DATA WAREHOUSING MODEL FOR INTEGRATING FRAGMENTED ELECTRONIC HEALTH RECORDS FROM DISPARATE AND HETEROGENEOUS CLINICAL DATA STORES

Saliya Nugawela (07699255)

B.E. (Computer Science and Engineering)

Principal Supervisor: Dr. Tony Sahama

Associate Supervisor: Ms. Susan Smith

Submitted in fulfilment of the requirements for the degree of

IT60-Master of Information Technology (Research)

School of Electrical Engineering and Computer Science

Faculty of Science and Engineering

Queensland University of Technology

June 2013

Keywords

Clinical Data Integration, Clinical Data Warehousing, eHealth, Electronic Health Records (EHRs), Electronic Medical Records, Health Data, Health Informatics, Medical Informatics

Abstract

Availability of timely and accurate data is vital to make informed medical decisions. Every type of health care organisation faces a common problem with the considerable amount of data they have in several systems. Such systems are unstructured and unorganised, demanding computational time for data and information integration. Data required to make informed medical decisions are trapped within fragmented and disparate clinical and administrative systems that are not properly integrated or fully utilised. The process of synthesising information from these multiple heterogeneous data sources is extremely difficult and time consuming. Ultimately, health care begins to constrain because, medical practitioners and health care providers are unable to access and use this information to improve patient care.

Without the ability to synthesise the information from disparate sources, health care providers and patients will not be able to leverage the knowledge stored in different repositories. Overall, varieties of approaches are available; however, they are limited in their application. What is now required is a data integration approach based on proven technologies. This research identifies and documents the main obstacles for healthcare data integration and proposes a data-warehousing model suitable for integrating fragmented data in a Cardiac Surgery Unit.

This research is based on the use of proven technologies in respective fields and research published in reputable journals and conferences. The result will contribute to the advancement of knowledge in the field of medical informatics.

If implemented with proper technology, the proposed data-warehousing model will help in efficient medical decision-making process.

The research undertaken is an Applied Research in the field of Translational Research Informatics (TRI). This research employs literature review and Data Warehousing Life Cycle as methods. Literature review is used to evaluate and analyse the literature relevant to the proposed research topic. Data Warehousing Life Cycle methodology is used to build the data warehousing model.

Outcomes of this research are:

1. Documenting the comparisons of data warehousing architectures, logical and conceptual data warehousing models.
2. Deriving and documenting Cardiac Surgery data table structures using data manual published by the Australasian Society of Cardiac and Thoracic Surgeons (ASCTS) under ASCTS National Cardiac Surgery Database Program (ASCTS, 2008).
3. Proposing a data-warehousing model suitable for a Cardiac Surgery Unit.
4. Identifying and documenting 13 main obstacles for healthcare data integration.
5. During the research period, presenting 3 conference papers among which 2 were included in conference proceeding publications. (Three papers are given in the Appendix – A)

Table of Contents

List of Figures

List of Tables

List of Abbreviations

BI	Business Intelligence
CDW	Clinical Data Warehouse
DBMS	Database Management System
DICOM	Digital Imaging and Communications in Medicine
DS	Decision Support
DWH	Data Warehouse
EHR	Electronic Health Record
EMR	Electronic Medical Record
ETL	Extract, Transform, Load
HIS	Health Information Systems
HIT	Health Information Technology
IT	Information Technology
LDM	Logical Data Model
NEHTA	National E-Health Transition Authority
NHHRC	National Hospital and Health Reform Commission
NHIMPC	National Health Information Management Principle Committee
OLAP	On Line Analytical Processing
PACS	Picture Archival and Communication Systems
PCEHR	Personally Controlled Electronic Health Records
PDM	Physical Data Model
RIS	Radiology Information Systems
SAS	Statistical Analysis System
SDM	Semantic Data Model
SNOMED CT	Systematized Nomenclature of Medicine -- Clinical Terms

Statement of Original Authorship

The work contained in this thesis has not been previously submitted to meet requirements for an award at this or any other higher education institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made.

Signature: ______________________

Date: 17 - JUNE -2013

Acknowledgments

I would like to firstly thank and acknowledge the guidance and vision provided by my principal supervisor, Dr. Tony Sahama to conduct this research. I would also like to thank my associate supervisor, Ms. Susan Smith for advice and support given throughout this project. I would also like to express my gratitude to Mr. Hiran Cooray and Mr. Shyam Dissanayake. I appreciate and thank all the support provided by QUT.

Dedication

This Thesis is Dedicated to My Mother Agnes

And

My Wife Tharindi

For

Their Love and Support.

Chapter 1: Introduction

This chapter outlines the background (section 1.1) and the research problem to be addressed in the research project (section 1.2) and research questions (section1.3). Section 1.4 describes the research context. Section 1.5 describes the significance and scope of this research. Finally, section 1.6 includes an outline of the remaining chapters of the thesis.

1.1 BACKGROUND

Informed decision making in health care is vital to provide timely, accurate, and relevant advice to the right person, to reduce the cost of health care and to improve the overall quality of health care services. Since medical decisions are very complex, making choices about medical decision-making processes, procedures and treatments can be overwhelming (Demetriades, Kolodner, & Christopherson, 2005).

One of the major Information Technology (IT) challenge in clinical practice is how to integrate several disparate, standalone information repositories into a single logical repository to create a single version of truth for all users (Chenhui et al., 2008; Kanup et al., 2007; Mann, 2005;Zheng et al., 2008; Goldstein et al.,2007; Shepherd, 2007). A massive amount of health records, related documents and medical images created by clinical diagnostic equipment are generated daily (Zheng, Jin, Zhang, Liu, & Chu, 2008). Medical documents are owned by different hospitals, departments, doctors, technicians, nurses, and patients. These valuable data are stored in various medical information systems such as HIS (Hospital Information System), RIS (Radiology Information System), PACS (Picture Archiving and Communications System) and etc., in various hospitals, departments and laboratories being primary locations (Zheng, Jin, Zhang, Liu, & Chu, 2008). These medical information systems are distributed and heterogeneous (utilising various software and hardware platforms including several configurations). Such processes and data flows have been reported by Zheng et al., (2008) (Zheng, Jin, Zhang, Liu, & Chu, 2008).

'*All medical data are located in different hospitals or different departments of single hospital. Every unit may use different hardware platforms, different operating systems, different information management systems, or different network protocols. Medical data is also in various formats. There are not only a tremendous volume of imaging files (unstructured data), but also many medical information such as medical records, diagnosis reports and cases with different definitions and structures in information system (structured data).*' (Zheng, Jin, Zhang, Liu, & Chu, 2008).

This causes Clinical Data Stores (CDS) with isolated information islands across various hospitals, departments, laboratories and related administrative processes, which are time consuming and demanding reliable integration (Sahama, & Croll, 2007). Data required to make informed medical decisions are trapped within fragmented, disparate, and heterogeneous clinical and administrative systems that are not properly integrated or fully utilised. Ultimately, health care begins to suffer because medical practitioners and health care providers are unable to access and use this information to perform activities such as diagnostics, prognostics and treatment optimisation to improve patient care.

Cardiac Surgery was chosen for this research since there are well defined and widely accepted data elements and table structures available for Cardiac surgery. This information is useful for data to be simulated and for designing a database to meet required data structures and standards. The supervisory team is familiar with this table structures, in particular an associate supervisor in the supervisory team engaged as a researcher at Cardiac Surgery Clinical Information Service, at the Prince Charles Hospital, Brisbane, Australia. This associated clinical environment enabled verification of the correctness of the proposed model.

1.2 PROBLEM DEFINITION

The main aim of this research is identifying the main obstacles for healthcare data integration and proposing a data-warehousing model suitable for integrating fragmented data in a Cardiac Surgery Unit.

1.3 RESEARCH QUESTION

Main questions concerned in this research are:

1. What are the main obstacles for healthcare data integration?
2. What is the most suitable data warehousing deployment architecture for a Cardiac Surgery unit?
3. What is the most suitable Conceptual Data Warehousing model for a Cardiac Surgery data set?
4. What is the most suitable Logical Data Warehousing model for a Cardiac Surgery data set?

1.4 RESEARCH CONTEXT

The preliminary literature review conducted revealed that though there are many significant research carried out in the topics of data warehousing and data integration, there have been very little research done on the relevance of the data warehousing and integration to the field of medical informatics. Most of the health data integration research done had focused on discovering the reasons for data fragmentation and proposing solutions. Very little research had been carried out to implement solutions to data fragmentation issue and test the validity of the solutions.

The previous research done on data warehousing and integration on other fields are helpful for this research. However, more research has to be carried out to ascertain the validity of the methods applied in other fields to health care domain. Health data are very complex. The figure 1.1 represents the various components of health data.

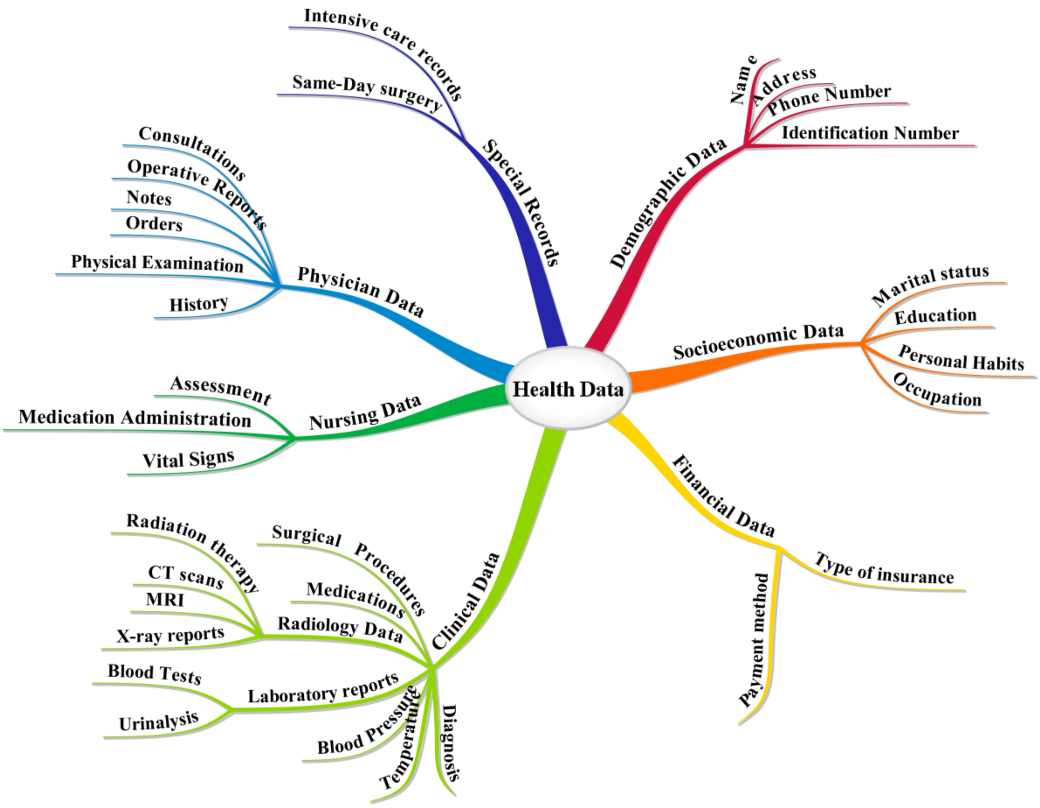

Figure 1.1 Components of health data

Continuous generation of very large amounts of unstructured data in health care demands the establishment of data warehouses with features beyond what is available in the traditional ones. Utilising the research already carried out in the field of data warehousing helps in finding a solution to the fragmentation of structured health data while more research is needed to find ways to minimise the fragmentation of unstructured health data.

1.5 SIGNIFICANCE OF THE RESEARCH

This research identifies and documents the main obstacles for the healthcare data integration and proposes a data-warehousing model suitable for integrating fragmented data in a Cardiac Surgery Unit. This research is based on the use of proven technologies in respective fields and research published in reputable journals and conferences. The result will contribute to the advancement of knowledge in the field of medical informatics.

If implemented with proper technology, the proposed data warehousing model will help in efficient medical decision making process.

1.6 THESIS OUTLINE

Chapter 1 provides details about the research background, research problem, its purpose and outcomes of the research.

Chapter 2 evaluates and analyses the literature relevant to the proposed research topic. This chapter presents a summary, classification and critical evaluation of the literature relevant to the research problem.

Chapter 3 describes the plans and methodology followed in order to achieve the aims and objectives of this research. This chapter discusses the characteristics of the research undertaken, Research Plan, research methodologies followed, resource requirements, ethical considerations, intellectual property rights, and health and safety issues of the research project.

Chapter 4 outlines the development of a data-warehousing model to integrate disparate and heterogeneous Cardiac Surgery data sets. This chapter describes the each stage of data warehouse model building in detail.

Chapter 5 outlines the analysis of the data-warehousing model proposed and documents the main obstacles for healthcare data integration.

Chapter 6 concludes the thesis by providing information on the research process, its benefits to healthcare data integration and future directions.

Chapter 2: Literature Review

Purpose

The purpose of this literature review is to evaluate and analyse the literature relevant to the proposed research topic. This chapter presents a summary, classification and critical evaluation of the literature relevant to the research problem.

Scope of the Literature Review

The main objectives of this literature review are to:

1. Present a brief overview about the field of medical informatics.
2. Identify the problems and challenges in medical informatics.
3. Develop an understanding about data integration and data warehousing technologies and their relevance to the medical domain.
4. Understand the application of Electronic Health Records and related technical issues.
5. Understand the state of current use and plans for medical informatics in Australia.
6. Present an overview about the main technologies used in this research.

Structure of the Literature Review

This literature review is structured under the following topics:

1. Information Technology (IT) in Health Care and Challenges to Adoption
2. Electronic Health Records (EHRs)
3. Current State and Future Plans for Electronic Health Data in Australia
4. An Introduction to Data Warehouses
5. Technologies

Review Methodology

The methodology implemented for this literature review was predominantly a bottom-up methodology, which emphasises on first searching for research papers and conference proceedings as opposed to the top-down approach of first starting with prominent academics and practitioners. The following sections detail the methodology used in the review.

Selecting Information Sources

The Internet has created a universal information explosion, as it facilitates an easier propagation of information. Therefore, there are many new tools at a researcher's disposal for the purpose of searching data.

MEDIA TYPE	SOURCE TYPE	INFORMATION SOURCE
Conference and journal papers	Publisher databases	ScienceDirect
		Web of Science
		ACM portal
		SpringerLink
		ProQuest
		IEEE Xplore
		CiteSeerX
		Elsevier
		JAMA
		ISI Web of Knowledge
	Citation indexes	ScienceDirect
		Google Scholar
Books	Libraries	Queensland University of Technology
		University of Queensland
	Online providers	Google Books
Technology and Product Reviews	Web Sites	http://tdwi.org/
		http://www.sas.com/
		http://www.oracle.com/

Table 2.1: Literature Search Sources

The primary literature search source was publisher databases. Numerous journals and conference proceedings are hosted by these sites; a large number of resources can be searched through in a short duration of time. The databases shown in Table 2.1 were selected due to their large collection of publications related to the computer science domain.

University libraries were used as the traditional method of information searching and retrieval, which still form the core source for reference knowledge. Furthermore, the general web search engines such as Google Scholar, Google, and Yahoo provided important e-books, peer reviewed articles as well as non-peer reviewed industry white papers that are related to my research field.

Searching Information Sources

The search strategy for each information source started with identifying relevant key words and terms around the topic. The search terms “data warehouse”, “data warehousing”, “Medical Informatics”, “Health Informatics” and “Clinical Data Warehousing” produced a plethora of search results. The next step was to combine words and phrases related to the research topic with the help of Boolean operators and proximity operators to narrow down the search. Once an article was found, abstracts were reviewed and if the article is related to the research topic, then it was fully reviewed and key points were extracted.

Managing Information

References were managed with Mendeley Desktop (a free reference management software) in order to reduce duplication of reading and summarisation. All relevant references from the searches were stored within the software. Full papers were stored in the Mendeley Desktop online database to facilitate future searches.

2.1 INFORMATION TECHNOLOGY (IT) IN HEALTH CARE AND CHALLENGES TO ADOPTION

2.1.1 INTRODUCTION

This section presents how IT is used in modern health care service provider organisations and the main problems related to the adoption of IT. The recent developments in health care have heightened the need for application of IT. The field of science which studies about the application of IT in health care is known as Medical Informatics or Health Informatics or E-health. In recent years, there has been an increasing interest in health care organisations to adopt IT in to their service process. However, there is increasing concern about the effective adoption of IT in health care domain. Questions have been raised about the challenges listed in the section 2.1.3. To date there has been very little progress in medical informatics compared to the rate of progress in other IT fields.

2.1.2 MEDICAL INFORMATICS BACKGROUND

In their book Goldstein et al. (2007, p.8) have defined medical informatics as the interconnection of medical, business, consumer centered care and IT disciplines for making significant and measurable improvements in both healthcare quality and cost effectiveness. Clinical guidelines, formal medical languages, standards, interoperability, and communication systems are part of medical informatics tools, technology and tactics apart from computers and information systems. (Goldstein, et al., 2007, p.8). The information in hospital medical records are immensely varied; they include: personal data of patients, referral and discharge letters from different doctors; investigation findings from laboratories and diagnostic imaging departments, medication charts, nursing notes, ECG traces, radiology data, examination, progress notes, and administrative data. (Benson, 2010, p.12) One question that needs to be asked is about the approach to convert massive piles of paper based data into electronic format.

According to Goldstein et al. (2007, p.305) Institute of Medicine has underlined six aims of improvements to be achieved in healthcare by utilising modern IT. The six aims are:

- **Safe:** Avoid causing injuries to patients from care that is intended to help them.
- **Effective:** Providing services using proven scientific knowledge tools.
- **Patient-centered:** Providing care that is responsive to individual patient preferences, needs and values, and assuring that patient values guide all medical decisions.
- **Timely:** Reducing waits and delays which can sometimes cause harm to both those who receive care and those who give care.
- **Efficient:** Avoiding waste of equipment, supplies, ideas and energy.
- **Equitable:** Providing quality care irrespective of personal characteristics such as gender, ethnicity, geographic location, or socio economic status.

As mentioned in above six aims, the proper implementation and use of IT in health care organisations can drastically improve the efficiency of medical decision making by providing relevant data at the point of care. According to Benson (2010, p.8), medical decisions such as ordering tests, therapy, referrals, and care, determine the cost and outcome of care. Only a small percentage of cost difference in health care can be attributed to service efficiency that is doing things right. Most cost and outcome inconsistency is the result of differences in clinical management, individual doctors' patterns of treatment and investigation. It is always important to do tests efficiently, but if the test is inappropriate, that is as an example it is repeated without good reason, it is a waste of resources irrespective of how efficiently it is performed.

In order to make patient data ubiquitously available worldwide for health care as well as for clinical research, epidemiological research, bio-surveillance and population health reporting, E-health ignores institutional boundaries and tries to share patient data among institutions. (Knaup, Bott, Kohl, Lovis, & Garde, 2007) The key problem with this explanation is that it has not considered the fact that still the health care institutions are not well connected to share the data.

2.1.3 CHALLENGES IN MEDICAL INFORMATICS

According to Goldstein et al. (2007), Kanup et al. (2007), Shepherd, (2007), Chenhui et al. (2008), Mann (2005), Zheng et al. (2008), Sahama & Croll (2007) and Leslie (2007) the main challenges in medical informatics are:

- Massive implementation and maintenance costs of health information systems
- Incompatible health information technology systems and proprietary software
- Lack of clinical data standards for sharing information
- Application systems which are not integrated to work collaboratively
- Fragmentation of health information over disparate and heterogeneous information systems

Since the research presented in this thesis is about integrating fragmented electronic health records only the last mentioned challenge is described in detail in this literature review.

2.1.4 FRAGMENTATION OF HEALTH INFORMATION OVER DISPARATE AND HETEROGENEOUS INFORMATION SYSTEMS

Multiple information systems are used by healthcare provider organisations in order to facilitate the process and information management needs from different departments. To make the best medical decisions and provide cost-efficient and quality patient care services healthcare workers need access to the complete medical information from these information systems. It is very important to maintain the consistency of medical information across departments to realise the benefits of digital workflow in order to increase the productivity. However, these information systems are been developed by different vendors with different technologies such as different computer languages, database management systems and different operating systems. (Chenhui et al., 2008; Kanup et al., 2007) This leads to wide spread incompatibility among systems. Thus, creating a hindrance to the smooth flow of information.

According to Chenhui et al. (2008), there are various system vendors who already provide solutions in an attempt to integrate with other systems, but healthcare facilities today are still facing the integration challenges of information systems with several problems:

- Access to the information is decentralised in multiple proprietary medical information systems
- Integration interfaces vary dramatically from system to system
- The number of system interfaces increases geometrically as the number of related systems increases. Therefore, multiplying the complexity to an unmanageable extent which would increase the total integration cost

Patients can get different forms of treatment from various healthcare sources for the same condition, with the same outcome but different costs. Fragmentation of care, when occurs even within healthcare institutions, is unproductive. Care is often discontinuous, since knowledge and activity being housed in separate data silos. (Mann, 2005) A serious consequence of this is that the patient being unable to receive the care expected.

In medical field, there are various sets of data such as patient’s medical records stored in databases, inspecting results saved as medical imaging files, clinical cases or doctor's advices saved as documents, and other video or images. Medical information systems, HIS (Hospital Information System), RIS (Radiology Information System), PACS (Picture Archiving and Communications System) and etc., are used to manage these varied medical data in hospitals or different departments. These systems are distributed and heterogeneous. They are spread across different hospitals and various departments within a hospital. The data within these systems are owned by different doctors, inspectors, nurses or patients. They cannot be changed or read by other persons. (Zheng et al., 2008) This leads to lack or contradicting information at the point of care. This causes the unsatisfactory treatment of the patients due to uninformed medical decision making.

All health care related data are located in different hospitals or different departments of one hospital. Every unit may use different hardware and software platforms. Medical data can also be in various formats. There are not only a tremendous volume of imaging files, but also much medical information such as medical records, diagnosis reports and cases with different definitions and structures in information system. The heterogeneities must be safeguarded and a unified, standard accessing pattern must be designed for the users. (Zheng et al., 2008) There is, therefore, a definite need for data integration in health care. But many difficulties arise when an attempt is made to integrate fragmented data.

