & Laleci, G.B. (2005). A Survey and Analysis of Electronic Healthcare Record Standards. *ACM Computing Surveys, 37(4)*, 277-315.

European Committee for Standardization (CEN). Brussels, Belgium. Retrieved April 24, 2007, from http://www.cen.eu/cenorm/index.htm.

Fernandez, E. & Sorgente, T. (2005). An analysis of modeling flaws in HL7 and JAHIS. In H. Haddad, L.M. Liebrock, A. Omicini & R.L. Wainwright (Eds.), *Proceedings of the 2005 ACM Symposium on Applied Computing (SAC), New Mexico, USA, March 13-17* (pp.216-223).

Fernández-Breis, J.T., Menárguez-Tortosa, M., Moner, D., Valencia-García, R., Maldonado, J.A., Vivancos-Vicente, P.J., et al. (2006). An Ontological Infrastructure for the Semantic Integration of Clinical Archetypes. In A. Hoffmann, B-H. Kang, D. Richards & S. Tsumoto (Eds.), *Advances in Knowledge Acquisition and Management: Pacific Rim Knowledge Acquisition Workshop, Pkaw 2006 Guilin, China, August 7 - 8, 2006 Revised Selected Papers* (156-167).

Ferranti, J.M., Musser, R.C., Kawamoto, K. & Hammond, W.E. (2006). The Clinical Document Architecture and the Continuity of Care Record: A Critical Analysis. *Journal of the American Medical Informatics Association, 13*, 245–252

Fowler, M. (2004). *UML Distilled: A Brief Guide to the Standard Object Modeling Language, 3rd Edition*. USA: Addison-Wesley.

Fowler, M. Language Workbenches: The Killer-App for Domain Specific Languages? Retrieved 24 April 2007, Last update June 12, 2005, from http://www.martinfowler.com/articles/languageWorkbench.html.

Garde, S., Knaup, P. & Hovenga, E. (2005b). openEHR Archetypes in Electronic Health Records: the Path to Semantic Interoperability? In R. Klar, W. Köpcke, K. Kuhn, H. Lax, S. Weiland & A. Zaiß (Eds.), *Tagungsband der 50. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie, 12.-15. September 2005 und 12. Jahrestagung der Deutschen Arbeits-gemeinschaft für Epidemiologie, 14.-15. September 2005, Freiburg im Breisgau* (263-265). Düsseldorf, Köln: German Medical Science.

Garde, S., Knaup, P., Hovenga, E.J.S. & Heard, S. (2007). Towards Semantic Interoperability for Electronic Health Records. *Methods of Information in Medicine, 46(3)*, 332-343.

Garde, S., Knaup, P., Schuler, T. & Hovenga, E. (2005a). Can openEHR Archetypes Empower Multi-Centre Clinical Research? In R. Engelbrecht, A. Geissbuhler, C.

Lovis, G. Mihalas (Eds.), *Connecting Medical Informatics and Bio-Informatics: Section 13: Public Health Informatics, Clinical Trials. ENMI European Notes in Medical Informatics,I(1), 971-976.* Amsterdam: IOS Press.

Garson, A. (2004). U.S. Healthcare: The Intertwined Caduceus of Physicians, Coverage, Quality, and Cost. *Journal of the American College of Cardiology, 43(1),* 1-5.

GNUmed Project. Retrieved April 23, 2007, from http://www.gnumed.org.

Godlee, F., Pakenham-Walsh, N., Ncayiyana, D., Cohen, B. & Packer, A. (2004). Can we achieve health information for all by 2015? *Lancet, 364*, 295–300.

Grassi, A. & Delvaux, M. (2003). Minimal Standards in Endoscopy Reporting: Structure of the Records, Terms, Definitions and Photo Documentation. *World Gastroenterology News - Spring 2003*. Retrieved March 22, 2007, from http://www.omge.org/publications/archive/2003_1/omed/omed2.htm.

Grimshaw, J.M., Russell, I.T. (1993). Effect of clinical guidelines on medical practice ? a systematic review of rigorous evaluations. *Lancet, 342,* 317-322.

Grimshaw, J.M., Russell, I.T. (1994a). Achieving health gain through clinical guidelines I: Developing scientifically valid guidelines. *Quality in health care, 2,* 243-248.

Grimshaw, J.M., Russell, I.T. (1994b). Achieving health gain through clinical guidelines I: ensuring that guidelines change medical practice. *Quality in health care, 2,* 243-248.

Grimson, J. (2001). Delivering the electronic healthcare record for the 21st century. *International Journal of Medical Informatics, 64*, 111–127.

Groenen, M.J.M., Kuipers, E.J., van Berge Henegouwen, G.P., Fockens, P. & Ouwendijk, R.J.Th. (2006). Computerisation of endoscopy reports using standard reports and text blocks. *The Netherlands Journal of Medicine, 64(3),* 78-83.

Hammond, W.E. (2002, December). *What if we really had an electronic health record?* Presented at, EuroRec 2002 Conference, European Institute for Health Records, Berlin, Germany.

Hammond, W.E. (2003). Making the boundaries clearer: revisiting information systems with fading boundaries. *International Journal of Medical Informatics, 26,* 99-104.

Harden, R.M. (2005). A New Vision for Distance Learning and Continuing Medical Education. *The Journal of Continuing Education in the Health Professions, 25,* 43-51.

HARP Project - Harmonisation for the secuRity of web technologies and applications. EU Contract No: IST-1999-10923. Retrieved April 23, 2007, from http://www.telecom.ntua.gr/~HARP/HARP/HARP.htm.

Haux, R. (2006). Health information systems— past, present, future. *International Journal of Medical Informatics, 75*, 268-281.

Health Level 7 Organization (HL7). Retrieved April 24, 2007, from http://www.hl7.org.

Health Level 7 Reference Information Model (HL7 RIM). Retrieved April 10, 2007 from http://www.hl7.org/library/data-model/.

Heeks, R., Mundy, D. & Salazar, A. (2007). *Why Health Care Information Systems Succeed or Fail*. Information Systems for Public Sector Management - Working Paper Series Paper No. 9. Institute for Development Policy and Management.

Hovenga, E., Garde, S. & Heard, S. (2005). Nursing constraint models for electronic health records: A vision for domain knowledge governance. *International Journal of Medical Informatics, 74*, 886-898.

Human Genome Project. Retrieved July 7, 2007 from National Human Genome Research Institute site on the World Wide Web: http://www.nhgri.nih.gov.

Humphreys, B.L. & Lindberg, D.A.B. (1993). The UMLS project: making the conceptual connection between users and the information they need. *Bulletin of the Medical Library Association 81(2)*, 170.

Institute of Electrical and Electronic Engineers, Inc. (IEEE). Retrieved April 24, 2007, from http://www.ieee.org.

Institute of Medicine (IOM), Committee on Improving the Patient Record. (1997). Dick, S.R., Steen, E. B. & Detmer, D.E. (Eds.). *The Computer-Based Patient Record: An Essential Technology for Health Care, Revised Edition*. Washington, D.C.: National Academy Press.

Institute of Medicine (IOM), Committee on Quality of Health Care in America. (2000). Kohn, L. T., Corrigan, J. M. & Donaldson, M. S. (Eds.). *To Err is Human: Building a Safer Health System.* Washington D.C.: National Academy Press.

Institute of Medicine (IOM), Committee on Quality of Health Care in America. (2001). *Crossing the Quality Chasm*. Washington D.C.: National Academy Press.

Institute of Medicine of the National Acedemies (IOM). Retrieved April 23, 2007, from http://www.iom.edu.

Integrating the Healthcare Enterprise (IHE). Retrieved April 24, 2007, from http://www.ihe.net.

International Organization for Standardization (ISO). Geneva, Switzerland. Retrieved April 24, 2007, from http://www.iso.org

ISO TR 20514:2005 Health Informatics - Electronic Health Record Definition, Scope and Context Standard. International Organization for Standardization (ISO), Geneva, Switzerland.

ISO TS 18308:2004 Health informatics -Requirements for an electronic health record architecture. International Organization for Standardization (ISO), Geneva, Switzerland.