As mentioned in their paper Sahama & Croll (2007), states that medical data is available in "information islands" spread across various departments, laboratories and related administrative processes. Accessing data from these islands are a time consuming and laborious task. The results of their study imply that, if proper policies are in place then data integration is technologically possible.

According to Benson (2010, p.8), the patient is the sole reason for healthcare activity, therefore the main focus should be on meeting the patient's needs. Each patient wants to live longer, feel well, and be able to do what he or she wants. Since data stored in proprietary information systems in a fragmented manner are unable to communicate with each other (Leslie, 2007), data required by medical practitioners to make informed decisions will not be readily available at the point of care.

2.1.5 CONCLUSION

The books and recent research papers reviewed provide concrete evidence that the fragmentation of medical data across multiple disparate and heterogeneous information systems is a real problem yet to be solved completely. The fragmentation exists among and within hospitals and various other health care service providers.

As Shepherd (2007) states in his paper, without the ability to synthesise information from disparate sources, health care practitioners and patients will not be able to leverage the knowledge stored in different repositories. Since health information is complex it must be made available at many different levels of knowledge and expertise. The knowledge must be synthesised and presented to solve the information need of the user and be presented to the user in an understandable format and at a correct level for comprehension. Without such data retrieval, the results of the information gathering may be at an inappropriate level for the user or contain information not applicable for the task or to the user. (Shepherd, 2007) An implication of this is, integrating data alone is not enough; the data must be presented in an appropriate level to suit the person.

Integrating medical information from fragmented data silos require each organisation to follow commonly accepted interoperability standards. This is another challenge faced by the medical domain. Medical data integration attempts mainly fail not because of technical incompetency, but due to various other reasons such as unavailability of policies and resistance from medical practitioners to get adapted to latest technology.

Strong relationship between data fragmentation and lack of interoperability standards has been reported in many literature reviewed.

2.2 ELECTRONIC HEALTH RECORDS (EHRS)

One of the most significant current discussions in health care is EHRs. Proper documentation of the health status of the patients is very important. It is becoming increasingly difficult to ignore the importance of having centrally managed and shareable patient health data. Traditional paper based method which has been used for centuries has many drawbacks. When implemented properly, EHRs can support the current information requirements of the health care industry. However, far too little attention has been paid to EHRs by both patients and health care organisations.

This section presents an overview about what an EHR is, key benefits to health care organisations and patients and the problems associated with the use of EHRs.

2.2.1 DEFINITION

According to Handler et al. (2003) an EHR is defined as:

'*The Electronic Health Record (EHR) is a secure, real-time, point-of-care, patient centric information resource for clinicians. The EHR aids clinicians' decision making by providing access to patient health record information where and when they need it and by incorporating evidence-based decision support. The EHR automates and streamlines the clinician's workflow, closing loops in communication and response that result in delays or gaps in care. The EHR also supports the collection of data for uses other than direct clinical care, such as billing, quality management, outcomes reporting, resource planning, and public health disease surveillance and reporting*.' This definition encompasses all the attributes and behaviours of EHRs.

An EHR can be considered as a computer based data collection created by physicians and other clinicians at the point of care to retrieve data later for reporting and use in research or administrative decision making. Health care workers use various computing devices such as desktop computers, laptops, handheld computers and bedside terminals to record the data. (Davis & LaCour, 2007, p.359)

2.2.2 KEY BENEFITS ASSOCIATED WITH THE USE OF EHR

According to Goldstein et al. (2007, p.198, 419) and Pearce & Haikerwal (2010) tangible and intangible benefits associated with EHRs to health care providers and patients are:

1. Improved quality of care
2. Reduced errors by migrating from paper based system
3. Availability of timely information about patients
4. Medical and medication safety with easy access to historical information
5. Satisfying reporting requirements of the organisations such as regulatory agencies, internal stake holders, and external stake holders
6. Ease of sharing information among everybody concerned

7. Saving time and administrative hassle by reducing the paper work
8. Increased revenue through timely and appropriate billing
9. Deep down satisfaction for clinicians about their care delivery
10. Overall health care cost reduction
11. Improved productivity of health care workers
12. Improved patient satisfaction
13. Reduced length of stay of a patient
14. Ability to utilise community based and evidence based best practices

The paper records can be used only in one place by one person at a time. Most of the time it is not available at where it is needed. When referring a paper record, it is hard to find what a practitioner wants in a disorganised, illegible, inconsistent, incomplete, badly sorted bundle of papers. The user has to spend lot of time and effort to do all of the work to glean any useful information. The electronic patient record can overcome these problems. Information quality is a paramount concern in the health care field. Computer-based patient records are legible and the information can be displayed in many different ways to suit the task at hand. Several people can work at different places at the same time to save the delays and effort required to locate, retrieve, and transport paper. Quality and safety can be improved, prevent key data being omitted, and save time by avoiding data replication. (Benson, 2010, p.12)

The EHR is the central component of the IT infrastructure in many modern healthcare organisations. It has become the common tool used by all healthcare practitioners working in the organisation. It is the entry point of most patient information, and it provides access to the data born in other systems, e.g., laboratory or financial systems. Data in the EHR is used and organised according to operational purposes, where many kinds of data about one patient is presented to get an overview of the health status of the patient. (Pedersen & Jensen, 1998) However, very little attention has been paid to date to integrate EHRs belongs to a patient which are scattered over multiple organisations.

All the above mentioned advantages can be realised only if the workforce in a health care organisation and patients fully support the implementation of EHR.

Another important practical implication is that all the concerned parties should follow a common standard to maintain EHRs.

2.2.3 BARRIERS TO IMPLEMENTATION OF ELECTRONIC HEALTH RECORDS

Although EHR systems are increasingly used in many health care organisations, there still remain some barriers to their adoption.' *There is a huge gap between the availability of innovative technologies and their application in daily health care. Reasons for this include the inherent complexity of the field, costs, as well as ethical and legal requirements.*' (Kanup et al., 2007). The main barriers to the implementation of EHRs as identified by Davis & LaCour (2007), Kanup et al. (2007), Zenios (2005), Sandsmark (2008), and Hersh (2009) are:

1. **Cost of Conversion**

The cost involved in converting the paper based records systems to an electronic system is massive. This cost includes the acquisition of hardware, software and trained technicians. (Davis & LaCour, 2007). Certainly the biggest impediment is financial since most of the small scale healthcare organisations are unable to bear the above mentioned costs without the help of a funding agency.

2. **Concerns about Privacy and Confidentiality of Data**

Many healthcare organizations provide services beyond hospitals to neighbourhood clinics, home-health providers, and off-site services such as radiology interpretation. In this distributed environment, EHRs are always on the move, and the security of critical information infrastructure becomes more difficult. (Sandsmark, 2008) Main implication of this is the need to follow stringent standards.

Special cultural environment in a healthcare organisation demands security to be a fine-tuned balance between technologies, human elements, standard practices and procedures (Sandsmark, 2008). In many instances the same data set of a patient is accessed by administrative staff, physicians, nurses and laboratory in order to make decisions regarding patient's healthcare. Therefore all those who are involved in the process have shared responsibility and accountability to maintain the security and integrity of a patient record.

3. **System Interoperability**

Various clinical and administrative systems within and beyond a healthcare organization must work together in a smooth manner to give optimum performance. But this does not happen since most of the proprietary software systems by various vendors do not communicate with each other effectively.

Most health care data, whether on paper or electronic format, are trapped in multiple data silos in multiple vendor products. As a result, a patient may have a physician or health system with an advanced EHR, yet if that patient requires care elsewhere, there is little likelihood the information from that advanced system will be accessible when it is needed (Hersh, 2009).

4. **Lack of a Well-Trained Medical Informatics Workforce to Lead the Process**

To maintain an efficient electronic health record keeping system, an organisation must have a well trained workforce with a clear understanding of the requirements of both the worlds of medicine and IT. They should be highly motivated to carry out operations and make innovations to support rapidly changing requirements of healthcare industry. This type of a workforce building is a highly time consuming and costly effort.

5. **Data Storage Requirements**

The need for data storage in healthcare grows rapidly. Systems such as Picture Archival and Communications Systems (PACS), which handle digital X-ray, CT, and MRI images use significant storage. These systems are frequently used by most of the healthcare organisations today. Improving the storage environment means more than simply adding better storage hardware. Centralised, standardised storage-management software, which are independent of hardware and are able to manage the diverse, heterogeneous environments that exist in real-world data centers, are important ingredients in the ideal storage prescription. This also should include maximising utilisation of existing storage, improving backup and recovery performance, and classifying structured and unstructured data to improve archiving and retrieval. This last benefit is of particular importance to clinicians, who need the right information at the right time. (Sandsmark, 2008).

6. **Fragmentation of EHRs**

EHRs are fragmented across various disparate and heterogeneous organisational systems. They are fragmented among and within hospitals. Health care practitioners, providers and patients often make decisions about medical treatments without complete understanding of the circumstances. The main reason for this is that medical data are available in fragmented, disparate and heterogeneous data silos. Without a centralised data warehouse structure to integrate these data silos, it is highly unlikely and impractical for the users to get all the information required on time to make a correct decision. (Shepherd, 2007)

2.2.4 CONCLUSION

The active participation of leadership, clinicians, patients and all those involved in healthcare is a must to make EHRs and information technology an essential resource to solve the problems in medical field. This suggests the need for every stakeholder in health care to participating in implementing working framework for EHRs.

Benefits of adopting EHRs in clinical practices largely outweigh the efforts required to overcome the barriers. The literature reviewed under this section also clearly highlights that the fragmentation of medical data is a problem currently faced by the health care organisations. Fragmented health record lacks the ability to deliver its purpose. Rather it might cause more damage to patients by not revealing some critical information at the point of care.

2.3 CURRENT STATE AND FUTURE PLANS FOR ELECTRONIC HEALTH DATA IN AUSTRALIA

2.3.1 INTRODUCTION

According to Goldstein et al. (2007, p.20) there are four major challenges in medical informatics currently.

1. Widespread medical errors due to system complexity, lack of information technology facilities and communication gaps among systems and practitioners
2. Not using the latest medical knowledge for diagnosis and treatment

3. Partial, fragmented patient medical records due to paper records at numerous sites of care.

Recently, researchers have shown an increased interest in finding solutions for the above mentioned challenges.

In their book Goldstein et al. (2007, p.529-530) describe the following capabilities of a truly effective EHR system:

1. Should have the capability to capture clinical images, videos and other forms of unstructured data, and make them part of the patient's electronic medical records.
2. Both image and text data should integrated in order to facilitate the clinician's task of correlating the data and making patient care decisions in a timely and accurate manner.
3. Computers used by clinicians should be able to display the high resolution clinical images stored in the EHR.
4. Imaging systems should be able to use the Digital Imaging and Communications in Medicine (DICOM) standard to obtain images directly from image acquisition machines like CT, MRI, ultra-sound, and digital X-ray.
5. All captured unstructured clinical data should be associated with the text report of the procedure or the related progress note within the EHR.
6. Should serve as a tool to aid communication and consultation among physicians.

One question that needs to be asked, however, is whether the realisation of the above mentioned capabilities is just a technical endeavour. In order to achieve the above mentioned capabilities, there need to be a very high level of dedication from all the stakeholders. This section presents the over view of current state of medical informatics in Australia and the initiatives taken by the government for the improvement in future.

2.3.2 CURRENT STATE OF MEDICAL INFORMATICS IN AUSTRALIA

Health informatics is now seen as an increasingly important tool against combating disease. It is no more just for efficiency alone. There are mounting evidences that when used properly both health outcomes and consumer satisfaction can be improved by the use of information technology in health care. Health informatics is also an essential component of any quality and safety agenda for Australia. (Health Informatics Society of Australia, 2007, p.3)

Regardless of widespread use of modern technology in healthcare domain, Australia’s health care system lags behind all other sectors the economy in the use of computerised systems. Currently, Australia sits in the middle in rankings of health systems among industrialised nations, and the use of modern electronic technologies for communication and clinical information transfer within health systems is low. The uncoordinated implementation information systems within hospitals, between hospitals in a region and across boundaries compounds a lack of national coordination and so loses the benefits of drawing on expertise and knowledge across the nation. The difficulties of introducing a comprehensive e-health system are an example of how the fragmented health system cannot work across silos. (Pearce & Haikerwal, 2010) The key problem with this explanation is that it does not give a clear indication for the main reason for data fragmentation. It seems that the authors assume the problem is only a technical issue.

According to NHIMPC (2007) although the Australian health system delivers world class health outcomes, consumers often have difficulty finding relevant information since health records and information are primarily paper-based. Even researchers, administrators and policy makers often find it difficult to locate, interpret and validate the information they need to manage the system and enhance outcomes for consumers.

2.3.3 FUTURE PLANS

Australia has one of the top 20% health systems in the world based on the health outcomes of its citizens. Conserving or improving on the current state will be challenging in the future as the current system is struggling to deal with increasing costs, and supply and demand pressures such as a shortage of skilled healthcare workers and an ageing population. (NEHTA Blueprint, 2010, p.10) A serious weakness of with this argument, however, is that it does not consider the applicability of information technology in health care.

While the Australian health system has its quality strengths, it is under growing pressure and the continued delivery of healthcare using the existing fragmented health system is ill-equipped to cope with these pressures. The Australian government has recognised the challenges ahead for the Australian health system and commissioned the National Hospital and Health Reform Commission (NHHRC) to develop a national plan for health reform. The vision of the NHHRC is for a sustainable, high quality, responsive health system for all Australians, now and into the future. (NEHTA Blueprint, 2010, p.10)

E-health is a key element to revolutionising the healthcare system as general adoption of e-health technologies by healthcare consumers and providers has the capacity to:

- Ensure the right health information is electronically made available to the right person at the right place and time to assist informed treatment decisions.
- Improve communication between individuals and providers and giving individuals electronic access to their health information.
- Supporting individuals and providers with the ability to monitor individual care plans and health status.
- Enable the Australian health sector to more efficiently operate as an inter-connected system overcoming the current fragmentation and duplication of information.
- Enable multi-disciplinary teams to electronically communicate and exchange information and provide better coordinated health care.
- Provide individuals with confidence that their personal health information is managed in a secure and confidential manner.
- Enable electronic access to appropriate health care services for consumers within remote, rural and disadvantaged communities.
- Facilitate continuous improvement of the health system through more effective reporting and sharing of health outcome information.

- Improve the quality, safety and efficiency of healthcare facilities by giving providers and individuals better access to health information, clinical evidence and clinical decision support tools.
- Support more informed policy, investment and research decisions through access to timely, accurate and comprehensive reporting on Australian health system activities and outcomes. (NEHTA Blueprint, 2010, p.12)

These proposals suggest several courses of action for the proper implementation of medical informatics. In their proposals NEHTA has pointed out number of important changes which need to be made to the current scenario.

Interoperable and compatible health information management systems are critical to achieving health reform that is designed to address present and future requirements. This will enable better integration of services across the care range to achieve improved quality of care for individuals and communities and assist in reducing cost increases driven by advances in medical technologies. (NHIMPC, 2007, p.2)

Figure 2.1 shows the proposed national eHealth architecture. This architecture takes into consideration all the parties involved in health care providing activities.

Figure 2.2 shows the architecture to integrate various health information systems fragmented and spread all over hospitals, state and federal government authorities in disparate and heterogeneous platforms into a single logical unit to make the data available to the required practitioners.

Figure 2.3 shows the interaction of various components involved in the proposed Personally Controlled Electronic Health Records (PCEHR) infrastructure. Under PCEHR initiative each patient will get a 16 digit identification number to identify their health records. Patients will have the authority to determine who will have the access to their health records.

Establishing EHRs as a patient oriented approach will involve the patient more deeply in the care process as he becomes responsible for keeping and recording his own health data and for making the right data available to the right persons at the right point in time. Access to information can be considered as one major element of patient empowerment. (Kanup et al., 2007) Future EHR systems should provide more power to the patient to control their health records. Considerably more work has to be done to achieve this goal.

Goldstein et al. (2007) states that, future healthcare system should be a one that is patient-centered, one with every person has a secure, private Electronic Health Record (EHR) that is available at the point of care, enabling the highest degree of coordinated medical care based on the latest medical knowledge and evidence.

The introduction of a person-controlled electronic health record for each Australian is one of the most important opportunities to improve the quality and safety of health care and improve continuity and health outcomes for patients. Giving people better access to their own health information through a person-controlled electronic health record is also essential to promoting patient participation, and supporting self-management and informed decision-making. (Report of NHHRC, 2009, p.8)

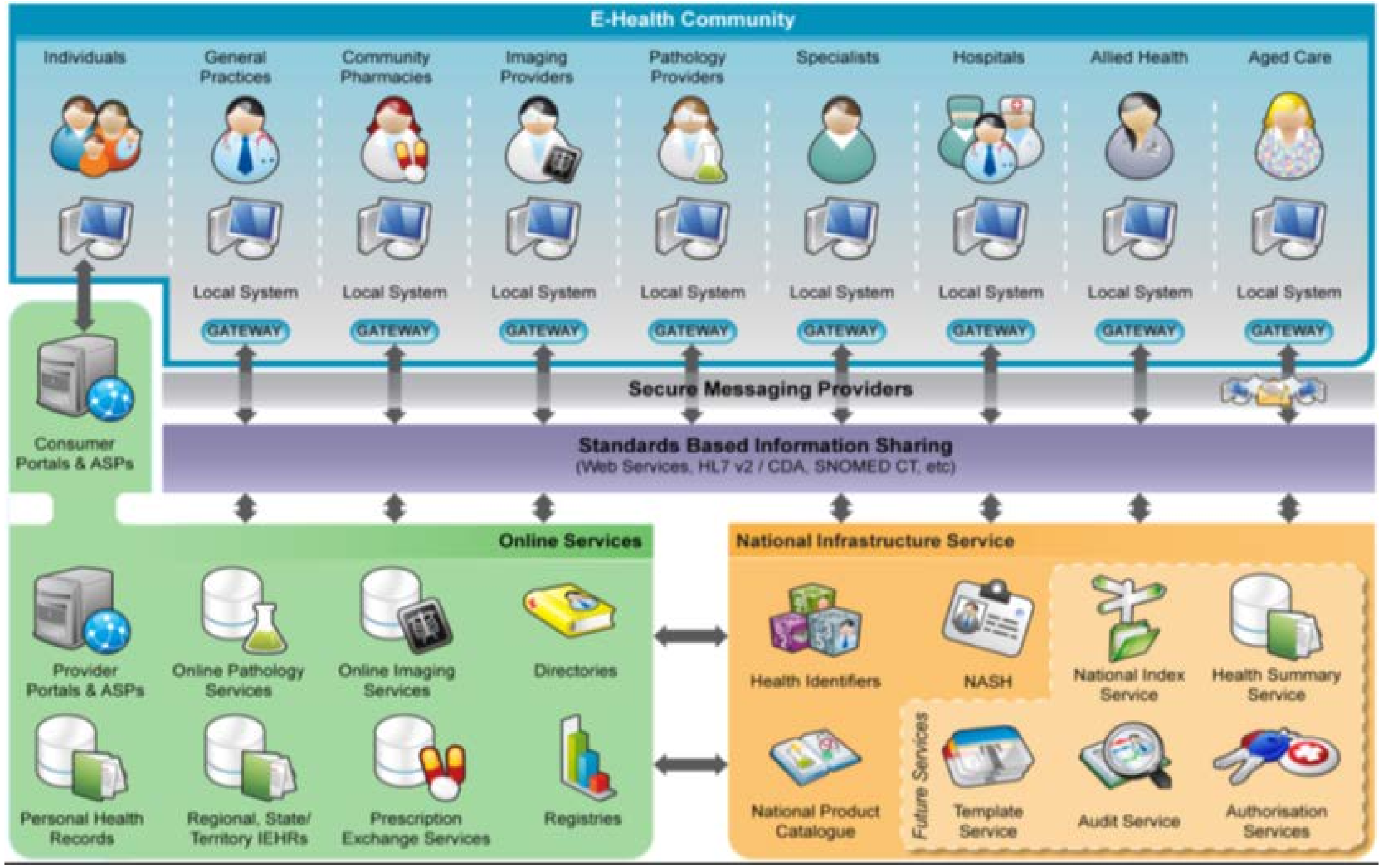

Figure 2.1: Proposed National eHealth Architecture

(NEHTA Blueprint, 2010, p.153)

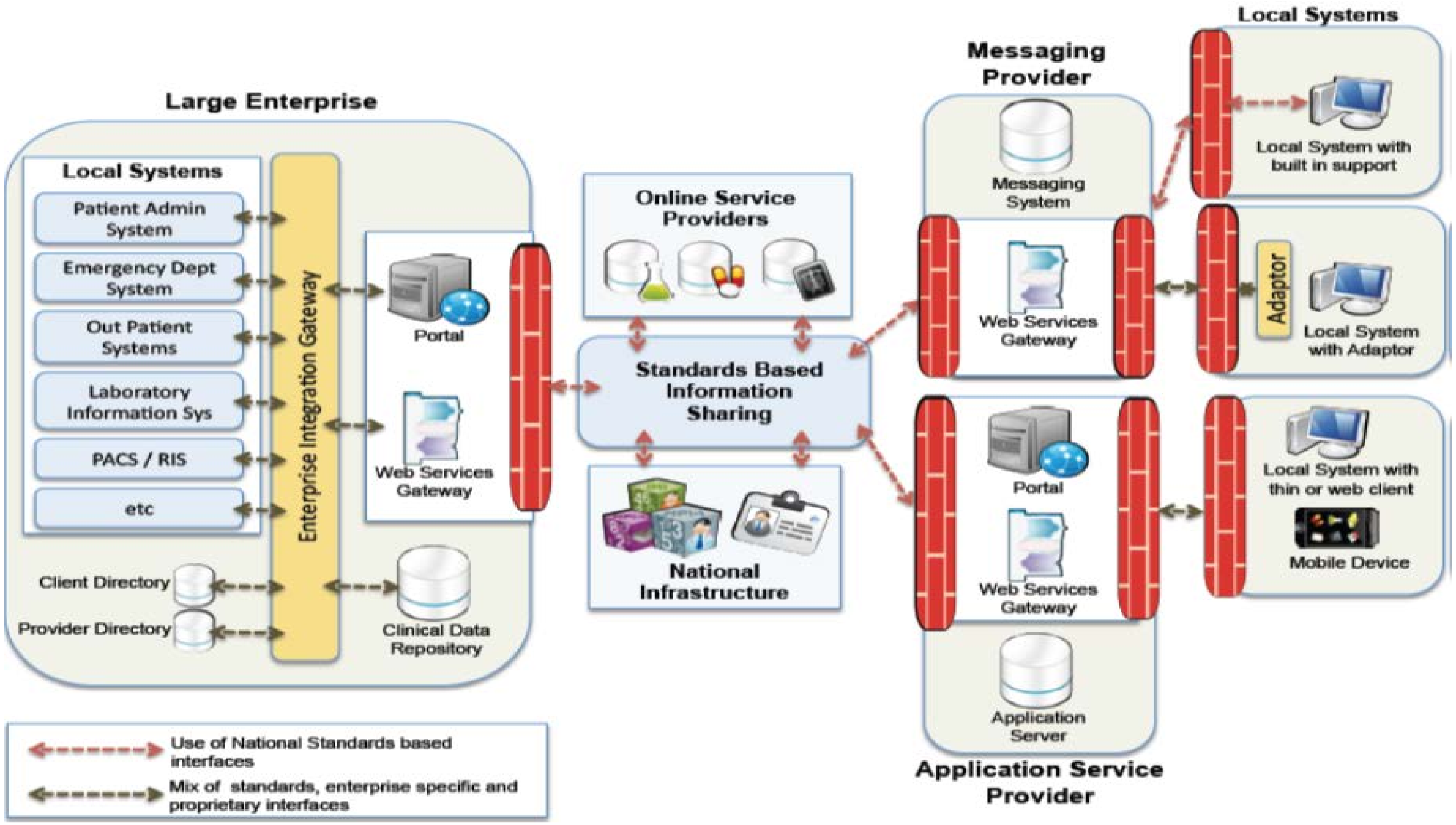

Figure 2.2: Proposed High Level Integration Architecture

(NEHTA Blueprint, 2010, p.159)

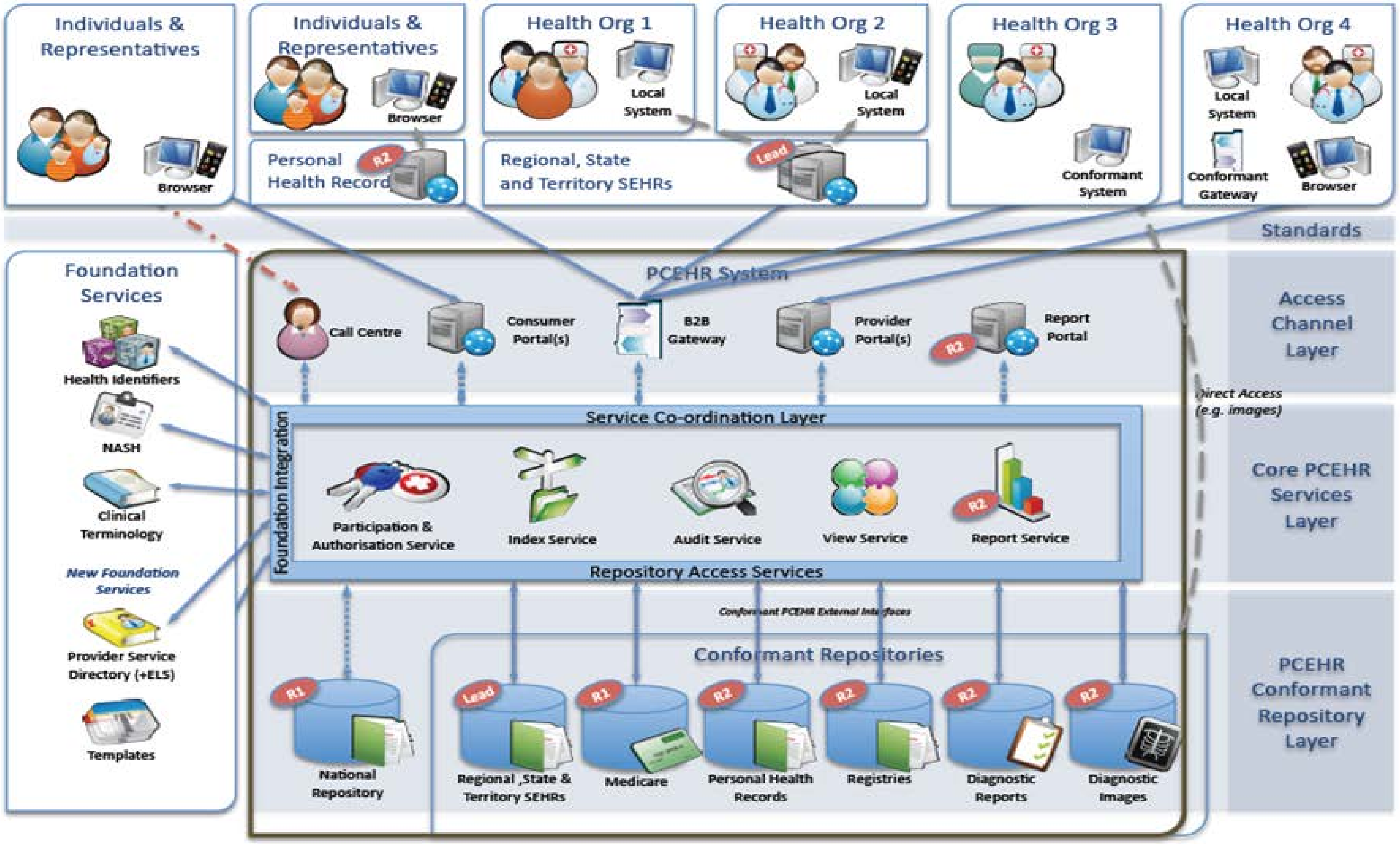

Figure 2.3: Core Components of Proposed PCEHR Infrastructure

(DRAFT Concept of Operations PCEHR System, 2010, p.36)

2.3.4 CONCLUSION

The literature reviewed under this section gives an understanding about the problems faced by eHealth in Australia and steps taken to improve the situation in the future. The NEHTA and NHHRC reports clearly indicate that the fragmentation of health information is a major challenge to the progress of eHealth in Australia. It would be interesting to assess the effects of the proposals made by NEHTA and NHHRC to the future development of medical informatics in Australia.

Future research should therefore concentrate on the investigation of the problems mentioned in the above mentioned government reports. This research is aimed at building an Enterprise Integration Gateway as indicated in Figure 3.3.2.

2.4 AN INTRODUCTION TO DATA WAREHOUSES AND THEIR USE IN THE MEDICAL FIELD

2.4.1 A DATA WAREHOUSE DEFINITION

According to Inmon (2002, p.31-32) '*A data warehouse is a subject-oriented, integrated, non-volatile, and time-variant collection of data in support of management's decisions. The data warehouse contains granular corporate data.*

Subject-oriented: *Classical operations systems are organised around the applications of the company. Each type of company has its own unique set of subjects.*

Integrated: *Data is fed from multiple disparate sources into the data warehouse. As the data is fed it is converted, reformatted, resequenced, summarized, and so forth. The result is that data once it resides in the data warehouse has a single physical corporate image.*

Non-volatile: *Data warehouse data is loaded and accessed, but it is not updated. Instead, when data in the data warehouse is loaded, it is loaded in a snapshot, static format. When subsequent changes occur, a new snapshot record is written. In doing so a history of data is kept in the data warehouse.*

Time-variant: *Every unit of data in the data warehouse is accurate as of some one moment in time. In some cases, a record is time stamped. In other cases, a record has a date of transaction. But in every case, there is some form of time marking to show the moment in time during which the record is accurate.*'

This definition of data warehousing is the most widely accepted one in the subject domain. This section provides a description about the need for data warehousing and integration in the medical field.

2.4.2 DATA IN SUPPORT OF MANAGEMENT DECISIONS

In order to serve the decision making process of the management the data warehouse has to supply the following primary functionality:

- The Data Warehouse (DWH) is a reflection of the business rules of the enterprise which can be applied to strategic decision support (DS) information.
- It is the collection point for the integrated, subject- oriented strategic information that is handled by the data acquisition process.
- It is the historical store of strategic information.
- It is the source of stable data regardless of how the processes may change.
- Additionally the data warehouse should support ad hoc queries.

2.4.3 BENEFITS OF DATA WAREHOUSING

Data warehousing is being hailed as one of the most strategically significant developments in information processing in recent times. One of the reasons for this is that it is seen as part of the answer to information overload. The points highlighted in Bill Inmon's definition give some of the reasons why data warehousing is regarded as important.

- Has a subject area orientation
- Integrates data from multiple, diverse sources
- Allows for analysis of data over time
- Adds ad hoc reporting and enquiry
- Provides analysis capabilities to decision makers
- Relieves the development burden on IT
- Provides improved performance for complex analytical queries
- Relieves processing burden on transaction oriented databases
- Allows for a continuous planning process

- Converts corporate data into strategic information

2.4.4 COSTS OF DATA WAREHOUSING

There are few down sides of the Data Warehousing technology. However, data warehouses can be extremely expensive to build and maintain, and have a high failure rate unless there is the right mix of high need, powerful sponsor, and reasonably short time scale. Also, the impact they can have on traditional views of data ownership and organisational structures can be quite disruptive. Some of the more noticeable costs associated with data warehousing are listed below:

- Time spent in careful analysis of measurable needs
- Design and implementation effort
- Hardware costs
- Software costs
- On-going support and maintenance
- Resulting re-engineering effort

2.4.5 DATA WAREHOUSE USERS

One of the remarkable aspects of the data warehouse evolution into its different parts is that a community of users has evolved along with the different forms of warehousing. In the 1989 time frame, there was simply the ‘end user’. Today in the world of DSS data warehouse users, there are farmers, explorers, tourists, power users, ad hoc users, top management users, analytical users, data miners, and many more. As the different data warehouse structures have unfolded, so have the different types of users begun to appear (Inmon, 2002).

Data warehouse users can be divided into four groups. These four groups include power users, standard users, occasional users and scientific users.

Power Users

Include data miners and explorers, who transform business queries into complex queries to the DWH. Additional tools and techniques are often employed for their search for data patterns. But they include also data analysers and farmers, who create queries to the DWH with existing/traditional tools and make them available to other users.

Standard Users

Include management personnel to whom the data warehouse supplies highly aggregated graphically represented information.

Occasional Users

This category consists of personnel who benefit from being able to access the data warehouse on a non-regular basis.

Clinical / Scientific Users

The group of clinical or scientific users is a blend of the characteristics that describe each of the previously mentioned three groups. They are in parts explorers or data miners for the use the DWH to gather new insights in an occasional or frequent manner.

2.4.6 THE STRUCTURE OF A DATA WAREHOUSE

According to William Inmon five different data components can be found within a Data Warehouse are:

- Metadata
- Current detail data
- Old detail data
- Lightly summarised data
- Highly summarised data

(Inmon, 2002)

Metadata

Metadata are data which describe data warehouse components and their relationship within the data warehouse. Since metadata contain no data directly taken from the operational environment, in many ways metadata sits in a different dimension than other DWH data. (Inmon, 2002). Metadata play an important role in a DWH.

Current detail data

Current detail data are the major concern of DWH applications, because they reflect the most recent happenings (Inmon, 2002). They are however capacious due to the fact that they are data at their lowest level of granularity. Current detail data are usually stored on high capacity disk storage devises.

Old detail data

Old detail data are current detail data that have passed the time frame of immediate importance and are therefore, migrated to a different, less costly storage space such as magnetic tapes or high capacity optical storage devices.

Lightly summarised data

These data are a distilled version the current detail data. They are an intermediary step towards highly summarised data and “almost always stored on disk storage” (Inmon, 2002).

Highly summarised data

These data are the highly distilled and compact version of data that were aggregated by analysts and other power users to serve the information needs of an organisation. Occasionally highly summarised data are found in the data warehouse environment and in other cases highly summarised data are even found outside of the technology that houses the DWH. Irrespective of where the data are physically stored, highly summarised data are part of the data warehouse’ (Inmon, 2002).

2.4.7 DATA WAREHOUSE MODELLING

The fundamental modelling techniques for a DWH are relational or multidimensional modelling or hybrid approach.

The three different levels of data modelling

To support the different requirements of a DWH project three different views are taken upon the data in a DWH and consequently translated into a three level model of any DWH. The Semantic Data Model (SDM) shows a high level view of the DWH project; a model in the language of the DWH end-user. The Logical Data Model (LDM) shows entities and their relationships in a logically sound manner, to serve as model for physical implementation. And finally the Physical Data Model

(PDM) shows the actual representation of the physical tables in the database as they are implemented.

The semantic level

The semantic level, sometimes also labelled conceptual level, is described by a semantic data model. An SDM aims to create an abstract representation of the situation under investigation, or more precisely, the way users think about it. Analysing the domain of user concepts should be done independently of any software technological considerations (Inmon, 2002).

The logical level

The logical level is described by the logical data model. An LDM's purpose is to describe entities and their relationships to facilitate the integration of data from the disparate source systems and to enable the data analyst to normalise data elements found in those source files. It is completely process independent and only captures the objects and actions as they exist in the real world for the organisation (Inmon, 2002).

The physical level

The physical level is described by the Physical Data Model. A PDM shows the actual representation of the tables in the database as they are implemented in the project. The PDM serves as a guide through an existing database, for programmers, administrators and analysts (Inmon, 2002).

Relational data modelling and the Entity-Relationship model

The basis for the Relational Data Model was sketched out by E. F. Codd. Relational database theory is centred on the process of data normalisation. Normalisation delivers a set of storage structures that minimise the effect of data anomalies. The inputs to the normalization process are the so-called data dependencies. Much of relational database theory is about the formal manipulation of data dependencies. Relational theory assumes the existence of a universal data dependency that is decomposed in the design process in its entities and their relations. This process is called 'Normalization by Decomposition' because in progressing from one normal form to the next larger tables are decomposed into smaller tables. Beyond Third Normal Form the resulting design can depend on the order of decomposition. This, though, violates the Conceptualisation Principle because the choice of a designer has effect on the structure of the designed model. No conclusive guiding line has been presented so far in theory.

The Entity-Relationship model, however, is the most popular technique to visualise relational data models. Classes of objects are identified and modelled as entities. Entities have properties that are either attributes that directly describe entities or can be relationships that model facts about entities. Entities, relationships and attributes are depicted visually in an Entity-Relationship diagram.

Multidimensional data modelling

The multidimensional data model is based on the key concepts cube, dimension and hierarchy. The cube is the concept that describes the whole of the data which are presented along labelled edges of that cube, the dimensions. Whereas hierarchies define the way in which dimensions are grouped. This dimensional modelling approach results in a database design that is consistent with the paths by which users wish to enter and navigate a DWH. Frequently requested aggregates, or calculated measures, are stored in the database, creating useful data redundancies that make it possible to avoid performance-inhibiting repetitive calculations every time a report is prepared.

Cube

A data cube is a representation of multidimensional data defined by a set of ***k*** different dimensions. Data can be organized into a data cube by calculating all its possible aggregations [SQL statement GROUP-BY]. The set of aggregations form a ***k***-dimensional data cube.

Dimension

The dimensions of a cube depend on the data to be modelled in the database. Typically, each dimension is an independent entry point or mechanism for selecting data. For example, a cube for a classic DWH application has the following three dimensions: product, market and time. A cube in a medical context on the other hand could have time, diagnosis and treatment as dimensions.

Hierarchy

A Hierarchy defines the way in which a dimension can be grouped in a DWH. For example, a possible hierarchy for the time dimension could consist of year, quarter, month, week and day.

Multidimensional modelling techniques

The two prominent data modelling techniques for the multidimensional model are the Snowflake Schema and the Star Schema.

The Snowflake Schema

A snowflake schema is the traditional multidimensional model implemented on a relational database management system (DBMS). It consists of a central fact table containing the measures or some other content of interest, and is surrounded by dimension entities containing conformed context for the analysis of the fact. The dimensions usually relate to the facts in one to many relationships and the snowflake schema exposes them as fully normalised structures usually consisting of many entities with often complex intra dimensional relationships.

An advantage of the Snowflake Schema is the clear exposure of the dimensional hierarchy which often directly relates to the aggregation levels that can be applied to the fact. For example, the time dimension may consist of a hierarchy representing year, quarter, month, week and day. This hierarchy is directly related to the temporal aggregations that can be applied to the fact, that is, view the facts consolidated by year, quarter, month, week or day. This is easy for the end user and supports an ad hoc analysis environment. The main disadvantage of the Snowflake Schema is the large number of tables that have to be joined to support even the most basic queries.

The Star Schema

Star Schemas were a direct reaction to that disadvantage of Snowflake Schemas. They were designed to produce models that boast a minimum join distance. Therefore the Star Schema simply consists of the fact surrounded by a single level of collapsed or consolidated dimension entities. For the example, the information representing year, quarter, month, week and day would all have to be effectively represented in a single entity. It is obvious that this will usually lead to better query performance as queries will now join fewer tables, a benefit to the RDBMS that does not perform multi-table joins efficiently.

The disadvantage of the Star Schema, like any de-normalised model, is the hiding of relationships like the dimensional hierarchy or consolidation path so evident in the Snowflake Schema. The Star Schema will also introduce redundancy and processing anomalies that require a more skilful handling of queries. Which modelling technique is better suited therefore depends on the project at hand.

2.4.8 NEED FOR MEDICAL DATA WAREHOUSING

Electronic Health Records (EHRs) which describe the diseases and treatments of patients are normally stored in the hospitals or clinics, where they are created. However, patients may be treated in different hospitals, clinics and, therefore, there is a need for integrating health records from different hospitals to enable any hospital to obtain a total overview of a patient's health history. Two different types of heterogeneity problems have to be solved in order to integrate EHR systems from different hospitals in a consistent way. The first problem is that different hospitals normally do not use a common database management system (DBMS) and therefore, the traditional ACID (Atomicity, Consistency, Isolation and Durability) database properties are missing across the different hospital locations. This may cause performance, autonomy, and consistency problems. The second heterogeneity problem is that there are multiple incompatible standards for how to make EHR entries. (Frank & Andersen, 2010)

In order to fulfil the information needs of medical practitioners and patients it is a necessity to Integrate physician records, hospital services, medication histories, and other medical information into a unified digital record that is available to needed persons at home or at the point of care. To realise the benefits of joined-up health care, to provide the right information at the right time and place, health care organisations must deploy and use standards that enable computer systems to exchange information in a way that is safe, secure, and reliable. (Goldstein et al., 2007 p.3; Benson, 2010, p.26)

An effective healthcare establishment requires the availability of efficient software tools that enable easy searching, retrieval and analysis of all gathered patient data, which are generally stored in distributed and heterogeneous databases.

Medical data integration is a significant challenge due to the great variety of data types, formats and platforms present in the medical scenario. (Masseroli et al, 2004)

Most of the times data warehouses fail because they do not meet the needs of the particular application domain, or are too difficult/expensive to change with the evolving needs of the domain. (Sen & Sinha, 2005)

2.4.9 DATA WAREHOUSING IN MEDICAL FIELD

Health care organisations require data warehousing solutions in order to integrate the valuable patient and administrative data fragmented across multiple information systems within the organisation. As stated by Kerkri, et al. (2001), at a technical level, information sources are heterogeneous, autonomous, and have an independent life cycle. Therefore, cooperation between these systems needs specific solutions. These solutions must ensure the confidentiality of patient information.

To achieve sufficient medical data share and integration, it is essential for the medical and health enterprises to develop an efficient medical information grid (Zheng et al., 2008).

A medical data warehouse is a repository where healthcare providers can gain access to medical data gathered in the patient care process. Extracting medical domain information to a data warehouse can facilitate efficient storage, enhances timely analysis and increases the quality of real time decision making processes. Currently medical data warehouses need to address the issues of data location, technical platforms, and data formats; organisational behaviours on processing the data and culture across the data management population. Today’s healthcare organisations require not only the quality and effectiveness of their treatment, but also reduction of waste and unnecessary costs. By effectively leveraging enterprise wide data on labour expenditures, supply utilisation, procedures, medications prescribed, and other costs associated with patient care, healthcare professionals can identify and correct wasteful practices and unnecessary expenditures. (Sahama & Croll, 2007).

Medical domain has certain unique data requirements such as high volumes of unstructured data (e.g. digital image files, voice clips, radiology information, etc.) and data confidentiality. Data warehousing models should accommodate these unique needs. According to Pedersen & Jensen (1998) the task of integrating data from several EHR systems is a hard one. This creates the need for a common standard for EHR data.

According to Kerkri et al (2001), the advantages and disadvantages of data warehousing are given below.

Advantages:

1. Ability to allow existing legacy systems to continue in operation without any modification
2. Consolidating inconsistent data from various legacy systems into one coherent set
3. Improving quality of data
4. Allowing users to retrieve necessary data by themselves

Disadvantages:

1. Development cost and time constraints

2.4.10 CONCLUSION

'*A data repository that gathers information from all areas that represent health and social care stakeholders leads to the possibility of having a holistic view of the system.*' (Szirbik et al., 2006) There is, therefore a definite need for data warehousing and integration in medical domain. However, the goal of providing a holistic view of the system cannot be achieved unless the governments implement policies which make medical organisations to follow a commonly accepted set of standards and freely share the information about a patient among organisations.

According to the literature reviewed, the main aim of a medical data warehouse should be to provide a repository of medical data related to patients and administrative processes extracted from multiple disparate and heterogeneous systems which are not compatible with each other.

2.5 TECHNOLOGIES

2.5.1 INTRODUCTION

This section presents a description about a technology popular in data integration.

2.5.2 SAS® DATA INTEGRATION TECHNOLOGIES

For many years the process known as extract, transform and load (ETL) is in use for any business intelligence solution. The critical factor in building a data warehouse or data mart to support an organization's business intelligence projects is selecting the right tool to bring data from disparate sources and transform it before loading into a target destination. (The New Data Integration Landscape, 2007)

Many companies manage their information in data silos that are spread throughout the enterprise and use inconsistent storage formats. The rise of heterogeneous platforms has made it hard to identify inconsistent and duplicate data, which contributes to the rising cost of data management. (SAS® Data Integration, 2009) This explanation clearly indicates the reasons for data fragmentation.