ISO/IEC Guide 2 (1996) Standardization and related activities - General vocabulary. International Organization for Standardization (ISO), Geneva, Switzerland.

ISO/TC215 (2007). International Organization for Standardization Technical Committee 215 "Health Informatics". Retrieved July 6, 2007, from http://isotc.iso.org/livelink/livelink.exe?func=ll&objId=529137&objAction=browse&sort=name.

i2010 - A European Information Society for growth and employment. Retrieved April 9, 2007 from http://europa.eu.int/information_society/eeurope/i2010/index_en.htm.

Johan, F., Hoorn, J. F., Konijn, E. A., van Vliet, H. & van der Veer, G. (2007). Requirements change: Fears dictate the must haves; desires the won't haves. *Journal of Systems and Software, 80*, 328-355.

Kalra, D. (2003). *Clinical Foundations and Information Architecture for the Implementation of a federated Health Record Service*. Ph.D. Thesis, University College London, 2003.

Kalra, D., Austin, A., O'Connor, A., Patterson, D., Lloyd, D. & Ingram, D. (2001). Design and Implementation of a Federated Health Record Server. *Toward an Electronic Health Record Europe (TEHRE) 2001 m-Health Conference Proceedings, 1st Annual Conference on Mobile & Wireless Healthcare Applications, 11 - 14 November 2001, London, UK* (pp. 1-13). Medical Records Institute for the Centre for Advancement of Electronic Records Ltd.

Katehakis, D.G. (2005) The PICNIC Story. In S. Saranummi, D. Piggott, D.G. Katehakis, M. Tsiknakis & K. Bernstein (Eds.), *Regional Health Economies And ICT Services: The Picnic Experience, Studies in Health Technology and Informatics* , 115,4-36.

Klein, D. & Ware, M. (2003). E-learning: new opportunities in continuing professional development. *Learned Publishing, 16,* 34-46

Kontrac Whitepaper (2003): *From coding to modelling: Past, Present and Future (?).* Retrieved April 24, 2007, from http://www.kontrac.com/legacy/pdf/modeling.pdf.

Korman, L.Y. & Bidgood, W.D. Jr. (1997). Representation of the Gastrointestinal Endoscopy Minimal Standard Terminology in the SNOMED DICOM microglossary. *Proceedings : a conference of the American Medical Informatics Association / ... AMIA Annual Fall Symposium. AMIA Fall Symposium,* (pp.434-438). Los Angeles, CA: Hanley & Belfus.

Korman, L.Y., Delvaux, M. & Crespi, M. (2001). The minimal standard terminology in digestive endoscopy: perspective on a standard endoscopic vocabulary. *Gastrointestinal Endoscopy, 53,* 392-396.

Korman, L.Y., Delvaux, M., Bidgood & W.D. Jr. (1998). Structured reporting in gastrointestinal endoscopy: Integration with DICOM and minimal standard terminology. *International Journal of Medical Informatics, 48,* 201-206.

Kuperman, G.J. & Gibson, R.F. (2003). Computer Physician Order Entry: Benefits, Costs, and Issues. *Annals of Internal Medicine, 139,* 31-39.

Langlois, B., Exertier, D. & Devda, G. (2006). Toward Families of QVT DSL and Tool. In J. Gray, J.P. Tolvanen & J. Sprinkle (Eds.), *Proceedings of the 6th OOPSLA Workshop on Domain-Specific Modeling (DSM'06)* (pp:87-97).

Liebowitz, J. (1999). A look at why information systems fail. *Kybernetes, 28(1),* 61-67.

Lindberg, C. (1990). The Unified Medical Language System (UMLS) of the National Library of Medicine. *Journal of the American Medical Record Association, 61(5),* 40-42.

Logan, J.R. & Klopfer, K.C. (2000). The use of a standardized terminology for comparison of free text and structured data entry. *Proceedings / AMIA Annual Symposium* (pp. 512-516). Los Angeles, CA: Hanley & Belfus.

Logan, J.R., McCashland, T. & Lieberman, D.A. (2004). Evaluation of Reliability in the Use of Endoscopic Terminology. In M. Fieschi, E. Coiera & Y.-C.J. Li) (Eds.) *Proceedings of the 11th World Congress on Medical Informatics, MEDINFO 2004* (pp. 396-400). Amsterdam: IOS Press

Los, R.K., van Ginneken, A.M. & van der Lei, J. (2005). OpenSDE: A strategy for expressive and flexible structured data entry. *International Journal of Medical Informatics, 74,* 481-490.

Los, R.K., van Ginneken, A.M., de Wilde, M. & van der Lei, J. (2004). OpenSDE: Row Modeling Applied to Generic Structured Data Entry. *Journal of the American Medical Informatics Association, 11*, 162-165.

Malterud, K. (2001). The art and science of clinical knowledge: evidence beyond measures and numbers. *Lancet, 358(9279),* 397-400.

Maratka, Z. (1999). *Terminology, definitions and diagnostic criteria in digestive endoscopy. 4th Edition*. Bad Homburg: Normed Verlag.

MEDCIN®. Medicomp Systems, Inc. Retrieved March 20, 2007, from http://www.medcin.com.

Menachemi, N., Brooks, R.G. (2006). Reviewing the Benefits and Costs of Electronic Health Records and Associated Patient Safety Technologies. *Journal of Medical Systems, 30*, 159-168.

Moorman, P.W. & Bernstein, K. (Eds.). (1999). *Regional Health Care Networks in Europe. The CoCo Project Report. Deliverable 4.2*, CoCo Project, HC 1074/ HC 1008.

Moorman, P.W. (1995a). *Towards formal medical reporting: An evaluation in endoscopy*. Master of Science Thesis, Erasmus University, Rotterdam.

Moorman, P.W., Siersema, P.D., de Ridder, M.A.J. & van Ginneken, A.M. (1995b). How often is large smaller than small? *Lancet, 345(8953),* 865.

Moorman, P.W., Siersema, P.D., van Ginneken, A.M. & van Blankenstein, M. (1994). Are referring physicians satisfied with endoscopy reports? *Zeitschrift für Gastroenterologie, 32*, 623-625.

Moorman, P.W., van Ginneken A.M., van der Lei, J. & van Bemmel, J.H. (1994). A model for structured data entry based on explicit descriptional knowledge. *Methods of Information in Medicine, 33(5),* 454-463.

Moorman, P.W., van Ginneken, A.M., van der Lei, J., Siersema, P.D., van Blankenstein, M. & Wilson, J.H.P. (1994). The contents of free text endoscopy reports: An inventory and evaluation by peers. *Endoscopy, 26,* 531-538.

Munoz, A., Somolinos, R., Pascual, M., Frague, J.A., Gonzalez, M.A., Monteagudo, J.L., et al. (2007). Proof-of-concept Design and Development of an EN13606-based Electronic Health Care Record Service. *Journal of the American Medical Informatics Association, 14*, 118-129.

Natalya, F.N. & McGuiness, D.L. (2002). *Ontology Development 101: A Guide to Creating Your First Ontology*. (Whitepaper). Retrieved March 12, 2007, from Stanford University: http://protege.stanford.edu/publications/ontology_development/ ontology101.pdf.

Nelson, A.R. (1998). Medicine: Business or profession, art or science? *American Journal of Obstetrics and Gynecology, 178(4),* 755-758.

Noy, N.F., Fergerson, R.W. & Musen, M.A. (2000). The Knowledge Model of Protégé-2000: Combining Interoperability and Flexibility. Rose Dieng, Olivier Corby (Eds.), *Knowledge Acquisition, Modeling and Management, 12th International Conference, EKAW 2000, Juan-les-Pins, France, October 2-6, 2000, Proceedings. Lecture Notes in Computer Science, 1937,* 17-32. Europe: Springer.

Object Management Group (OMG) Healthcare Domain Task Force (HDTF). Retrieved July 6, 2007, from http://healthcare.omg.org.

Object Management Group (OMG). Retrieved April 24, 2007, from http://www.omg.org.