SAS® Data Integration provides a solid foundation for supporting information needs of a health care organisation by helping to transform patient and organisational data into useful information. It provides capabilities for enterprise data access and processing across systems and platforms – including internal processes, patient information, and third-party or regulatory sources, the Web and more. SAS® Data Integration also helps to ensure information excellence through data quality lifecycle management. And an interactive, visual data integration development environment enables collaboration and easy reusability across the organisation. (SAS® for Health Care, 2010)

SAS® Data Integration includes:

- An interactive GUI-based data integration process development environment.
- Sophisticated metadata management facilities.
- Capabilities for creating, running and managing flexible data integration services.
- Read/write capabilities to third-party databases.

- Access to unstructured and semi structured data sources.
- Deployment and management facilities.
- ETL and ELT for in-database processing.
- Integrated data quality and cleansing capabilities.
- The ability to convert existing SAS programs to graphical data integration jobs.

(SAS® Data Integration, 2009)

In addition to transforming data into intelligence, SAS® ensures that this information is made available to any person in the organisation who needs to see it in whatever format they choose. Whether they require electronic distribution of reports, interactive query environments, delivery of content via a Web-based portal or publish and subscribe channel distribution of information through e-mail. SAS® ensures that the applications which are deployed today will serve the organisation in the years to come, despite any changes in the infrastructure. (SAS® for Health Care, 2010)

2.6 CONCLUSION OF PRELIMINARY LITERATURE REVIEW

The books, research papers, white papers, government reports and websites referred in the preparation of this literature review provide a clear idea about the field of medical informatics, unsolved challenges in the domain and the technological and other concepts which can be helpful in dealing with the challenges. This research is aimed at proposing and developing a solution for a main technological challenge.

The challenges listed in section 2.1 hinder the efficient implementation of health information technology; thus, creating a resistance among patients and clinicians to use this helpful technology. Though the field of medical informatics emerged about four decades back, it had proved how useful it can be for providing better health care.

One of the main building blocks of medical informatics is the Electronic Health Record (EHR) which is used for recording patient information electronically. The full potential of EHRs have not been realised yet due to various technical and non-technical issues. One major obstacle is the fragmentation of health data across multiple disparate and heterogeneous information systems. This is the research problem focused in this research. Various books and recent research papers used for this literature review conforms that the fragmentation of health data is still an unresolved problem. This problem requires both solutions of technical and non-technical nature. Focus of this research is narrowed into developing a solution to integrate the fragmented health information within a health care service provider organisation. Without integrating the information within individual organisations, national level integration cannot be achieved.

This research will investigate about the application of data warehousing and integration technologies into medical domain. As previously mentioned, data warehousing had been used in many other fields successfully to integrate data. In medical domain, data warehousing faces problems mainly due to certain non-technical issues.

One major non-technical issue related to the research problem is non-adherence to commonly accepted set of standards by care providers and information systems developers.

As mentioned in the section 2.4, NEHTA and NHHRC have come up with proposals and implementation plans to improve the state of medical informatics in Australia. One major initiative is the PCEHR (Personally Controlled EHR). This gives power to the patient to handle his or her EHR.

SAS® data integration technologies and openEHR architecture can be used to implement a solution for the research problem. SAS® is accepted all over the world as the leading business analytic tool and open source data integration tools are gaining popularity.

The research problem focused on is a challenge due to various reasons. Some of them are, heterogeneity of medical information systems, lack of widely accepted standards, incompatibility among proprietary software and lack of support from patients, clinicians and policy makers.

The literature review compiled, documents the current findings related to the research problem.

Chapter 3: Research Design

This chapter describes the plans and methodology followed in order to achieve the aims and objectives of this research. Section 3.1 discusses the characteristics of the research undertaken. Section 3.2 describes the Research Plan. Section 3.3 explains the research methodologies followed. Remainder of the chapter discuss the resource requirements, ethical considerations, intellectual property rights and health and safety issues of the research project.

3.1 CHARACTERSITICS OF THE RESEARCH UNDERTAKEN

1. Propose a data warehousing model to integrate clinical data derived from disparate systems into coherent Cardiac Surgery data.
2. In keeping with Applied Science processes, in order to solve the problem, well known and accepted theories, principles and tools relevant to data integration, data warehousing, medical informatics and medical data interoperability are employed.
3. The outcome of this research has immediate applicability in the problem domain.
4. The proposed solution can be applied without making any changes to the already existing information systems in the health care organisation, assuming verification of the simulated data to data available in real world clinical information systems .

The main aim of this research is to propose an applicable model solution for a practical problem which is suitable for immediate use. Therefore, considering all the characteristics, this research can be categorised as **Applied Research** in the field of **Translational Research Informatics (TRI)**. (Woolf S.H., 2007;Payne P.R., Embi P.J., & Sen C.K., 2009)

3.2 RESEARCH PLAN

Stage of the Research	Research Method Applied
Stage 1: Research Planning and Design	Literature Review
Stage 2: Investigate the Current state of Medical Data Fragmentation, Relevant Data Warehousing and Integration Technologies	Literature Review Observation Review Existing Systems
Stage 3: Create the table structures for Cardiac Surgery data Simulate data & tables	Data Warehouse Development Life Cycle
Stage 4: Designing the Data Warehouse Model for Cardiac Surgery Data	Data Warehouse Development Life Cycle
Stage 5: Preparation and Submission of the Thesis	Not Applicable

Table 3.1: Stages of the Research and Methods to be applied

Stages of the research are depicted in the Figure 3.1. There are six stages in the research.

Stage 1: Research Planning and Design

This stage consists of determining the research problem, conducting literature review and developing the research problem. Clearly identifying the research problem at the early stages helps in doing the research successfully.

The literature review is conducted in order to present a brief overview about the field of medical informatics, identify the problems and challenges in medical informatics, develop an understanding about data integration and data warehousing technologies and their relevance to the medical domain, understand the application of Electronic Health Records and related technical issues, understand the state of current use and future plans of medical informatics in Australia, present an overview about the two main technologies used in this research, and present a brief description about interoperability standards used in medical domain.

Research proposal document integrates research question, literature and method in a feasible research plan.

Outcomes of Stage 1: Problem Definition, Literature Review and Research Proposal

Stage 2: Investigate the Current State of Medical Data Fragmentation, Relevant Data Warehousing and Integration Technologies

This stage consists of:

1. Studying the structure of data distribution and communication in existing health information systems.

This facilitates the understanding of the underlying data related problems in the health care information systems.

2. Compare and analyse various data warehousing technologies in order to identify the most suitable model(s) to health care domain.
3. Compare and analyse various data integration technologies.

The findings of the processes in this stage are important to understand the data warehousing and integration technologies employed in various other fields and their relevance to health care domain.

Existing health information systems and research published in highly reputable journals and conferences will be used in this stage.

Outcomes of Stage 2: Current State of Health Data, Comparison and Analysis Results of Data Warehousing and Integration Technologies.

Stage 3: Create the Table Structures for Cardiac Surgery data and simulate data in laboratory environment

This stage involves developing table structures for cardiac surgery data using the data manual published by Australasian Society of Cardiac and Thoracic Surgeons (ASCTS) under ASCTS National Cardiac Surgery Database Program and populating the tables with simulated data.

Outcomes of Stage 3: Cardiac Surgery table structures populated with simulated data.

Stage 4: Designing the Data Warehouse Model for Cardiac Surgery Data

This stage involves developing a suitable data warehouse model to integrate cardiac surgery data tables created during the stage 3.

Outcomes of Stage 4: Cardiac Surgery Data Warehouse Model

Stage 5: Thesis Submission

Ultimate objective of this stage is to submit the final thesis. The steps included in this stage are:

1. Complete the drafting of the thesis based on the research findings and implementation results of the previous stages.
2. Draft thesis will be then submitted for the final seminar.
3. Final thesis document will be prepared based on the feedback from the final seminar.

Completed thesis will be submitted for the external evaluation

Outcomes of Stage 5: Final Thesis

3.3 METHODOLOGY

3.3.1 LITERATURE REVIEW

Literature review was done as explained in chapter 2.

3.3.2 DATA TABLES STRUCTURE DEVELOPEMTN AND SIMULATION PROCESS

Detailed explanation of how the data simulation was carried out according to the data definitions and format in the ASCTS tables is described in Chapter 4.

3.3.3 DATA WAREHOUSE DEVELOPMENT LIFE CYCLE

The data warehouse development life cycle chosen is the Kimball Lifecycle approach described in the book “The Data Warehouse Lifecycle Toolkit”. This approach has been practiced for decades. The concepts were originally conceived in the 1980s by members of the Kimball Group and several colleagues at Metaphor Computer Systems. It was referred to as the Business Dimensional Lifecycle because this name reinforced three fundamental concepts:

- Focus on adding business value across the enterprise
- Dimensionally structure the data delivered to the business via reports and queries
- Iteratively develop the solution in manageable lifecycle increments rather than attempting a Big Bang deliverable

Kimball Lifecycle consists of the nine steps mentioned below:

1. Program/Project planning
2. Program/Project Management
3. Business Requirements Definition
4. Technology Track
5. Data Track
6. Business Intelligence Application Track
7. Deployment
8. Maintenance
9. Growth

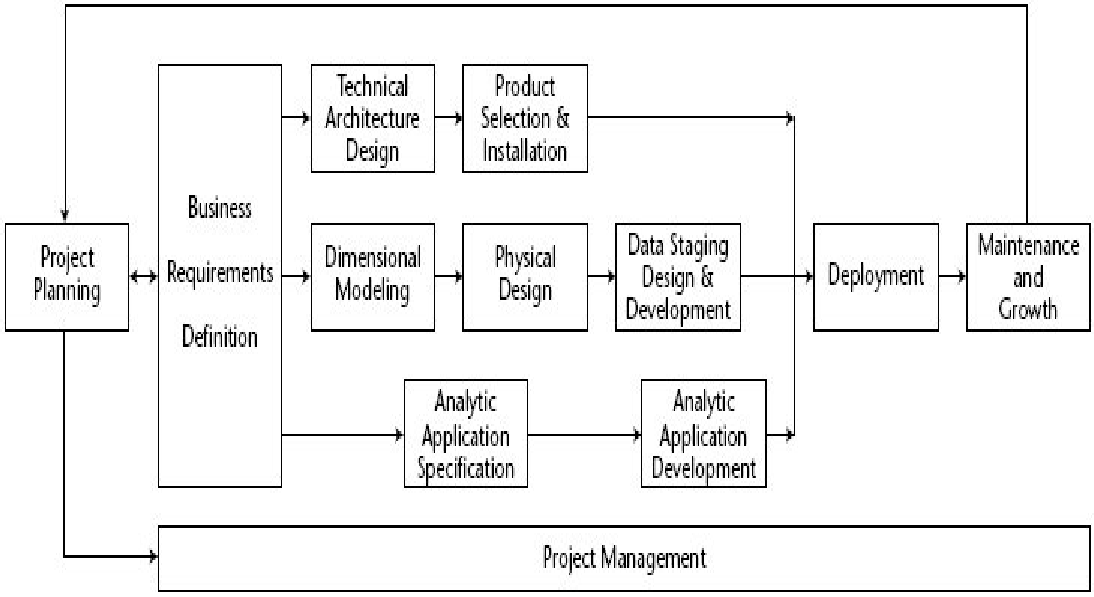

Figure 3.1: The Kimball Lifecycle Diagram

(Kimball,Ross,Thornthwaite,Mundy, & Becker,2008)

The description of the main steps as described in the book "The Data Warehouse Lifecycle Toolkit" (Kimball,Ross,Thornthwaite,Mundy, & Becker,2008) is given below.

Program/Project Planning and Management:

The first box on the roadmap focuses on getting the program/project launched, including scoping, justification and staffing. Throughout the Lifecycle, ongoing program and project management tasks keep activities on track.

Business Requirements:

Eliciting business requirements is a key task in the Kimball Lifecycle as these findings drive most upstream and downstream decisions. Requirements are collected to determine the key factors impacting the business by focusing on what business users do today (or want to do in the future), rather than asking “what do you want in the data warehouse?” Major opportunities across the enterprise are identified, prioritised based on business value and feasibility, and then detailed requirements are gathered for the first iteration of the DW/BI system development. Three concurrent Lifecycle tracks follow the business requirements definition.

Data Track:

The data track begins with the design of a target dimensional model to address the business requirements, while considering the underlying data realities. In dimensional modelling the data is divided into either measurement facts or descriptive dimensions. Dimensional models can be instantiated in relational databases, referred to as star schemas, or multidimensional databases, known as On Line Analytical Processing (OLAP) cubes. Regardless of the platform, dimensional models attempt to address two simultaneous goals: ease of use from the users’ perspective and fast query performance. The Enterprise Data Warehouse Bus Matrix is a key Kimball Lifecycle deliverable representing an organization’s core business processes and associated common conformed dimensions; it’s a data blueprint to ensure top-down enterprise integration with manageable bottom-up delivery by focusing on a single business process at a time. The bus matrix is tremendously important because it simultaneously serves as a technical guide, a management guide, and a forum for communication with executives.

The dimensional model is converted into a physical design where performance tuning strategies are considered, then the ETL system design and development challenges are tackled. The Lifecycle describes 34 subsystems in the extract, transformation and load process flow grouped into four major operations: extracting the data from the source, performing cleansing and conforming transformations, delivering the data to the presentation layer, and managing the backroom ETL processes and environment.

Business Intelligence (BI) Track:

While some project members are immersed in the technology and data, others focus on identifying and constructing a broad range of BI applications, including standardized reports, parameterized queries, dashboards, scorecards, analytic models, data mining applications, along with the associated navigational interfaces.

Deployment, Maintenance, and Growth:

The three Lifecycle tracks converge at deployment, bringing together the technology, data and BI applications. The deployed iteration enters a maintenance phase, while growth is addressed by the arrow back to project planning for the next iteration of the DW/BI system.

Applying Kimball Life Cycle approach to this research

Since this research was not intended to apply and test the solution in a healthcare institution, it is not practical to cover all the Life Cycle Stages. Only the Program/Project planning, Program/Project Management, Business Requirements Definition, Technology Track, Data Track and Business Intelligence Application Track stages are fully covered. Practical suggestions are provided for Deployment, Maintenance and Growth stages where applicable.

ASCTS registry is mainly used for monitoring outcomes in Cardiac Surgery Units. The above mentioned steps of the Kimball LifeCyscle Approach was applied to integrate the various ASCTS tables fragmented across disparate systems and facilitate efficient clinical decision making.

3.4 RESOURCE REQUIREMENTS

Table 3.2 illustrates the additional resources required at each stage of the project.

Stage of the Research	Additional Resource Requirements
Stage 1: Research Planning and Design	• Access to Journal and Conference Papers. • Books on Medical Informatics and Data Warehousing.
Stage 2: Investigate the Current state of Medical Data Fragmentation, Relevant Data Warehousing and Integration Technologies	• Access to Journal and Conference Papers. • Books on Medical Informatics and Data Warehousing.
Stage 3: Create the table structures for Cardiac Surgery data. Synthesize appropriate data in the tables	• Access to ASCTS Data Manual. • Computer lab facilities • Data simulation software • Consultation with cardiac surgery information users/clients • Database Design Software.
Stage 4: Designing the Data Warehouse Model for Cardiac Surgery Data	• Database Design Software.
Stage 5: Thesis Submission	• Word Processing Software

Table 3.2: Additional Resource Requirements

3.5 INTELLECTUAL PROPERTY STATEMENT

This research is being undertaken as part of a Master by Research course at the Queensland University of Technology. The research is not being conducted as part of a joint collaboration with an industrial, commercial, professional or research organisation. At the time of writing, no student intellectual property and confidentiality agreement has been signed.

3.6 HEALTH AND SAFETY STATEMENT

This research does not involve exposure to any high risk materials nor does the research involve the use of biological, microbiological, biomedical or biochemical material. Accordingly, this research does not give rise to any special health and safety training.

3.7 RESEARCH ETHICS CHECKLIST / STATEMENT

This research does not require obtaining any Ethical Clearance.

Chapter 4: Results

This chapter outlines the development of a data warehousing model to integrate disparate and heterogeneous Cardiac Surgery data sets. This chapter describes each stage of data warehouse model building in detail.

The data tables and table structures were derived from the data manual published by Australasian Society of Cardiac and Thoracic Surgeons (ASCTS) under ASCTS National Cardiac Surgery Database Program (ASCTS, 2008). The ASCTS Database records details of all adult cardiac surgical procedures performed in participating hospital units. The Hospitals participating in registry using the current manual are:

- Austin Hospital – VIC
- Geelong Hospital – VIC
- Monash Medical Centre – VIC
- Royal Melbourne Hospital – VIC
- The Alfred Hospital – VIC
- Mater Misericordiae Hospital - QLD
- Lake Macquarie Private Hospital - NSW
- Flinders Medical Centre – SA
- Canberra Hospital - ACT
- John Hunter Hospital - NSW
- Prince of Wales Hospital - NSW
- Westmead Hospital - NSW
- St George Hospital - NSW
- St Vincent's Hospital - NSW
- Liverpool Hospital - NSW
- Royal North Shore Hospital - NSW

- Royal Prince Alfred Hospital - NSW
- Cabrini Health – VIC
- Townsville Hospital – QLD

4.1 THE DATA TABLES

List of table names taken into consideration is as per below:

1. Patient Demographics
2. Patient Risk Factors
3. Pre-Operative Cardiac Status
4. Previous Interventions
5. Haemodynamic Data
6. Operative Status Category
7. Minimally Invasive
8. CPB and Support
9. Coronary Bypass
10. Valve Surgery
11. Post-Operative Data
12. Post-Operative Data - Complications
13. Mortality/Readmission
14. Automatic Data

4.2 THE CARDIAC SURGERY REGISTER TABLE STRUCTURES

Patient Demographics

Column Name	Meaning	Data Type
LNAME	The last name (surname) of the patient.	Text
FNAME	The first name (Christian name / Given name) of the patient.	Text
MNAME	The middle name of the patient.	Text
DOB	The date of birth of the patient.	Date
SEX	The sex of the patient.	Numeric
NOMEDC	Patient does not have a medicare number	Numeric

MEDC	The Full Medicare number of the patient (i.e. family number plus person number) if the patient is registered with Medicare	Numeric
ADDRESS	The Street Number, Name and Type of the patient's residential address	Text
SUBURB	The Suburb of the patient's residence	Text
STATE	The State of the patient's residence	Text
PCODE	The Post Code of the patient's residence.	Numeric
PHONE1	The Contact Phone Number of the patient.	Text
PHONE2	The Contact Phone Number of the patient.	Text
RACE1	Is the patient Aboriginal and/or Torres Strait Islander?	Numeric
RACE2	Does racial group include Aboriginal?	Numeric
RACE3	Does racial group include Torres Strait Islander?	Numeric
UR	Patient medical record number at the hospital where surgery occurred.	Text
DOSA	Patient admitted for scheduled elective procedure on same day as procedure. [Day of admission into acute hospital in which surgery is to be performed- thus patients admitted to Medihotel the night prior to surgery may still qualify as DOSA patients].	Numeric
DOA	Date Patient admitted/transferred to hospital where surgery performed.	Date
DOD	Date Patient discharged from being an inpatient at the hospital where the procedure was performed. Discharge to Hospital in the Home, rehabilitation hospital or unit or to a local referring hospital is considered as discharge from hospital.	Date
INSUR	Select the category which most accurately describes the patient's insurance status: Private Patient has private health insurance DVA Patient is funded by Department of Veteran Affairs Self- Patient is self-funded (private patient without Insured Private health insurance) Overseas patient is an overseas visitor Medicare Patient is funded by Medicare Other all other payment classes (eg TAC, Australian Military, Seamen, Work cover)	Numeric
DOP	Date on which the first surgical incision was made for the current Cardiac Surgical Procedure	Date
PROCNO	Number of operation(s) done on the day for this patient	Numeric
VerifiedComment	Comment of verification on surgery	Text

Patient Risk Factors

Column Name	Meaning	Data Type
SMO_H	A history confirming any form of tobacco use in the past	Numeric
SMO_C	Patients having smoked cigarettes within one month of surgery are considered to be current smokers.	Numeric
FHCAD	Family Historyof CAD. Whether any direct blood relatives (parents, siblings, children) have had any of the following at age <55: a. Angina b. Myocardial infarction (MI) c. Sudden cardiac death presumed to be from ischaemic heart disease because of no other obvious cause.	Numeric

	d. Coronary intervention	
DB	A history of diabetes, regardless of duration of disease or need for anti-diabetic agents.	Numeric
DB_CON	Method of diabetic control, at time of intervention. The most aggressive therapy should be indicated as per the following order: insulin > oral > diet. Section requirement one choice only: None: No treatment for diabetes Diet: Diet treatment only Oral: Oral agent treatment Insulin: Insulin treatment (includes any combination of above with insulin)	Numeric
HCHOL	Whether the patient has a history of hypercholesterolaemia diagnosed and/or treated by a physician, and/or Cholesterol > 5.0 mmol/L, HDL <1.0 mmol/L or Triglycerides >2.0 mmol/L.	Numeric
PRECR	Last serum creatinine recorded prior to surgery.	Numeric
DIAL	Is the patient on dialysis pre-operatively?	Numeric
TRANS	Did the patient have prior renal transplant?	Numeric
HYT	Does the patient have a diagnosis of hypertension documented by one of the following: a. Documented history of hypertension diagnosed and treated with medication, diet and/or exercise. b. Blood pressure >140 systolic or >90 diastolic on at least 2 occasions. c. Currently on antihypertensive medication.	Numeric
CBVD	Whether the patient has had Cerebro-Vascular Disease, documented by any one of the following: a. Unresponsive coma >24 hrs b. CVA (symptoms >72 hrs after onset) c. RIND (recovery within 72 hrs) d. TIA (recovery within 24 hrs) e. Non-invasive carotid test with 50% diameter stenosis (equivalent to 75% cross-sectional area stenosis).	Numeric
CBVD_T	What type of Cerebro-Vascular Disease does the patient have? Choose one of the following: 1. Unresponsive Coma >24 hrs: Patient experienced complete mental unresponsiveness and no evidence of psychological or physiologically appropriate responses to stimulation. 2. CVA: Patient has a history of stroke, i.e. loss of neurological function with residual symptoms >72 hours after onset. 3. RIND: Patient has a history of loss of neurological function with symptoms >24 hours after onset but with a complete return of function within 72 hours. TIA: Patient has a history of loss of neurological function that was abrupt in onset but with complete return of function within 24 hours. 4. Non-invasive/invasive carotid test with 50% diameter stenosis (equivalent to 75% cross-sectional area stenosis).	Numeric
CVA_W	Those events occurring within two weeks of the surgical procedure are considered recent, while all others are considered remote.	Numeric
PVD	The patients history of PVD either aneurysmal or chronic or acute occlusion or narrowing of the arterial lumen of the aorta or extremities. Includes the following: • claudication either with exertion or rest • amputation for arterial insufficiency	Numeric