O'Carroll, P.W. (2002). Introduction to Publich Health Informatics. In P.W. O'Carroll, W.A. Yasnoff, W.M. Elizabeth, L.H. Ripp & E.L. Martin (Eds), *Public Health Informatics and Information Systems*. New York: Springer-Verlag.

Ocean Informatics. Retrieved April 9, 2007, from http://www.oceaninformatics.biz.

Open Source Health Care Alliance (OSHCA). Retrieved April 24, 2007, from http://www.oshca.org.

openEHR Archetype Definitions and Principles Revision 1.0. (2007). Retrieved March 14,2007, from http://svn.openehr.org/specification/TAGS/Release-1.0.1/publishing/architecture/am/archetype_principles.pdf.

openEHR Archetype Object Model (AOM) Revision 2.0.1. (2007). Retrieved March 20, 2007, from http://svn.openehr.org/specification/TAGS/Release-1.0.1/publishing /architecture/am/aom.pdf.

openEHR Foundation. Retrieved April 1, 2007, from http://www.openEHR.org.

openEHR Reference Models. Retrieved April 1, 2007 from http://svn.openehr.org/ specification/BRANCHES/Release-1.0.1-candidate/publishing/architecture/rm/.

openEHR Template Object Model Revision 0.5. (2007). Retrieved March 13, 2007, from http://svn.openehr.org/specification/TAGS/Release-1.0.1/publishing/ architecture/ am/ tom.pdf.

openEMed Project. Retrieved April 23, 2007, from http://openemed.org.

OpenGALEN Foundation. Retrieved April 1, 2007, from http://www.opengalen.org.

OpenSDE Website. Retrieved March 13, 2007, from Erasmus University, Department of Medical Informatics Website: http://www2.eur.nl/fgg/mi/OpenSDE.

Özkan, S. (2006). *PB-ISAM: A process-based Framework for Information Systems Effectiveness Assessment in Organisational Contexts*. Ph.D. Thesis, Middle East Technical University (METU) Informatics Institute.

President's Advisory Comission of Consumer Protection and Quality in the Healthcare Industry, US Government (1998) Final Report, *Quality First: Better Health Care for All Americans*. Retrieved March 19, 2007 from http://www.hcqualitycomission.gov/final. Last update on June 15, 1998.

Protégé Website. Retrieved April 3, 2007, from Stanford University School of Medicine, Stanford Medical Informatics: http://protege.stanford.edu.

Quamar, R. & Rector, A. (2007). Semantic Issues in Integrating Data from Different Models to Achieve Data Interoperability. Paper Submitted to *MEDINFO Conference 2007*.

Rector, A.L. (1999). Clinical Terminology: Why is it so hard? *Methods of Information in Medicine, 38,* 239-252.

Rector, A.L., Glowinski A.J., Nowlan, W.A. & Rossi-Mori, A. (1995). Medical-concept Models and Medical Records: An Approach Based on GALEN and PEN & PAD. *Journal of the American Medical Informatics Association, 2*, 19-35.

Rumbaugh, J., Jacobson, I. & Booch, G. (1998). *The Unified Modeling Language Reference Manual*. USA: Addison Wesley.

Sackett, D.L., Rosenberg, W.M.C., Gray, J.A.M. & Richardson, W.S. (1996). Evidence based medicine: What it is and what it isn't. *British Medical Journal, 312*, 71-72.

Shabo, A., Vortman, P. & Robson, B. (2001). Who's Afraid of Lifetime Electronic Medical Records? *In proceedings of TEHRE – Towards Electronic Health Records Conference, London, UK, November 14*.

Sittig, D.F. & Stead, W.W. (1994). Computer-based Physician Order Entry: The State of the Art. *Journal of the American Medical Informatics Association, 1*, 108-123.

Smith, B. & Ceusters, W. (2006). HL7 RIM: An Incoherent Standard. *Studies in Health Technology and Informatics,124*, 133-138.

Sommerville, I. (2000). *Software Engineering, 6th Edition*. USA: Addison Wesley.

Stolberg, H.O. (2004). The Canadian Health Care System: Past, Present, and Future. *Journal of the American College of Radiology, 1*, 659-670.

Straus, S.E. & Sackett, D.L. (1999). Applying evidence to the individual patient. *Annals of Oncology 10*, 29-32.

SNOMED - Transforming Expressions to Normal Forms – August 2006 revision of External Draft Guide. Retrieved 23 June, 2007, from http://www.snomed.org/snomedct/documents/transformations_to_normal_forms.pdf.

Systematized Nomenclature for Medicine (SNOMED). Retrieved March 11, 2007, from http://www.snomed.org.

Tang, P.C. (2003). *Key Capabilities of an Electronic Health Record System*, Letter Report. Committee on Data Standards for Patient Safety, Board on Health Care Services, Institute of Medicine.

The European Pathway Association (E-P-A.org). Retrieved March 28, 2007, from http://www.e-p-a.org/000000979b08f9803/index.html.

The Institute for Formal Ontology and Medical Information Science (IFOMIS). Retrieved March 29, 2007, from http://www.ifomis.org.

The Joint Comission. Retrieved 27 March 2007, from http://www.jointcommission.org.

The Leapfrog Group for Patient Safety. Retrieved April 23, 2007, from http://www.leapfroggroup.org.

TORCH Project. Retrieved April 23, 2007, from http://sourceforge.net/projects/op-torch.

Tringali, M., Hole, W.T. & Srinivasan, S. (2002). Integration of a standard gastrointestinal endoscopy terminology in the UMLS metathesaurus. *Proceedings / AMIA Annual Symposium*, (pp. 801-805). Los Angeles, CA: Hanley & Belfus.

Unified Medical Language System (UMLS). National Institutes of Health (NIH), U.S. National Library of Medicine (NLM), 8600 Rockville Pike, Bethesda, MD 20894. Retrieved April 1, 2007, from http://www.nlm.nih.gov/research/umls. Last update June 25, 2007.

van Bemmel J.H. & Musen, M.A. (1997). *Handbook of Medical Informatics*. The Netherlands: Bohn Stafleu Van Loghum.

van Ginneken, A.M. (2002). The computerized patient record: balancing effort and benefit. *International Journal of Medical Informatics, 65*, 97-119.

van Ginneken, A.M., de Wilde, M., van Mulligan, E.M. & Stam, H. (1997). Can Data Representation and Interface Demands Be Reconciled? Approach in ORCA. In D.R. Masys (Ed.), *Proceedings : a conference of the American Medical Informatics Association Annual Fall Symposium, Philadelphia, PA* (pp. 779-183). USA: Hanley And Belfus.

van Mulligan, E.M. (1999). UMLS-based access to CPR data. *International Journal of Medical Informatics 53* , 125-131.

Veterans Health Information Systems and Technology Architecture (VistA). Retrieved April 23, 2007, from http://www.va.gov.

Webster's Online Dictionary. Retrieved April 24, 2007, from http://www.websters-online-dictionary.org.

Westbrook, J.I., Coiera, E.W. & Gosling, A.S. (2005). Do Online Information Retrieval Systems Help Experienced Clinicians Answer Clinical Questions? *Journal of the American Medical Informatics Association, 12*, 315–321.

Wikipedia definition: ICT. Retrieved April 23, 2007, from http://en.wikipedia.org/wiki/Information_and_Communications_Technology.

Wikipedia, the free encyclopedia. Retrieved April 24, 2007, from http://www.wikipedia.org.

World Health Organization (WHO) International Classification of Diseases (ICD). Retrieved April 23, 2007, from http://www.who.int/classifications/icd/en/.

World Health Organization (WHO). (2006). Building foundations for eHealth : progress of member states : report of the Global Observatory for eHealth. Geneva: WHO Press.

Yamazaki, S. & Satomura, Y. (2000). Standard Method for Describing an Electronic Patient Record Template: Application of XML to Share Domain Knowledge. *Methods of Information in Medicine, 39*, 50-55.

Yeh, R.T. (1991). System Development as a Wicked Problem. *International Journal of Software Engineering and Knowledge Engineering, 1(2)*, 117-130.