	• vascular reconstruction, bypass surgery, or percutaneous intervention to the extremities • documented aortic aneurysm • documented renal artery stenosis • positive non-invasive testing documented (e.g. ankle brachial index <0.8).	
LD	Whether the patient has chronic lung disease, and severity level according to the following classification: No: No disease Mild: On chronic inhaled or oral bronchodilator therapy Moderate: On chronic oral steroid therapy directed at lung disease Severe: Room Air p02 < 60 or Room Air pC02 > 50 or mechanical ventilation for chronic lung disease	Numeric
LD_T	Specify if the patient has chronic lung disease, and severity level according to the following classification: Mild: On chronic inhaled or oral bronchodilator therapy Moderate: On chronic oral steroid therapy directed at lung disease Severe: Room Air p02 < 60 or Room Air pC02 > 50 or mechanical ventilation for chronic lung disease	Numeric
IE	A patient presenting with valvular disease of infectious aetiology with past or present positive blood culture, or post-operative pathology confirmation.	Numeric
IE_T	Active: If the patient is currently being treated for endocarditis, the disease is considered active. Treated: If no antibiotic medication (other than prophylactic medication) is being given at the time of surgery, then the infection is considered treated.	Numeric
IMSRX	Use of any form of immunosuppressive therapy, including systemic steroid therapy equivalent to ≥ 5mg prednisolone within 30 days or less preceding the operative procedure.	Numeric

Pre-Operative Cardiac Status

Column Name	Meaning	Data Type
MI	Patient hospitalised at any time for a Myocardial Infarction documented in the medical record.	Numeric
MI_T	1. Non ST Elevation MI (NSTEMI) Must have at least one of the following: A. BIOCHEMICAL indicators of myocardial necrosis 1. Troponin T or I > the institutional decision limit on at least one occasion during the first 24 hrs after the index event. 2. CKMB >2x the upper limit of normal on one occasion during the first 24 hrs. 3. CKMB (or preferably the CKMB mass) > upper limit of normal on 2 successive samples. AND one of the following: B. ECG CHANGES – either ST segment depression OR T-wave abnormalities OR C. CLINCAL ISCHAEMIC SYMPTOMS such as: 1. Unexplained nausea or vomiting, &/or 2. Persistent SOB secondary to LVF, &/or 3. Unexplained weakness, dizziness or syncope 2. ST elevation MI (STEMI) Must have the following:	Numeric

	A. BIOCHEMICAL indicators as for Non-STEMI. AND B. ECG CHANGES 1. ST segment elevation: New or presumed new ST elevation at the J-point in two or more contiguous leads with cut-off points => 0.2 mV in leads V1, V2 or V3 OR => 0.1mV in other leads. OR 2. Development of any new Q wave in leads V1 through V3 OR a new Q wave with duration =>0.03 sec and => 1mm deep in any other two contiguous leads. That is: Mere elevations of troponin I, even to the extent defined above. If unaccompanied by ECG changes or CLINICAL CONCOMITANTS as described above are NOT to be reported as MI. So that, BIOCHEMICAL CHANGES alone do not define an MI. The accuracy with which MI can be defined increases with the proximity of the MI. Therefore, episodes of MI nominated as being <6hrs, 6-24hrs, 24hrs-7days, 8-21days demand absolute accuracy. Remote MI, >21 days, are historical. It may not be possible to diagnose STEMI unless there are residual ECG changes OR an epicardial scar is seen at operation.	
MI_W	Time period between the last documented myocardial infarction and surgery.	Numeric
CCS	Angina – CCS Class Classification. Canadian Cardiovascular Society Classification. The highest class leading to current episode of hospitalisation and/or intervention:	Numeric
ANGRXG	Treatment for Unstable Angina (on day of surgery) includes i-v GTN.	Numeric
ANGRXH	Treatment for Unstable Angina (given ≤ 12 hours prior to surgery) includes i-v Heparin.	Numeric
ANGRXC	Treatment for Unstable Angina (given ≤ 24 hours prior to surgery) includes s.c. clexane at ≥ 1mgm/Kg bd (include Other low Mol. Wt. Heparinoids), if equivalent dose.	Numeric
CHF	Whether a physician has ever diagnosed Congestive Heart Failure (CHF)	Numeric
CHF_C	The diagnosis and management of CHF was made this admission, OR The management changed due to deterioration in CHF.	Numeric
NYHA	NYHA: New York Heart Association Class - the highest level leading to current episode of hospitalisation and/or procedure. (Note: Dyspnoea in a patient who has only CAD, should be considered an angina equivalent and therefore indicated by the CCS class. The NYHA class should be marked as I in such patients)	Numeric
SHOCK	Cardiogenic Shock.Is the patient, at the time of procedure, in a clinical state of hypoperfusion according to either of the following criteria: Clinical criteria for cardiogenic shock are: a. Hypotension (a systolic blood pressure < 90 mmHg) &/or OR CI <2.0 for at least 30 minutes OR b. the need for supportive measures to maintain a systolic pressure > or = 90 mmHg or a CI > 2.0	Numeric
RESUS	The patient required cardiopulmonary resuscitation, or initiation of treatment for cardiogenic shock, within one hour before the start of the operative procedure.	Numeric
ARRT	Was there a pre-operative arrhythmia present by clinical documentation	Numeric
ARRT_A	Atrial fibrillation or flutter requiring treatment.	Numeric
ARRT_AT	The type of atrial fibrillation (AF)	Numeric
ARRT_H	Complete Heart Block (AV dissociation)	Numeric

ARRT_V	Sustained or recurrent ventricular tachycardia or ventricular fibrillation, requiring cardioversion and/or IV therapy	Numeric
ARRT_O	Other arrhythmia (e.g. Sick Sinus Syndrome, Wenckebach, 2:1 Heart Block) present.	Numeric
PACE	Patient has a permanent pacemaker implanted.	Numeric
MEDIN	Patient on inotropes on day of surgery (on drug when entered OR), for haemodynamic support excluding renal dose Dopamine.	Numeric
MEDNI	Patient on IV Nitrates on day of surgery (on drug when entered OR).	Numeric
MEDAC	Patient given warfarin/heparin/low MW heparinoid ≤ 24 hours prior to surgery	Numeric
MEDST	Patient given systemic steroids on day of surgery (on drug when entered OR).	Numeric
MED_ASP	Patient has taken aspirin in the 7 days prior to surgery.	Numeric
	(ATACAS is to be selected for those patients involved in the ATACAS trial where use is not yet known.)	Numeric
MED_CLOP	Patient has taken clopidogrel in the 7 days prior to surgery.	Numeric
MED_ABCI	Patient has taken abciximab in the 7 days prior to surgery	Numeric
MED_AGG	Patient has taken aggrostat therapy in the 7 days prior to surgery	Numeric
MED_OTH	Patient has taken other antiplatelet therapy in the 7 days prior to surgery.	Numeric
ASPdays	Time when last documented aspirin given.	Numeric
CLOPdays	Time when last documented clopidogrel therapy given.	Numeric
ABCIdays	Time when last documented abciximab given.	Numeric
AGGdays	Time when last documented aggrostat given	Numeric
OTHdays	Time when last documented other antiplatelet therapy given.	Numeric

Previous Interventions

Column Name	Meaning	Data Type
POP	Has the patient undergone any previous cardiovascular intervention, surgical or non-surgical including those done during the current admission. Includes all forms of percutaneous angioplasty and thrombolytic therapy for cardiac indications. If the patient has had for example a PTCA Stent at another hospital and was then transferred to this hospital for surgery; ie. same admission episode.	Numeric
PBYP	Prior to this operation, how many cardiac surgical operations were performed on this patient utilising cardiopulmonary bypass.	Numeric
PBH	Prior to this operation, how many cardiac surgical operations were performed on this patient without cardiopulmonary bypass.	Numeric
PCAB	Previous Coronary Artery Bypass surgery by any approach.	Numeric
POPCAB	Previous coronary artery bypass surgery performed without the use of cardiopulmonary bypass.	Numeric
PVAL	Previous surgical replacement and/or repair of a cardiac valve, by any approach.	Numeric
PCOP	Any other previous cardiac surgery, including operation on the ascending aorta and/or aortic arch, including pericardiectomy?	Numeric
PCI	Was Percutaneous Transluminal Coronary Angioplasty, Coronary Atherectomy, and/or coronary Stent done at any time prior to this surgical procedure (which may include during the current admission)?	Numeric

PCI_ADM	Indicate whether PTCA/STENT was done at this admission or prior admission.	Numeric
PCI_H	The time between PTCA/Atherectomy/Stent and surgical repair of coronary occlusion. No. of hours if during same admission (includes direct transfer patients from hospital where stent was inserted to hospital where operation was performed).	Numeric
THRM	Was Thrombolytic treatment given for cardiac indications at any time prior to this surgical procedure, during this admission? (includes direct transfer patients from hospital where thrombolysis was given to hospital where operation was performed).	Numeric
THRM_H	The time between thrombolysis treatment and surgical repair of coronary occlusion. No. of hours if during same admission (includes direct transfer patients from hospital where stent was inserted to hospital where operation was performed). No. of hours	Numeric
PBALL NUMERIC N 1 = Yes 0 = No Must not be Null if POP = 1	Has the patient had a previous non-surgical Balloon Valvuloplasty at any time?	Numeric
PCI_ASD	**PREVIOUS PERCUTANEOUS INTERVENTION – ASD DEVICE CLOSURE** Closure by percutaneous technique of Atrial Septal Defect at any time	Numeric
PCI_VSD	**PREVIOUS PERCUTANEOUS INTERVENTION – VSD DEVICE CLOSURE** Closure by percutaneous technique of Ventricular Septal Defect, at any time	Numeric
SVT	Patient had percutaneous ablation procedure for any form of SVT or VT.	Numeric

Haemodynamic Data

Column Name	Meaning	Data Type
HTM	Height in centimetres in bare or stockinged feet.	Numeric
WKG	Weight in kilograms in light clothing and stockinged feet.	Numeric
CATH	Has the patient had a cardiac catheter for angiogram or pressure study.	Numeric
CATH_W	The date the patient had a cardiac catheter inserted.	Date
EF_T	Was the Left Ventricular Ejection Fraction measured, and how was this information obtained?	Numeric
EF	The percentage of the blood emptied from the left ventricle at the end of the contraction. Use the most recent determination prior to intervention.	Numeric
EF_EST	**EF ESTIMATE** If Nuclear scan, echo or angiogram did not yield a digital EF %, provide an estimate from reviewing the study. Choose ONE of: **Normal** (LV-EF > 60%) **Mild Impairment**(EF 46-60%) **Moderate**(EF 30-45%) **Severe** (EF<30%)	Numeric

LMD	**LEFT MAIN STENOSIS > 50%** Any stenosis that involves any parts of the Left Main. Left Main Coronary stenosis is present when there is > 50% compromise of vessel diameter in any angiographic view.	Numeric
DISVES	The number of major coronary systems	Numeric

Operative Status/Category

Column Name	Meaning	Data Type
SURG	The surgeon whom is ultimately responsible for the operation. (As per individual surgeon code)	Numeric
PROC	The surgeon who performs all or the majority of the "specific" part of the operation. In this context, "specific" refers to the actual coronary artery graft, valve or other cardiac procedure. Which of the following was the operating surgeon? Consultant Senior Registrar: (FRACS) Trainee or accredited Registrar Overseas Fellow Oversight: Foreign graduate oversight supervision	Numeric
STAT	Status. Select one of the folowing: a. Elective: The procedure could be Deferred without increased risk of compromised cardiac outcome. b. Urgent: Not routine – medical reason for operating this admission – a) within 72 hours from angiography if on the same admission that angiography was performed (in this case, "same admission" includes the situation when angiography is performed at another hospital and the patient is transferred directly to the hospital where surgery is to be performed) OR b) within 72 hours after an unplanned admission (in a patient who had a previous angiogram and was scheduled for surgery but was admitted acutely). c. Emergency: Unscheduled surgery required in next available theatre on same day due to refractory angina or cardiac compromise d. Salvage: The patient is undergoing CPR en route to the operating room, that is, prior to surgical incision.	Numeric
DTCATH	DIRECT TRANSFER FROM CATH LAB TO THEATRE As a result of a cardiac catheter lab complication, in the opinion of the operator or the responsible physician, the patient needed to be moved directly to surgery from the cath lab or hospital ward. Typically due to indications such as ongoing ischaemia, rest angina despite maximal treatment, pulmonary oedema requiring intubation, or shock.	Numeric
CCAB	Current Surgical Procedure is Coronary Artery Bypass	Numeric
CVLV	Current Surgical Procedure is Valve Surgery	Numeric
COTH	Current Surgical Procedure is Cardiac surgery other than Valve surgery or Coronary Artery Bypass.	Numeric
LVA	Current Surgical Procedure is for Left Ventricular Aneurysm (LVA).	Numeric
VSD	Current Surgical Procedure is for the correction of an acquired (usually ischaemic) Ventricular Septal Defect	Numeric
ASD	Current Surgical Procedure is for the correction of an Atrial Septal Defect (Excludes closure of incidental PFO)	Numeric
TRAUMA	Current Surgical Procedure is for the repair of Cardiac Trauma.	Numeric

COTHNS	Current Surgical Procedure is a cardiac procedure not otherwise specified.	Numeric
HOCM	This procedure is performed for either hypertrophic obstructive cardiomyopathy or left ventricular muscular dynamic LVOT obstruction, or in cases of tunnel stenosis in the left ventricular outflow tract. This procedure involves excision of left ventricular endocardial muscle from the left ventricular outflow tract.	Numeric
LVRR	This is ischaemic rupture of the free wall of the left ventricle. Therefore does not include traumatic rupture.	Numeric
PCAR	Current Surgical Procedure is a pericardiectomy	Numeric
PTE	Operation performed for chronic pulmonary thrombo-embolic disease. It involves cardiopulmonary bypass, and mostly hypothermic circulatory arrest, and incisions are made in the right and left (or both) pulmonary arteries, and an endartectomy performed out into the distal branches.	Numeric
LVRECON	Reshaping of the left ventricle by lateral excision (Bastista) or antero-septal reconstruction (Dor). Does not include resection and repair of chronic left ventricular aneurysm, by whatever technique.	Numeric
PULEMB	Current surgical procedure is for pulmonary embolectomy	Numeric
TUMOUR	Current Surgical Procedure is for removal of Cardiac Tumour(s).	Numeric
CT	Current Surgical Procedure is Cardiac Transplant.	Numeric
OTHCON	Current Surgical Procedure is for a congenital complication not otherwise specified.	Numeric
PLVEL	Insertion of a permanent LV Epicardial Lead in association with a cardiac procedure.	Numeric
AAS	Current surgical procedure is for paroxysmal, persistent or permanent atrial tachy arrhythmia. If Yes, completion of “Lesion Set” and “Energy Source” is compulsory.	Numeric
LESION1	ATRIAL ARRHYTHMIA SURGERY- LESION SET Choose the ONE predominant Lesion Set from the following: 1 = Cox-Maze III 2 = Radial 3 = Mini-Maze 4 = Left Atrial Reduction 5 = Pulmonary Vein Isolation 6 = Left Atrial Only 7 = Right Atrial only 8 = Other	Numeric
ES1	ATRIAL ARRHYTHMIA SURGERY- ENERGY SOURCE. Choose the ONE predominant technique used of creating the lines of conduction block from the following: 1 = Cut & Sew 2 = Unipolar RF 3 = Bipolar RF 4 = Cryoblation 5 = Microwave 6 = Laser 7 = Ultrasound 8 = Other	Numeric
AO	Current Surgical Surgery is Aortic Procedure.	Numeric
AOAN	Aortic Aneurysm repair.	Numeric
AOAN_A	Aortic Aneurysm involving the ascending aorta.	Numeric
AOAN_H	Aortic Aneurysm involving the aortic arch.	Numeric
AOAN_D	Aortic Aneurysm involving the descending thoracic aorta.	Numeric
AOAN_T	Aortic Aneurysm involving the thoraco-abdominal aorta.	Numeric
AODS	Aortic Dissection repair.	Numeric
AODS_T	Aortic dissection type. Choose one of the following: Ascending / Descending	Numeric

AODS_A	Was the dissection acute – occurring within last 2 weeks? Choose one of the following: Acute (≤2 weeks) Non-Acute (>2 weeks)	Numeric
AOTS	Indication for surgery is traumatic aortic transection – occurring within last 2 weeks.	Numeric
NCOTH	Current Surgical Procedure is Non Cardiac	Numeric
CEND	Surgical removal of stenotic atheromatous plaque	Numeric
ROLU	Surgical removal of a damaged or diseased portion of a lung	Numeric
VOTH	Procedures correcting peripheral vascular occlusion	Numeric
TOTH	Procedures involving Thorax/pleura.	Numeric
OTH	Any other concomitant surgery not covered in the non-cardiac procedures above.	Numeric

Minimally Invasive

Column Name	Meaning	Data Type
MIN	Was a non-standard incision used to minimise thoracic trauma, either for beating heart off-pump coronary artery procedure or for an on-pump cardiac procedure utilising any form of cardiopulmonary bypass?	Numeric
MIN_I	What was the indication for attempting a minimally invasive technique? Select ONE of the following: Surgeon/Patient choice Contraindications for standard incision Combined with Catheter Intervention	Numeric
OFFP	The response to the off pump procedure question will determine the setting/response for CPBP section.	Numeric
ROBOT	Any procedure performed with the assistance of a robot	Numeric

CPB and Support

Column Name	Meaning	Data Type
CPB	Was cardiopulmonary bypass used?	Numeric
CPLEG	Was cardioplegia used?	Numeric
CCT	CUMULATIVE CROSS CLAMP TIME. Total number of minutes the aorta is completely cross-clamped and the heart was ischaemic during bypass. Enter zero if no cross clamp was used.	Numeric
PERF	CUMULATIVE CARDIOPULMONARY BYPASS TIME (PERFUSION TIME) Total number of minutes on cardiopulmonary bypass. Enter zero if no cardiopulmonary bypass was used.	Numeric
IABP	Was the patient placed on Intra-Aortic Balloon Pump (IABP)?	Numeric
IABP_W	What was the time of earliest IABP insertion? Choose ONE of the following: Pre-Operatively: (before patient enters the operating theatre) Intra-Operatively Post-Operatively	Numeric
IABP_I	What was the PRIMARY reason for inserting the IABP? Choose one of the following: Haemodynamic Instability PTCA Support Unstable Angina CBP Wean: Cardiopulmonary bypass (CPB) weaning failure Prophylactic	Numeric
RPUMP	Was a Rota-pump used?	Numeric
RPUMP_W	What was the time of earliest use of the rota-pump? Choose ONE of the following: Pre-Operatively Intra-Operatively Post-Operatively	Numeric

PUMP_I	What was the PRIMARY reason to use the rota-pump? Choose one of the following: Haemodynamic Instability PTCA Support Unstable Angina CBP Wean Prophylactic	Numeric
VAD	Was a VAD/ECMO used at the time the patient left the operating room?	Numeric
VAD_W	What was the time of earliest use of the VAD/ECMO? Choose ONE of the following: Pre-Operatively Intra-Operatively Post-Operatively	Numeric
VAD_I	What was the PRIMARY reason to use the VAD/ECMO? Choose one of the following: Haemodynamic Instability PTCA Support Unstable Angina CBP Wean Prophylactic	Numeric
IOTOE	Trans-oesophageal Echocardiography performed during cardiac surgery procedure?	Numeric
ELECT	Indicate the insertion method. Elective Insertion = Routine Insertion of TOE, planned before commencement of operation. Non-Elective Insertion = Unplanned insertion of TOE, for whatever reason.	Numeric
ANTIFIB	Antifibrinolytic used	Numeric
ANTIFIB_T	INTRA-OPERATIVE ANTIFIBRINOLYTIC USE- TYPE	Numeric

Coronary Bypass

Column Name	Meaning	Data Type
IDGCA	Decision made to graft coronary artery(ies) for intra-operative reasons.	Numeric
IMA	Was an Internal Mammary Artery Used for Coronary Bypass?	Numeric
LIMA	Was a Left Internal Mammary Artery Used for Coronary Bypass?	Numeric
RIMA	Was a Right Internal Mammary Artery Used for Coronary Bypass?	Numeric
DAN_AC	The total number of distal anastomoses with arterial conduits, whether IMA, GEPA, radial artery, etc.	Numeric
DANIMA	Total number of distal anastomoses done using internal mammary artery grafts.	Numeric
RAC	Total number of radial artery conduits used.	Numeric
RDAN	Total number of radial distal anastomoses.	Numeric
DANV	The total number of distal anastomoses with venous conduits, e.g. saphenous veins.	Numeric
DANGEP	Total number of Gastro-Epiploic Artery (GEPA) distal anastomoses.	Numeric
ATY	Was any form of T or Y graft used between segments of arterial conduit?	Numeric
DAN	The total number of distal anastomoses.	Numeric