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name:	Atalağ, Koray		
Date and Place of Birth:	12 March 1971, Ankara, TURKEY		
Nationality:	Turkish (TC)		
Marital Status:	Married and one daughter		
Phone:	+90 (312) 440 31 40	Fax:	+90 (312) 440 29 95
email:	atalagk@yahoo.com	Mobile:	+90 (532) 274 42 85

EDUCATION

Degree	Institution	Year of Graduation
MD.	Hacettepe University School of Medicine	1997
High School	Ankara Atatürk Anadolu Lisesi (AAAL)	1989

WORK EXPERIENCE

Self Employed 2006-

Freelance Health Information Systems Consultant and Developer

CrimsonLogic Global Pte. Ltd. Co. 2006

e-Government Gateway Project Office Coordinator/Translator

Tepe Teknolojik Servisler A.Ş. / corTTex Europe B.V. 2003-2004

Health Information Systems Consultant

Hacettepe University School of Medicine 2000-2003

ICT expert and lecturer in Department of Medical Education and Informatics

MEDİKOD Health Information Systems Ltd. Co. 1999-2000

Founder of the company and head of software development team.

Boğaziçi University, Department of Computer Engineering 1998-1999

System and Network Administrator / Web Admin (Research Assistant)

FOREIGN LANGUAGES

English – Advanced in written and verbal (TOEFL 623)

PUBLICATIONS and PROJECTS

Publications:

1. Atalağ, K., Bilgen, S. (2007). Multi-level Modeling and the Role of Archetypes in the Design of Health Information Systems: A Modeling Example in Endoscopy. In L.E. Akman, N. Baykal, T. Can, M. Iscan & U.E. Mumcuoglu (Eds.), *HIBIT '07 Proceedings of the International Symposium on Health Informatics and Bioinformatics, April 30 - May 2, 2007*. Ankara, Turkey: Middle East Technical University (METU) Press.

2. Atalağ, K., Bilgen, S., Gür, G. and Boyacıoğlu S. (2007b). Evaluation of the Turkish translation of the Minimal Standard Terminology for Digestive Endoscopy by development of an endoscopic information system. *The Turkish Journal of Gastroenterology* [submitted].

3. Atalag, K., Sincan, M., Celasun, B. and Karaagaoglu, E. (2004). Effects of lossy image compression on quantitative image analysis of cell nuclei. *Analytical and Quantitative Cytology and Histology, 26(1*), 22-7.

4. Atalağ, K. (2002). *Computer Implementation Of Bethesda System 2001 And Integration With A Pathology Information System With Automatic Snomed Coding Capability*. Poster presentation in 28th European Congress of Cytology in Antwerpen - Belgium.

5.Atalağ, K. (1998). *Establishment of a Human Mutation Database in Turkey:*

http://bioserver.bio.boun.edu.tr.Poster presentation in HUGO HGM'98 meeting, Turin/Italy.

6. Atalağ, K. (1991). Natural Gas Alarm Project. *TÜBİTAK Science and Technology Journal (Bilim ve Teknik), 24(278*), 57-58.

Hacettepe University Projects:

Design and Development of Course Management System	2000-2003
Hacettepem.org Health Portal for Provision of Publich Health	2001-2003
HÜTF Que-Bank Project (Web based question-bank system)	2001-2003

Private Projects

PATHOS-WEB Anatomic Pathology Information System 1996-

Free and Open Source Project at www.pathos-web.org

Currently being used at major Turkish institutions nationwide and worldwide

GASTROS-WEB Endoscopic Gastroenterology HIS Development 2000-

In collaboration with Başkent University Hospital Dept. of Gastroenterology

KlinikOS Clinical IS Development 2000-2001

In collaboration with Haydarpaşa Numune State Hospital

HOBBIES

Alpine ski, snowboard, skin and SCUBA diving, trekking, music, travel and technology

AFFILIATIONS

Turkish Informatics Accociation (TBD)

Turkish Medical Informatics Association (TURKMIA)

openEHR member

Human Genome Organization (HUGO) Mutation Database Initiative (MDI) member

Health Level 7 (HL 7) Turkey founding member.