Valve Surgery

Column Name	Meaning	Data Type
AOPROC	Was a surgical procedure done on the Aortic Valve, and if so, what? Select ONE of the following: 1: No 3: Replacement 5: Repair/Reconstruction without annuloplasty 6: Root Reconstruction with Valve Conduit (Bentall procedure) 7: Root Reconstruction with Valve Sparing (David procedure) 8: Resuspension Aortic Valve 9: Resection Sub-Aortic Stenosis 12: Repair Paravalvular leak 14: Valvotomy 15: Ross Procedure 16: Inspection only 17: Decalcification of valve only	Numeric
AOIM	AORTIC VALVE PROSTHESIS – IMPLANT – Manufacturer's model number Select from the prosthesis list	Text
AOIM_SR	AORTIC VALVE PROSTHESIS – IMPLANT – Serial number	Text
AOIM_S	The size of the aortic prosthesis implant.	Numeric
AOEX	AORTIC VALVE PROSTHESIS - EXPLANT – Manufacturer's model number Select from the prosthesis list	Text
AOEX_SR	AORTIC VALVE PROSTHESIS – EXPLANT - Serial number	Text
AOEX_S	The size of the aortic prosthesis explant.	Numeric
AOSTEN	Aortic stenosis that clinically warrants valve replacement.	Numeric
AOREG	Is there evidence of Aortic valve regurgitation? 0: None 1: Trivial 2: Mild 3: Moderate 4: Severe	Numeric
AOPATH	Where there is a congenital pathology/aetiology such as a bicuspid aortic valve in combination with either a myxomatous degeneration or calcific pathology then this should be recorded as congenital.	Numeric
	1 Rheumatic 2 Congenital 3 Ischaemic 4 Idiopathic Calcific 5 Myxomatous degen. 6 Failed prior repair 7 Prosthetic valve failure 8 Peri-prosthetic leak 9 Prosthetic valve thrombosis 10 Active infection	
	11 Previous infection 12 Marfans 13 Annuloaortic ectasia 14 Other degen. Disease 15 Dissection 16 Tumour 17 Trauma 18 Iatrogenic 99 Other	
MIPROC	Was a surgical procedure done on the Mitral Valve, and if so, what? Select ONE of the following: 1: No 2: Annuloplasty Only 3: Replacement. 4: Repair/Reconstruction with Annuloplasty 5: Repair/Reconstruction without Annuloplasty 10: Commissurotomy with annuloplasty ring 11: Commissurotomy without annuloplasty ring 12: Repair Paravalvular leak 16: Inspection only 17: Decalcification of valve only	Numeric
MIIM	MITRAL VALVE PROSTHESIS - IMPLANT – Manufacturer's model number Select from the prosthesis list	Text
MIIM_SR	MITRAL VALVE PROSTHESIS – IMPLANT – Serial number	Text
MIIM_S	The size of the mitral prosthesis implant.	Numeric
MIEX	MITRAL VALVE PROSTHESIS - EXPLANT – Manufacturer's model number Select from the prosthesis list	Text
MIEX_SR	MITRAL VALVE PROSTHESIS – EXPLANT - Serial number	Text
MIEX_S	The size of the mitral prosthesis explant.	Numeric
MISTEN	Mitral stenosis warranting surgical correction.	Numeric
MIREG	Is there evidence of Mitral valve regurgitation? 0: None 1: Trivial 2: Mild 3: Moderate 4: Severe	Numeric

MIPATH	What is the aetiology of the mitral valve lesion? 1 Rheumatic 2 Congenital 3 Ischaemic 4 Idiopathic Calcific 5 Myxomatous degen. 6 Failed prior repair 7 Prosthetic valve failure 8 Peri-prosthetic leak 9 Prosthetic valve thrombosis 10 Active infection 11 Previous infection 12 Marfans 14 Other degen. Disease 16 Tumour 17 Trauma 18 Iatrogenic 19 “Functional” or Isolated Annular Dilatation (Def: MR due to annular dilatation, without leaflet or sub-valvular abnormality). 99 Other	Numeric
TRPROC	Was a surgical procedure done on the Tricuspid Valve, and if so, what? Select ONE of the following: 1: No 2: Annuloplasty Only 3: Replacement 4: Repair/Reconstruction with Annuloplasty 5: Repair/Reconstruction without Annuloplasty 10: Commissurotomy with annuloplasty ring 11: Commissurotomy without annuloplasty ring 12: Repair Paravalvular leak 13: Valvectomy (no replacement) 16: Inspection only	Numeric
TRIM	TRICUSPID VALVE PROSTHESIS – IMPLANT – Manufacturer’s model number Select from the prosthesis list	Text
TRIM_SR	TRICUSPID VALVE PROSTHESIS – IMPLANT – Serial number	Text
TRIM_S	The size of the tricuspid prosthesis implant.	Numeric
TREX	TRICUSPID VALVE PROSTHESIS - EXPLANT – Manufacturer’s model number Select from the prosthesis list	Text
TREX_SR	TRICUSPID VALVE PROSTHESIS – EXPLANT - Serial number	Text
TREX_S	The size of the tricuspid prosthesis explant.	Numeric
TRSTEN	Tricuspid stenosis warranting surgical correction	Numeric
TRREG	Is there evidence of Tricuspid valve regurgitation? 0: None 1: Trivial 2: Mild 3: Moderate 4: Severe	Numeric
TRPATH	What is the aetiology of the tricuspid valve lesion? 1 Rheumatic 2 Congenital 3 Ischaemic 4 Idiopathic Calcific 5 Myxomatous degen. 6 Failed prior repair 7 Prosthetic valve failure 8 Peri-prosthetic leak 9 Prosthetic valve thrombosis 10 Active infection 11 Previous infection 12 Marfans 14 Other degen. Disease 16 Tumour 17 Trauma 18 Iatrogenic 99 Other 20 Functional	Numeric
PUPROC	Was a surgical procedure done on the Pulmonic Valve, and if so, what? Select ONE of the following: 1: No 3: Replacement 5: Repair/Reconstruction without annuloplasty 11: Commissurotomy without annuloplasty ring 12: Repair Paravalvular leak	Numeric
PUIM	PULMONARY VALVE PROSTHESIS - IMPLANT – Manufacturer’s model number Select from the prosthesis list	Text
PUIM_SR	PULMONARY VALVE PROSTHESIS – IMPLANT - Serial number	Text
PUIM_S	The size of the pulmonary prosthesis implant.	Numeric
PUEX	PULMONARY VALVE PROSTHESIS - EXPLANT – Manufacturer’s model number Select from the prosthesis list	Text
PUEX_SR	PULMONARY VALVE PROSTHESIS – EXPLANT - Serial number	Text
PUEX_S	The size of the pulmonary prosthesis explant.	Numeric
PUSTEN	Pulmonary stenosis warranting surgical correction.	Numeric
PUREG	Is there evidence of pulmonary valve regurgitation? 0: None 1: Trivial 2: Mild 3: Moderate 4: Severe	Numeric
PUPATH	What is the aetiology of the pulmonary valve? 1 Rheumatic 2 Congenital 3 Ischaemic 4 Idiopathic Calcific 5 Myxomatous degen. 6 Failed prior repair 7 Prosthetic valve failure 8 Peri-prosthetic leak 9 Prosthetic valve thrombosis 10 Active infection 11 Previous infection 12 Marfans 14 Other degen. Disease 16 Tumour 17 Trauma 18 Iatrogenic 99 Other	Numeric

Post-Operative Data

Column Name	Meaning	Data Type
RBC	Were Red Blood Cells transfused intra- and/or post-operatively, that is, from the commencement of surgery to discharge?	Numeric
NRBC	Was a transfusion of blood products other than RBC (eg. FFP, Platelets) given intra- and/or post-operatively, that is, from the commencement of surgery to discharge?	Numeric
RBCUnit	Indicate the number of Bank RBC Units used.	Numeric
PlateUnit	Indicate the number of Novo 7 Units used.	Numeric
NovoUnit	Indicate the number of Novo 7 Units used.	Numeric
FFPUnit	Indicate the number of Cryo Units used.	Numeric
CryoUnit	Indicate the number of Cryo Units used.	Numeric
AICU_D	Indicate the date and time of admission to ICU from OR.	Date/Time
DICU_D	Indicate the date and time of discharge from ICU to HDU or General Ward or death.	Date/Time
EXTU_D	Indicate the date post-operation when the patient was extubated.	Date/Time
REICU	Was patient readmitted to ICU following transfer to the HDU or General Ward?	Numeric
REINT	Indicate whether the patient was reintubated during hospital stay after the initial extubation.	Numeric
REINT_D	Indicate the date and time when the patient was reintubated.	Date/Time
REEXT_D	Indicate the date and time when the patient was extubated following the reintubation.	Date/Time
DRAIN_4	Indicate the fluid loss in mls from the Pericardial/mediastinal drains in the first 4hrs post-operation.	Numeric

Post-Operative Data – Complications

Column Name	Meaning	Data Type
RTT	Did patient return to the operating theatre for management of complications? Includes operative procedures done in the ICU that normally would be performed in the operating theatre.	Numeric
ROVD	Operative re-intervention was required for valve dysfunction.	Numeric
ROBL	Operative re-intervention was required for bleeding/ tamponade.	Numeric
ROGO	Operative re-intervention was required to refashion a graft or graft a previously ungrafted coronary.	Numeric
ROSI	Operative re-intervention for infection of sternal bone, muscle and/or mediastinum	Numeric
ROOC	Operative re-intervention was required for other cardiac reasons.	Numeric
RONC	Operative re-intervention was required for other non-cardiac reasons.	Numeric
NRF	Acute post-operative renal insufficiency resulting in two or more of the following: a. Increased serum creatinine to >0.2 mmol/l (>200 µmol/l). b. A doubling or greater increase in creatinine over baseline pre-operative value. c. A new requirement for dialysis/haemofiltration.	Numeric
HAEMOFIL	Acute institution of haemofiltration for renal failure. Excludes haemofiltration for removal of fluid with normal serum urea and creatinine.	Numeric
POSTCR	Highest Serum creatinine recorded after surgery.	Numeric

POMI	A peri-operative Myocardial Infarction (MI) is diagnosed by finding at least two of the following three criteria: a. Enzyme level elevation: either (1) CK-MB >30; or (2) troponin >20.0 micrograms /L or troponin level equivalent documented at your institution, provided operation does not involve myocardial incision. b. New wall motion abnormalities c. Serial EGG (at least two) showing new Q waves, duration ≥0.03ms in 2 contiguous leads.	Numeric
POCS	Clinical criteria for cardiogenic shock are: a. Hypotension (a systolic blood pressure < 90 mmHg &/or OR CI <2.0 for at least 30 minutes b. or the need for supportive measures to maintain a systolic pressure > or = 90 mmHg or a CI > 2.0	Numeric
CIUSE	Any inotrope use for longer than 4 hours post-operatively.(Includes Dopamine at > 300 μg/min)	Numeric
IULowOut	Inotrope use for Low Cardiac Output Syndrome: when an inotrope is administered with the intent to improve cardiac output, irrespective of the reasons for that decision.	Numeric
IULowSVR	Inotrope use for Low Systemic Vascular Resistance Syndrome: when a primarily alpha adrenergic agonist is given with the intent to increase SVR if SVR < 800. This is usually in presence of high cardiac output. Does not include Noradrenalin given with Milrinone.	Numeric
NARRT	Did any new form of cardiac arrhythmia occur that required treatment?	Numeric
HB	New heart block requiring the implantation of a permanent pacemaker prior to discharge.	Numeric
BA	New brady-arrhythmia not otherwise specified requiring the implantation of a permanent pacemaker prior to discharge.	Numeric
CA	A new cardiac arrest documented by one of the following: a. ventricular fibrillation b. rapid ventricular tachycardia with haemodynamic instability c. asystole.	Numeric
AFIB	New onset of atrial fibrillation/flutter (AF) requiring treatment. Does not include recurrence of AF, which was present pre-operatively.	Numeric
NARRTV	Did any new form of ventricular tachycardia (greater than 6 beat run) occur that required treatment?	Numeric
CVA_P	A new central neurologic deficit persisting for > 72 hours.	Numeric
CVA_T	A new transient central neurologic deficit that resolves completely within 72 hours (TIA, RIND).	Numeric
COMA	New post-operative coma that persists for at least 24 hours in a non-sedated patient	Numeric
VENT_P	Post Operative Pulmonary Insufficiency requiring ventilatory support - includes (but not limited to) causes such as ARDS and pulmonary oedema- for a total period of longer than 24 hours. Use cumulative period if patient re-intubated.	Numeric
PUEMB	New Pulmonary Embolism diagnosed by study such as V/Q scan or angiogram.	Numeric
PUPNU	Pneumonia diagnosed by one of the following: positive cultures of sputum or trans-tracheal aspirate and consistent with clinical findings of pneumonia (should include radiological changes).	Numeric
PURINT	Was re-intubation required for any reason during hospitalisation?	Numeric
INFDS	Involves muscle and bone, with or without mediastinal involvement, as demonstrated by surgical exploration. Must have: wound debridement and one of the following: a. Positive culture b. Treatment with antibiotics	Numeric
INFTH	An infection involving a thoracotomy or parasternal site.	Numeric
INFSP	Septicaemia requires positive blood cultures supported by at least two of the following indeces of clinical infection: A. Fever B. Elevated granulocyte cell counts C. Elevated and increasing CRP D. Elevated and increasing ESR, post-operatively.	Numeric

NAODS	Dissection occurring in any part of the aorta.	Numeric
LISCH	Any complication producing limb ischaemia?	Numeric
ACOAG	Any bleeding, haemorrhage, and/or embolic events related to anticoagulant therapy?	Numeric
GIT	Post-operative occurrence of any GI complication including: a. GI bleeding requiring transfusion b. Pancreatitis with abnormal amylase/lipase requiring nasogastric suction therapy c. Cholecystitis requiring cholecystectomy or drainage d. Mesenteric ischaemia requiring exploration e. Hepatic f. Other GI complication.	Numeric
MSF	Two or more of the following major organ systems to fail concurrently for at least 48 hours: Renal **-** New renal failure (defined previously); Respiratory **-** Requires endotracheal intubation for respiratory dysfunction; Cardiac **-** The use of inotropes and/or IABP to treat low cardiac output.	Numeric

Mortality/Readmission

Column Name	Meaning	Data Type
Dischar	Patient was discharged from the hospital following the admission during which the surgery occurred. 1 = Home 2 = Hospital in the Home 3 = Rehabilitation Unit/Hospital 4 = Local or referring hospital 5 = Hospital Mortality	Numeric
MORTPD	Specify whether the patient died within 30 days of surgery after discharge from hospital.	Numeric
MORT_D	Provide date of death in hospital during the index admission at any time after the procedure, or death after discharge from hospital within thirty days of the procedure.	Date
MORT_L	Specify the patient location at time of death: 1 = OR 2 = Hospital 3 = Home 4 = Other Facility	Numeric
MORT_R	Specify the PRIMARY cause of death, i.e. the first significant abnormal event which ultimately led to death; choose one of the following: 1 = Cardiac 2 = Neurologic 3 = Renal 4 = Vascular 5 = Infection 6 = Respiratory failure 8 = Multisystem failure 9 = Other 10 = Unknown 11 = Pulmonary 12 = Aortic	Numeric
MORT_SR	Specify SUBSEQUENT cause of death. This applies only when PRIMARY cause of death is cardiac or infection.	Numeric
WITHDRAW	Patient who was aware of the consequences to his/her actions, elected to withdraw treatment in circumstances where they would survive if treatment was continued. NOTE: Completing “YES” to this field implies automatic review of patient’s hospital file and permission for ASCTS personnel to review their case.	Numeric
READ	Patient was readmitted as an in-patient within 30 days from the date of surgery for ANY reason. (‘readmission’ means admission to general hospital not emergency, short-stay wards or planned transfer to rehabilitation facility) (Date of surgery counts as day 0)	Numeric
READAC	Reason the patient was readmitted as an in-patient within 30 days from the date of surgery was Anticoagulant Complication.	Numeric
READAR	Reason the patient was readmitted as an in-patient within 30 days from the date of surgery was Arrhythmia.	Numeric
READCHF	Reason the patient was readmitted as an in-patient within 30 days from the date of surgery was Congestive Heart Failure (CHF), by one of the following: a. Paroxysmal nocturnal dyspnoea (PND); b. Dyspnoea on exertion (DOE) due to heart failure; c. Chest X-ray (CXR) showing pulmonary congestion, OR d. Patient has received treatment for this – ACE inhibition, diuretics, Carvedilol or digoxin	Numeric

READVD	Reason the patient was readmitted as an in-patient within 30 days from the date of surgery was Valve Dysfunction.	Numeric
READPE	Reason the patient was readmitted as an in-patient within 30 days from the date of surgery was Pericardial Effusion.	Numeric
READCT	Reason the patient was readmitted as an in-patient within 30 days from the date of surgery was Cardiac Tamponade.	Numeric
READOTHC	Reason the patient was readmitted as an in-patient within 30 days from the date of surgery was Other Complication related to Cardiac Surgery (eg. renal, hepatic, GI, etc.).	Numeric
READDSI	Reason the patient was readmitted as an in-patient within 30 days from the date of surgery was Deep Sternal Infection; Involves muscle and bone, with or without mediastinal involvement, as demonstrated by surgical exploration.Must have: wound debridement and one of the following: a. Positive culture b. Treatment with antibiotics	Numeric
READIC	Reason the patient was readmitted as an in-patient within 30 days from the date of surgery was Other Incisional Complication.	Numeric
READRC	Reason the patient was readmitted as an in-patient within 30 days from the date of surgery was Pneumonia or other Respiratory Complication, diagnosed by one of the following: positive cultures of sputum or trans-tracheal aspirate and consistent with clinical findings of pneumonia (should include radiological changes).	Numeric
READMI	Reason the patient was readmitted as an in-patient within 30 days from the date of surgery was Myocardial Infarction (MI)	Numeric
READRA	Reason the patient was readmitted as an in-patient within 30 days from the date of surgery was Recurrent Angina. Objective confirmation that chest pain is due to ischaemia by exercise test (nuclear, echo, treadmill or angiography).	Numeric
READOTHNC	Primary reason the patient was readmitted as an in-patient within 30 days from the date of surgery was Other readmission unrelated to Cardiac Surgery.	Numeric

Automatic Data

Column Name	Meaning	Data Type
PatientID	PATIENT ID This is an arbitrary number, (not a recognisable ID like SSN or Medical Record Number) that uniquely and permanently identifies each patient. Once assigned to a patient, this can never be changed or reused. Generated by the system this must be blank for data transferring.	
AdmissionID	ADMISSION ID This is an arbitrary number, (not a recognisable ID like SSN or Medical Record Number) that uniquely and permanently identifies each admission. Once assigned to an admission, this can never be changed or reused. Generated by the system this must be blank for data transferring.	
OperationID	This is an arbitrary number, (not a recognisable ID like SSN or Medical Record Number) that uniquely and permanently identifies each operation. Once assigned to an operation, this can never be changed or reused. Generated by the system this must be blank for data transferring.	
AGE	Age of the patient at surgery.	Numeric
BMI	Body Mass Index calculated by the following equation. WKG / (HTM/100)2 Calculated automatically where height and weight is available.	Numeric
BSA	Body Surface area calculated by the following equation. (0.007184 x HTM0.725 x WKG0.425) Calculated automatically where height and weight is available.	Numeric
eGFR	ESTIMATED GLOMERULAR FILTRATION RATE	Numeric

ANG_T	Indicate the type of angina present at the time of surgery: Stable: Angina which is controlled by oral or transcutaneous medication. Unstable: The presence of ischemia that requires hospitalisation and use of intravenous nitrate, heparin therapy, s.c. clexane (include low Mol. Wt. Heparinoid) or intravenous Tyrofiban (Aggrostat) for control.	Numeric
PINT	The sum of all the intubation periods exceeds 24 hrs.	Numeric
VENT	Indicate the number of hours post operation for which the patient was ventilated. Calculate from the date and time of intubation to that of extubation. Round to the nearest hour eg 6 hours 25 minutes is rounded down to 6 hours, and 6 hours 35 minutes is rounded up to 7 hours. In the unlikely event that the time is exactly 30 minutes between the hour then round up. Use zero if the patient was extubated on the operating table. Do not count delayed re-intubation time.	Numeric
ICU	Number of hours spent by the patient in the ICU prior to transfer to the HDU or General Ward (does not include readmission to ICU). Round to the nearest hour eg 6 hours 25 minutes is rounded down to 6 hours, and 6 hours 35 minutes is rounded up to 7 hours. In the unlikely event that the time is exactly 30 minutes between the hour then round up.	Numeric
MORT30	Specify whether the patient died within 30 days after the procedure was performed. (Date of surgery counts as day 0; calculated from MORT_D-DOP)	Numeric
AOIM_T	AORTIC VALVE PROSTHESIS – IMPLANT - TYPE. Indicate the type of implant; choose ONE: N: None M: Mechanical B: Bioprosthesis A: Autograft H: Homograft/Allograft R: Ring/Ban	Text
AOEX_T	AORTIC VALVE PROSTHESIS – EXPLANT - TYPE. Indicate the type of implant; choose ONE: N: None M: Mechanical B: Bioprosthesis A: Autograft H: Homograft/Allograft R: Ring/Band	Text
MIIM_T	MITRAL VALVE PROSTHESIS – IMPLANT - TYPE. Indicate the type of implant; choose ONE: N: None M: Mechanical B: Bioprosthesis A: Autograft H: Homograft/Allograft R: Ring/Band	Text
MIEX_T	MITRAL VALVE PROSTHESIS – EXPLANT - TYPE. Indicate the type of implant; choose ONE: N: None M: Mechanical B: Bioprosthesis A: Autograft H: Homograft/Allograft R: Ring/Band	Text
TRIM_T	Indicate the type of implant; choose ONE: N: None M: Mechanical B: Bioprosthesis A: Autograft H: Homograft/Allograft R: Ring/Band	Text
TREX_T	TRICUSPID VALVE PROSTHESIS – EXPLANT - TYPE Indicate the type of implant; choose ONE: N: None M: Mechanical B: Bioprosthesis A: Autograft H: Homograft/Allograft R: Ring/Band	Text
PUIM_T	PULMONARY VALVE PROSTHESIS – IMPLANT - TYPE Indicate the type of implant; choose ONE: N: None M: Mechanical B: Bioprosthesis A: Autograft H: Homograft/Allograft R: Ring/Band	Text
PUEX_T	PULMONARY VALVE PROSTHESIS – EXPLANT - TYPE Indicate the type of implant; choose ONE: N: None M: Mechanical B: Bioprosthesis A: Autograft H: Homograft/Allograft R: Ring/Band	Text

4.3 PROCESS OF DEVELOPING DATA TABLE STRUCTURES AND SIMULATING DATA

Data Tables and Fields Chosen for Implementation

Patient Demographics	Patient Risk Factors	Pre-Operative Cardiac Status	Previous Interventions	Haemodynamic Data
LNAME	SMO_H	MI	POP	HTM
FNAME	SMO_C	MI_T	PBYP	WKG
MNAME	FHCAD	MI_W	PBH	CATH
OTHER FIELDS	OTHER FIELDS	OTHER FIELDS	OTHER FIELDS	OTHER FIELDS

Operative Status/Category	Minimally Invasive	CPB and Support	Coronary Bypass	Valve Surgery
SURG	MIN	CPB	IDGCA	AOPROC
PROC	MIN_I	CPLEG	IMA	AOIM
STAT	OFFP	CCT	LIMA	AOIM_SR
OTHER FIELDS	OTHER FIELDS	OTHER FIELDS	OTHER FIELDS	OTHER FIELDS

Post-Operative Data	Post-Operative Data/ Complications	Mortality/Readmission	Automatic Data
RBC	RTT	Dischar	PatientID
NRBC	ROVD	MORTPD	AdmissionID
RBCUnit	ROBL	MORT_D	OperationID
OTHER FIELDS	OTHER FIELDS	OTHER FIELDS	OTHER FIELDS

Data Tables Structure Development Process

Step 1:- SQL Data Definition Language (DDL) statements were created for all the selected tables. Used the given data types for each field in the ASCTS manual.

Step 2:- Grouped the chosen tables into three groups and implemented the DDL statements in three Database Management Systems as given below.

In SQL Server 2008 Instance

- Patient Demographics
- Patient Risk Factors
- Pre-Operative Cardiac Status
- Previous Interventions
- Haemodynamic Data

In Oracle 10g Express Instance

- Operative Status/Category
- Minimally Invasive
- CPB and Support
- Coronary Bypass
- Valve Surgery

In MySQL instance

- Post-Operative Data
- Post-Operative Data/ Complications
- Mortality/Readmission
- Automatic Data

Data Simulation Process

Step 1:- Created Temporary table structures for all the above mentioned tables in an SQL Server express instance.

Step 2:- Using appropriate regular expressions for each field in the tables, simulated 100 – 1000 records for each table using **Red Gate SQL Data Generator 2** software.

Step 3:- Populated the tables with simulated data.

Step 4:- Developed data migration scripts for migrating relevant tables for SQL Server 2008, Oracle 10g Express and MySQL.

Step 5:- Migrated the data into relevant tables in SQL Server 2008, Oracle 10g Express and MySQL Databases.

4.4 OVERVIEW OF THE PROPOSED DATA WAREHOUSING MODEL

Cardiac Surgery clinical data can be distributed across various disparate and heterogeneous clinical and administrative information systems. This makes accessing data highly time consuming and error prone

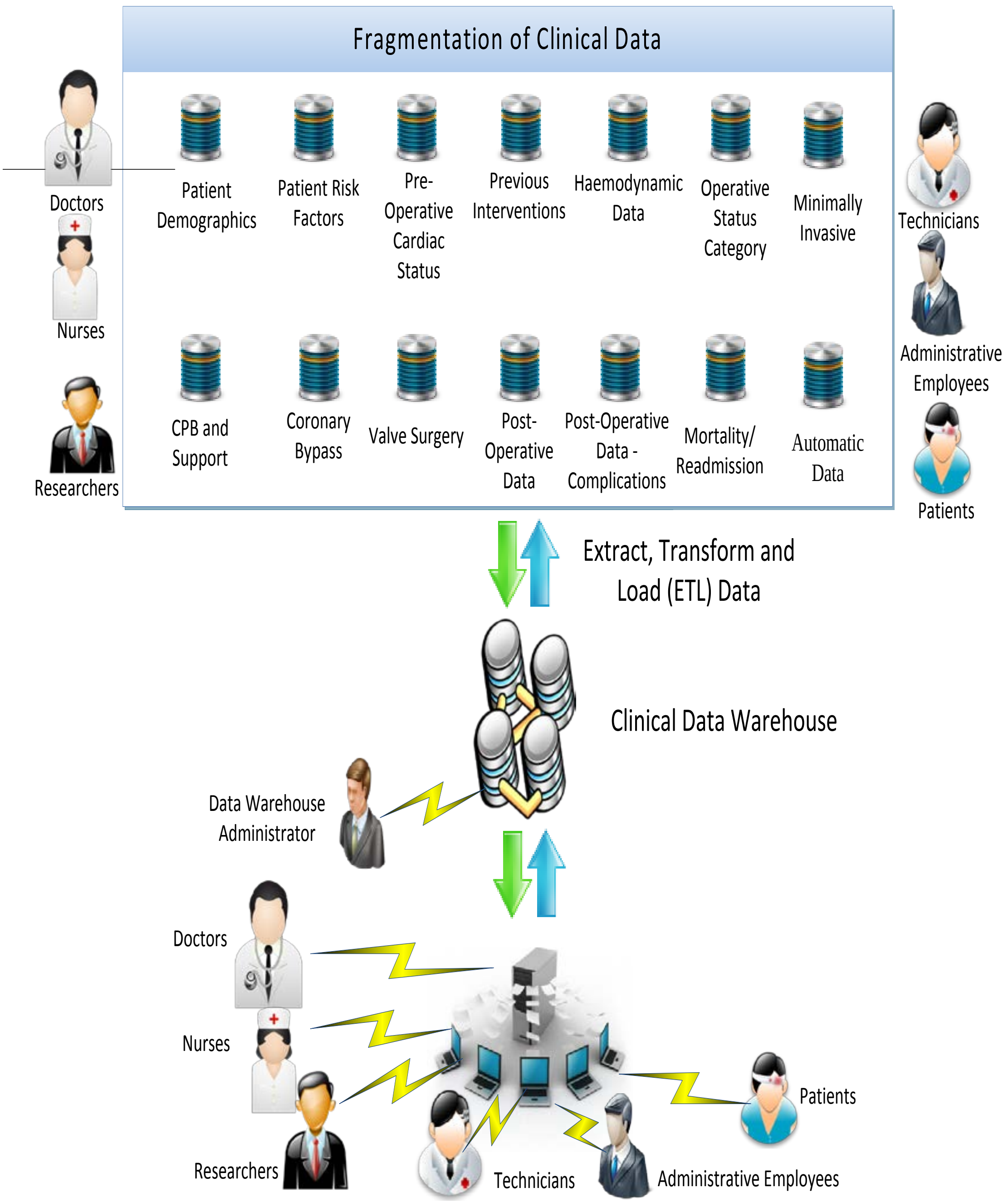

Figure 4.1: Data Warehousing Model for Integrating Fragmented Electronic Health Records from Disparate and Heterogeneous Cardiac Surgery Clinical Data Stores.

A data warehouse can be used to integrate the fragmented data sets. Once the data warehouse is created it should be populated with data through Extract, Transform and Load processes. Figure 4.1 shows a graphical overview of the proposed data warehousing model . The following sections of the chapter describe the selection of the suitable data warehousing model and creation of the selected model.

4.5 SELECTION OF THE DATA WAREHOUSING ARCHITECTURE AND THE DATA WARE HOUSING MODEL

4.5.1 META DATA

An important component of the DWH Environment is Meta data. Meta data, or data about data, provides the most effective use of the DWH. Meta data allows the end user to navigate through the possibilities. If a user approaches a data warehouse where there is no Meta data, the user does not know where to begin the analysis.

Meta data acts like an index to the data warehouse contents. It sits above the warehouse and keeps track of what is where in the warehouse. Typically, items the Meta data store tracks are as follows (Başaran, 2005):

- Structure of data as known to the programmer and to the DSS analyst
- Source data
- Transformation of data
- Data model
- DW
- History of extracts

4.5.2 DATA WAREHOUSES VS. DATA MARTS AND TYPES OF DATA WAREHOUSE ARCHITECTURES

Data Warehouses	Data Marts
Scope Application independent Centralized or Enterprise Planned	**Scope** Specific application Decentralized by group Organic but may be planned
Data Historical, detailed, summary Some denormalization	**Data** Some history, detailed, summary High denormalization
Subjects Multiple subjects	**Subjects** Single central subject area
Source Many internal and external sources	**Source** Few internal and external sources
Other Flexible Data oriented Long life Single complex structure	**Other** Restrictive Project oriented Short life Multiple simple structures that may form a complex structure

Table 4.1: Data Warehouses vs. Data Marts

(Kimball,Ross,Thornthwaite,Mundy,& Becker,2008 ;Inmon 2002)

Data warehousing methodologies share a common set of tasks, including business requirements analysis, data design, architecture design, implementation, and deployment (Kimball,Ross,Thornthwaite,Mundy,& Becker,2008 ;Inmon 2002).

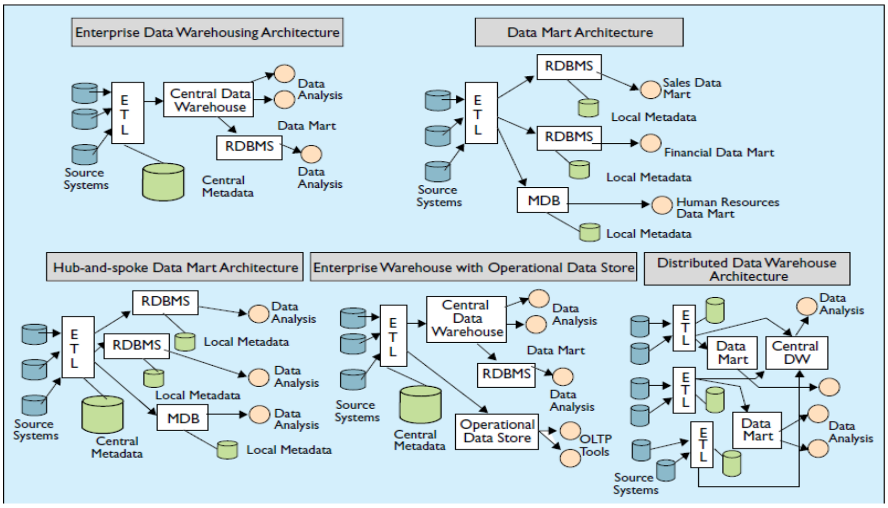

Figure 4.1: Different Types of Data Warehouse Architectures (Sen & Sinha, 2005)

In cardiac surgery related information systems, data repositories are scattered over limited disparate selected systems. They can be linked to data marts using ETL techniques. Considering the above characteristics of a Cardiac Surgery Unit, the most suitable deployment architecture is the **Enterprise Data Warehousing Architecture**.

4.5.3 SELECTION OF CONCEPTUAL AND LOGICAL DATA WAREHOUSE DESIGN APPROACHES

There are several approaches to design a data warehouse both in conceptual and logical design phases. The generally accepted conceptual design approaches are dimensional fact model, multidimensional E/R model, starER model and object-oriented multidimensional model. In the logical design phase, flat schema, terraced schema, star schema, fact constellation schema, galaxy schema, snowflake schema, star cluster schema and starflake schemas are widely used approaches. (Başaran, 2005)

Conceptual Design Models

The main objective of conceptual design modelling is to develop a formal, complete, abstract design based on the user requirements. The objective of creating a conceptual schema is to translate user requirements into an abstract representation understandable to the user that is independent of implementation issues, but is formal and complete, so that it can be transformed into the next logical schema without ambiguities (Phipps & Davis, 2002).

StarER Model

'*StarER model combines star structure with constructs of entity relationship (ER) model. Star structure has one fact table and a set of smaller dimension tables arranged around the fact table. The fact table is linked to all the dimension tables by one to many relationships. The main constructs of ER model are - the entity sets capturing real world objects, the relationship sets capturing associations among objects and, the attributes representing properties of entity or relationship set. StarER model has the following constructs: fact set, entity set, relationship set, attribute.*' (Mishra, Yazici & Başaran, 2008)

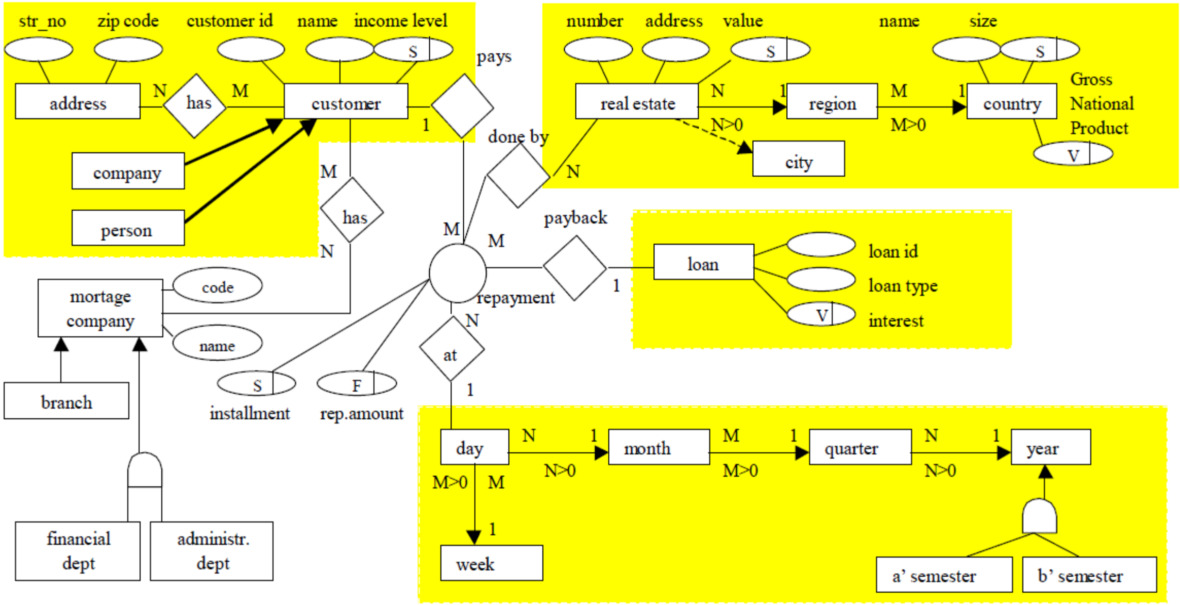

Figure 4.2: An Excerpt from a Mortgage Company Data Warehouse, Using the starER Model.

The financial and the administration department are outside the customer dimension as this is dependent on design decisions and semantic issues. (Tryfona, Busborg & Christiansen, 1999)

Multidimensional ER (ME/R) Model

'Multidimensional E/R (ME/R) model is an extension of the ER Model. There are few additional elements : A special entity set (dimension level), a special n-ary 'fact' relationship set and a special binary 'rolls-up to' relationship set.' (Mishra, Yazici & Başaran, 2008)

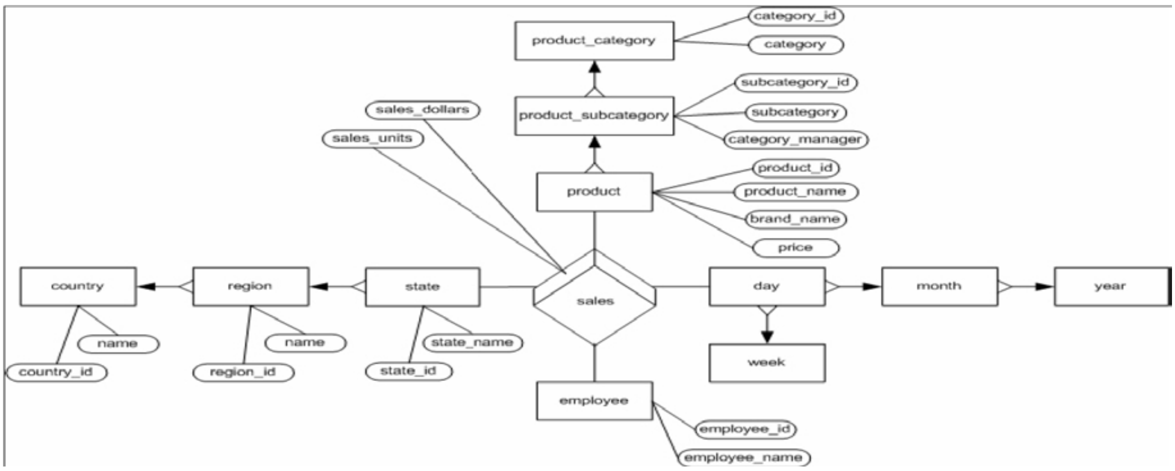

Figure 4.3: ME/R Model for Sales System

(Mishra, Yazici & Başaran, 2008)

Object Oriented Multidimensional (OOMD) Model

'*OOMD modelling approach is based on UML. In OOMD model, dimensions and facts are represented by dimension classes and fact classes. OOMD approach uses a generalization- specialization relationship to categorize entities that contain subtypes.*' (Mishra, Yazici & Başaran, 2008)

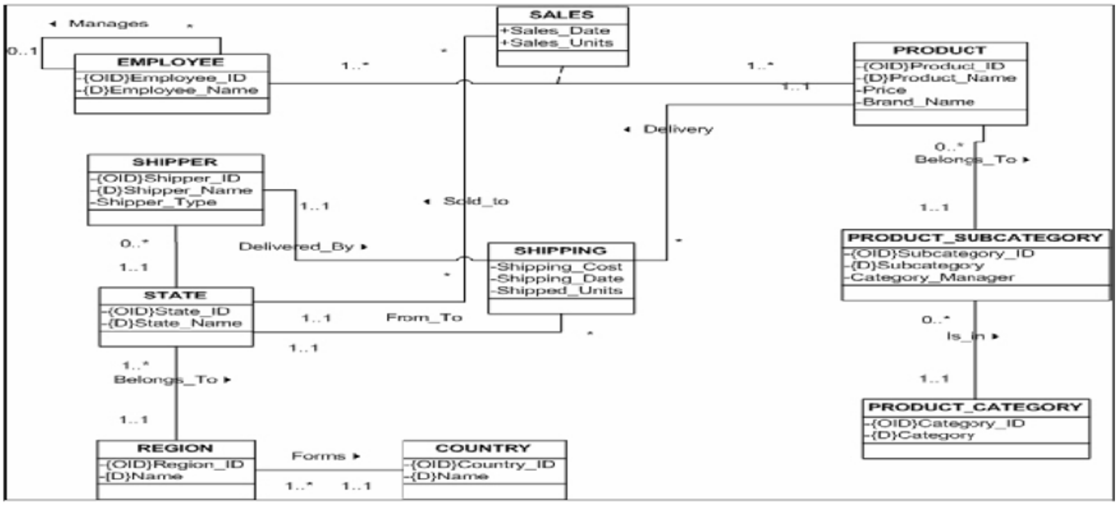

Figure 4.4: OOMD for Sales and Shipping System

(Mishra, Yazici & Başaran, 2008)

Dimensional Fact (DF) Model

'*The Dimensional Fact (DF) Model is a collection of tree structured fact schemas whose elements are facts, attributes, dimensions and hierarchies.*' (Mishra, Yazici & Başaran, 2008)

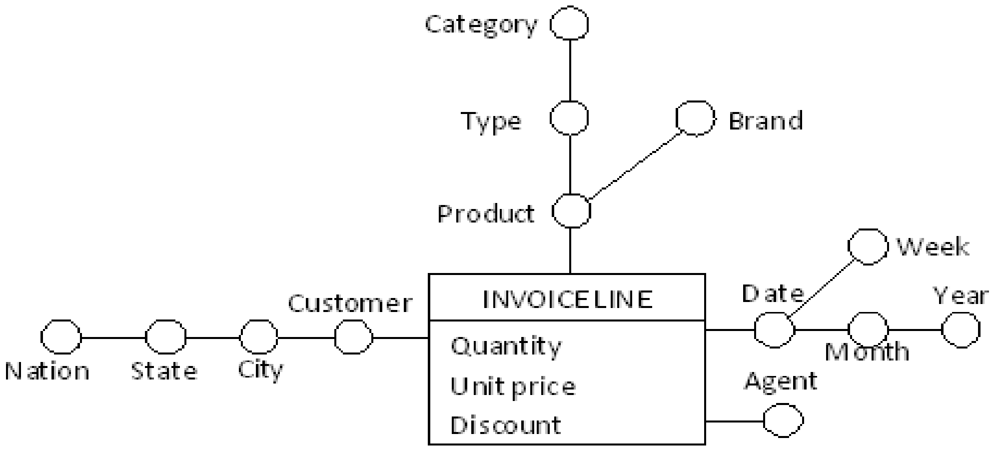

Figure 4.5: DF Model for the Invoice Fact

(http://en.wikipedia.org/wiki/Dimensional_Fact_Model)

Comparison of Conceptual Design Models

Property 1 - Additivity of measures

Property 2 - Many-to-many relationships with dimensions

Property 3 - Derived measures

Property 4 - Nonstrict and complete classification hierarchies

Property 5 - Categorization of dimensions - specialisation/ generalisation

Property 6 - Graphic notation and specifying user requirements

Property 7 - Case tool support

Property	DF	starER	ME/R	OOMD
1	✓	✓	✗	✓
2	✗	✓	✗	✓
3	✗	✗	✗	✓
4	✗	✓	✗	✓
5	✗	✓	✓	✓
6	✓	✓	✓	✓
7	WAND	✗	GRAMMI	GOLD, MS VISIO Rational Rose

Table 4.2: Comparison of Conceptual Design Models

(Mishra, Yazici & Başaran, 2008)

Data elements in the ASCTS Registry are updated every few years as medical science and clinical care methods change and improve. **Considering the above requirement of a Cardiac Surgery Unit, the most suitable Conceptual Data Warehousing Model is the *Object Oriented Multidimensional (OOMD) Model*.** OOMD approach satisfies most of the required. Also, it is more adaptable when user requirements are constantly changing. Using UML, OO model allows modelling of the business process, sub processes, use cases, system actors, classes, objects, collaboration of objects and relations.

<u>Logical Design Models</u>

‘*Data Warehouse logical design involves the definition of structures that enable an efficient access to information. The designer builds relational or multidimensional structures taking into account a conceptual schema representing the information requirements, the source databases, and non-functional requirements. There are many logical models like Flat schema, Terraced schema, Star schema, Starflake Schema, Star Cluster schema, Snowflake schema, Fact Constellation schema etc. Among them, star schema, snowflake schema and fact constellation schema are most commercially used models.*’ (Mishra, Yazici & Başaran, 2008)

Star Schema

It is the basic structure for a dimensional model. It has one fact table and a set of smaller dimension tables arranged around the fact table. The fact table is linked to all the dimension tables by one to many relationships. It contains measurements which may be aggregated in various ways. (Mishra, Yazici & Başaran, 2008)

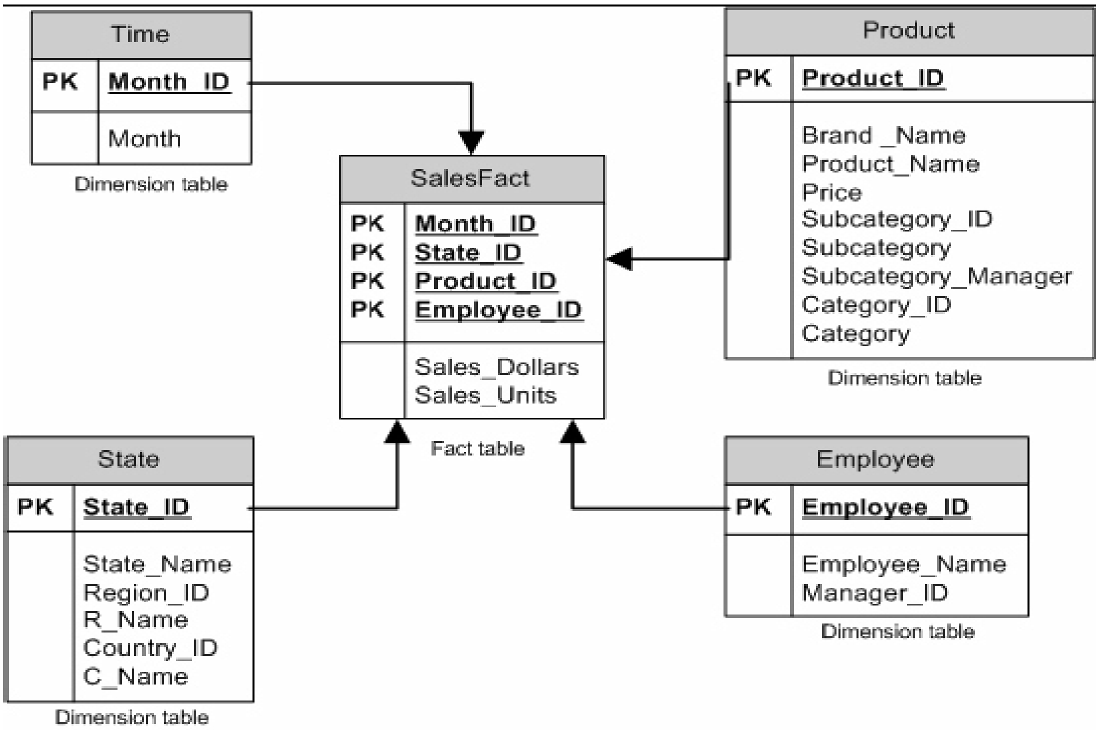

Figure 4.6: Star Schema for Sales Subsystem

(Mishra, Yazici & Başaran, 2008)

Fact Constellation Schema

A fact constellation schema consists of a set of star schemas with hierarchically linked fact tables. The links between the various fact tables provide the ability to “drill down” between levels of detail. (Mishra, Yazici & Başaran, 2008)

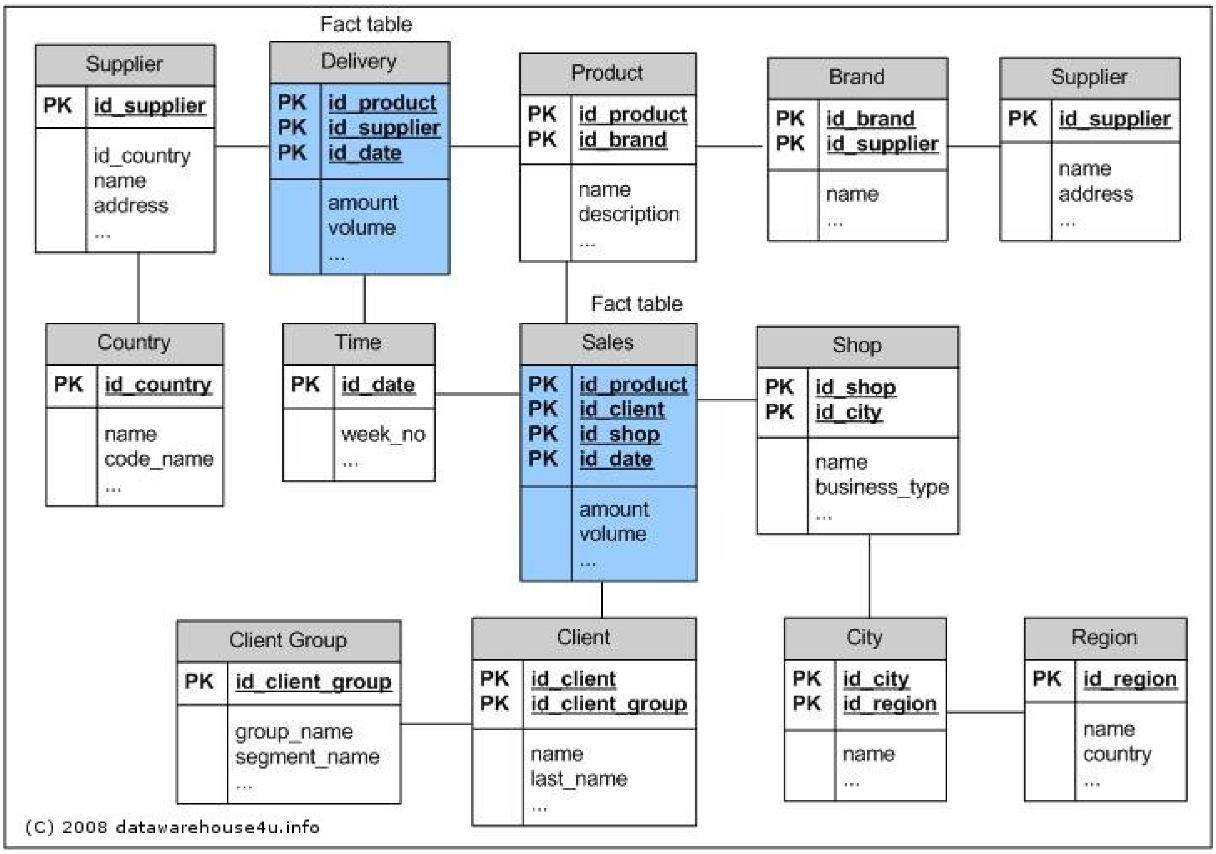

Figure 4.7: Fact Constellation Schema

(http://www.datawarehouse4u.info)

Snowflake Schema

A snowflake schema is a variant of star schema with all hierarchies explicitly shown and dimension tables do not contain denormalised data. (Mishra, Yazici & Başaran, 2008)

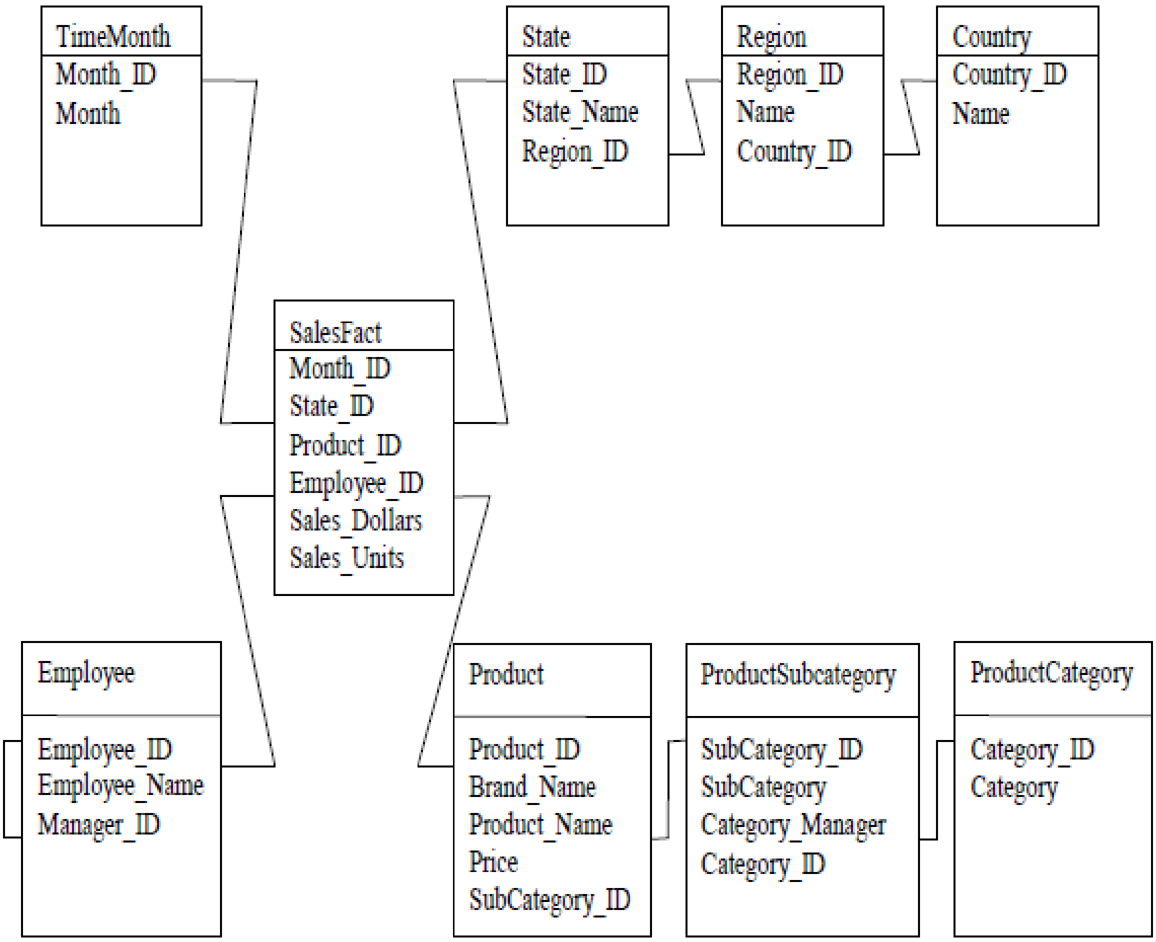

Figure 4.8: Snowflake Schema for the Sales Subsystem
(Mishra, Yazici & Başaran, 2008)

Comparison of Logical Design Models

	Star Schema	Fact Constellation Schema	Snowflake Schema
Efficiency	High	High	Moderate
Usability	High	Moderate	Moderate
Reusability	Low	Low	High
Flexibility	High	High	Moderate
Redundancy	High	High	Low
Complexity	Low	Moderate	Moderate

Table 4.3: Comparison of Logical Design Models (Mishra, Yazici & Başaran, 2008)

Efficiency is the most important factor in DW modelling because many queries access large amounts of data that may involve multiple join operations (Martyn, 2004). Cardiac surgery information systems require answering extremely complex clinical questions. **Considering the requirements of a Cardiac Surgery Unit, the most suitable Logical Data Warehousing Model is the *Star Schema*.**

A star schema is generally the most efficient design for two reasons. First, a design with denormalized tables need fewer joins. Second, most optimizers recognize star schemas and can generate efficient "star join" operations. The star schema is the simplest structure among the three schemas. In terms of flexibility, a star schema is more flexible in adapting to changes in user requirements. The star schema can adapt to changes in the user requirements easier, as all dimensions are equivalent in terms of providing access to the fact table. (Mishra, Yazici & Başaran, 2008)

4.6 DEVELOPING THE CARDIAC SURGERY DATA WAREHOUSING MODEL

Patient Demographics	Patient Risk Factors	Pre-Operative Cardiac Status	Previous Interventions	Haemodynamic Data
LNAME	SMO_H	MI	POP	HTM
FNAME	SMO_C	MI_T	PBYP	WKG
MNAME	FHCAD	MI_W	PBH	CATH
OTHER FIELDS	OTHER FIELDS	OTHER FIELDS	OTHER FIELDS	OTHER FIELDS

Operative Status/Category	Minimally Invasive	CPB and Support	Coronary Bypass	Valve Surgery
SURG	MIN	CPB	IDGCA	AOPROC
PROC	MIN_I	CPLEG	IMA	AOIM
STAT	OFFP	CCT	LIMA	AOIM_SR
OTHER FIELDS	OTHER FIELDS	OTHER FIELDS	OTHER FIELDS	OTHER FIELDS

Post-Operative Data	Post-Operative Data/ Complications	Mortality/Readmission	Automatic Data
RBC	RTT	Dischar	PatientID
NRBC	ROVD	MORTPD	AdmissionID
RBCUnit	ROBL	MORT_D	OperationID
OTHER FIELDS	OTHER FIELDS	OTHER FIELDS	OTHER FIELDS

Figure 4.9: Conceptual Design of the Initial Status of the Cardiac Surgery Data Sets

Figure 4. 10 show the initial status of the Cardiac Surgery Data Sets. They exist as independent data silos without any relationships to link them together. When a clinician wants to obtain a patient record he/she has to retrieve data from several independent silos and compile them to make the record. This process is highly time consuming and error prone. Also the patient record will not be available at the point of care.

The Logical Data Warehouse Design given below shows a data warehousing model which can be physically implemented to create a patient record when and where needed.

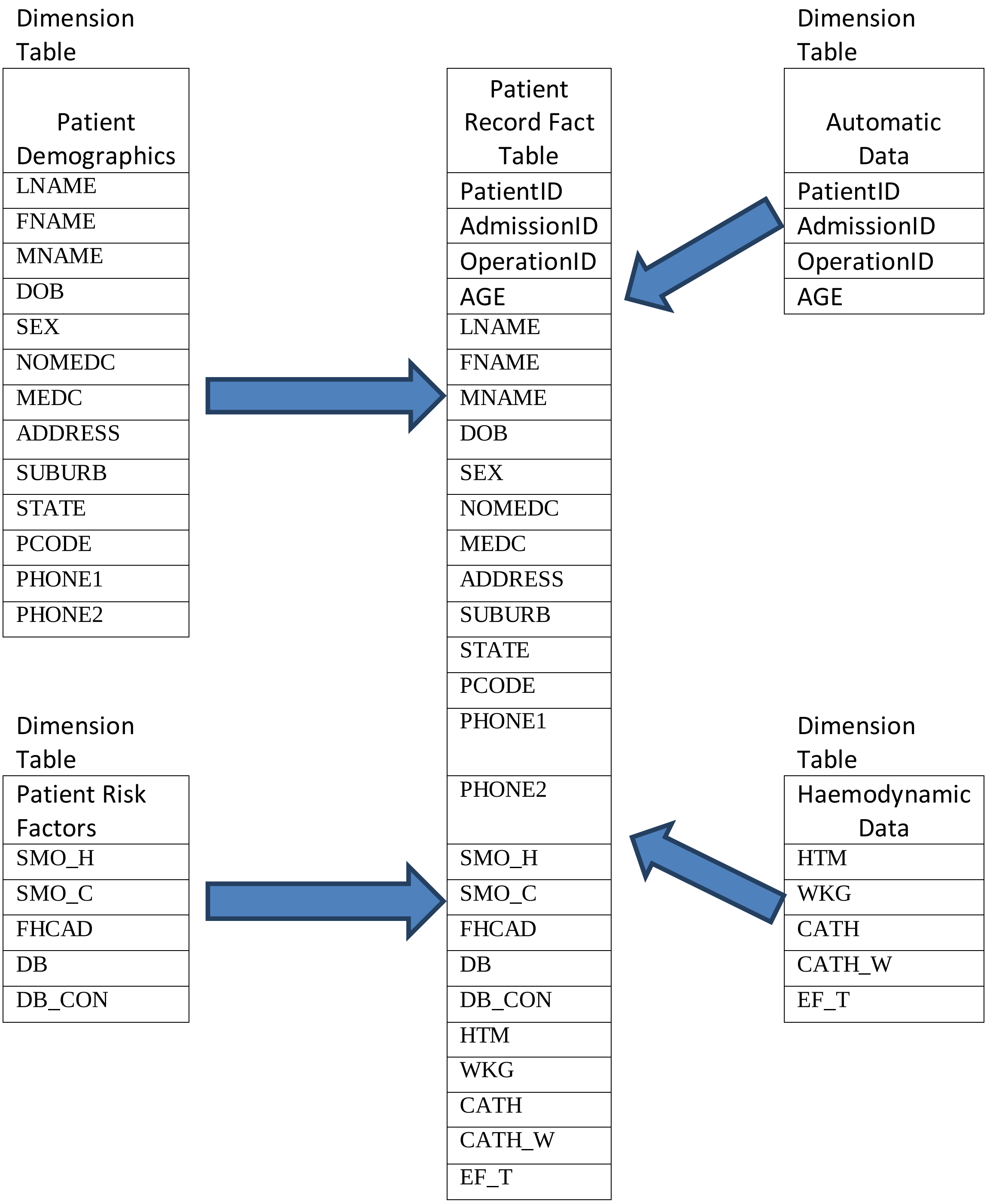

Figure 4.10 Patient Record Star Schema

Once the physical data warehouse structure has been created ETL tools must be used to populate the data. Following software can be used to build the data warehouse and extract, transform and load data to the data warehouse.

- SAS Data Integration Server.
- Oracle Data Warehouse Builder
- SQL Server 2008 R2
- IBM InfoSphere Warehouse

Chapter 5: Analysis

Data warehousing provides a suitable and practical approach in transforming fragmented operational healthcare data into useful and reliable information to support the decision making process and also provides the basis for performing data analysis tasks such as data mining and multidimensional analysis. Process of creating a data warehouse contains extraction of data from disparate and heterogeneous data sources, cleaning, filtering and transforming the extracted data into a common structure and storing data in a structure that is easily accessed and used for reporting and analysis purposes.

As the need for building a clinical data warehouse is clear, this research proposes a data warehousing model for integrating disparate and heterogeneous Cardiac Surgery data sets. The focus of this thesis is discussing the main challenges in medical informatics, identifying the causes for clinical data fragmentation and proposing a data warehousing model (conceptual and logical design models) to integrate fragmented data. The main reason for building a DWH is to improve the quality of information in the Cardiac Surgery unit. Data coming from both internal and external sources in various formats and structures is consolidated and integrated into a single repository. DWH system comprises the DWH and all components used for building, accessing and maintaining the DWH.

When a Doctor, Nurse, Patient and administrative staff of the hospital tries to obtain a report from a fragmented data set they have to face the three challenges listed below:

1. Need to obtain the service of programmers/analysts to locate, analyse and compile the data.
2. Problems arising due to data inconsistencies.
3. Having to go through every piece of data not just by name but by definition also.

Unless these processes are carried out properly, the report will end up creating yet another level of confusion and will lead to misinformed medical decision making.

Even though the report generation program should be simple to write, retrieving the data for the report is tedious due to lack of a proper data integration framework.

One major objective of this research was to identify the main barriers to healthcare data integration. Literature review conducted in order to achieve the objective resulted in discovering following reasons as the main obstacles for healthcare data integration.

1. Access to the medical information in proprietary software are decentralised in multiple information systems (Chenhui, Huilong, & Xudong, 2008).
2. Integration interfaces vary dramatically from system to system, especially for legacy information systems with proprietary interfaces (Chenhui, Huilong, & Xudong, 2008).
3. The number of system interfaces increases exponentially as the number of related systems increases. Thus, the complexity increases in an unmanageable way and the total integration cost will also increase dramatically (Chenhui, Huilong, & Xudong, 2008).
4. Heterogeneous Application Programming Interfaces (API) used by health care applications hinder the further advancement of communication and interoperation between different applications (Gong, Feng, & Lim, 2005).
5. Dissimilar information representations and meanings in diverse applications and their different syntaxes in heterogeneous data sources lead to semantic mismatch (Gong, Feng, & Lim, 2005).
6. Since most of integration methods focus on static weaving, they lack the flexibility needed to deal with changing requirements. As a result, new applications and related data sources with rapid change of data schemas are hard to be engaged in a running integration environment (Gong, Feng, & Lim, 2005).
7. The heterogeneity of data sources and that of applications are at all times mixed together in most integration practices. The heterogeneity of data sources increases the problem of communication and interoperation among distributed applications (Gong, Feng, & Lim, 2005).

8. Difficulty of adaptation of current users to a new system with an integration framework (Moner, et. al., 2006).
9. Unprecedented rate of increased cost in data integration (Giordano, 2011).
10. Absence of a clear, consistent, and effective approach to defining, designing, and building data integration components that lead to a more effective and cost-efficient data integration environment (Giordano, 2011).
11. Most of the data integration environments are poorly documented, with no repeatable method of understanding or clear ability to view the data integration processes or jobs. This leads to unnecessary reworks that result in massive redundancy in the number of data integration processes or jobs (Giordano, 2011).
12. Absence of managerial-level strategy and operational processes regarding the management and maintenance of corporate data assets (Giordano, 2011).
13. The complexity arising as a result of having to combine similar data from multiple and distinct source systems into one consistent and common data store (Giordano, 2011) for use by the healthcare practitioners and patients.

The clinical data warehousing model developed in this research is a highly customisable model. Star Schema presented can be customised to accommodate various data retrieval needs. Clinical Data Warehousing is a cost effective data integration methodology. In order to make use of the model presented in this research in a clinical environment, the conditions mentioned below should be satisfied.

1. There should be a document which clearly defines the structure of the data tables currently used by the Cardiac Surgery unit concerned.
2. Should clearly know what are the data retrieval operations going to be executed using the data warehouse.

5.1 LIMITATIONS OF THE STUDY

There are number of limitations to this study of the development of a clinical data warehouse model. Firstly, because of the time limitation and complexity of healthcare data and clinical decision-making processes the scope of the project is focused on discovering the main reasons for clinical data fragmentation and then developing a suitable data warehousing model.

Secondly, due to privacy issues, the proposed model was not implemented and tested in a clinical environment. However, this model can be easily implemented in a Cardiac Surgery clinical environment with the help of proper data warehousing and data integration tools.

Chapter 6: Conclusions

Successful healthcare data management is an important factor in developing support systems for the clinical decision-making process. Traditional database systems do not satisfy the requirements for critical data analysis tasks of the clinical decision-making users. An operational database supports daily data transactional operations. Operational databases contain detailed data but they do not include important historical patient data, and since they are usually highly normalized, they perform poorly for complex queries that need to join many relational tables or to aggregate large volumes of data in order to generate various clinical reports. This aggravates the data fragmentation.

A clinical data warehouse is a database that is different from the hospital's operational databases. It can be used for the analysis of consolidated historical data. According to Barry Devlin, IBM Consultant, a DWH is simply a single, complete and consistent store of data obtained from a variety of sources and made available to end users in a way they can understand and use it in a business context.

Implementing a Data Warehouse is a complex task containing two major phases. Firstly, in the configuration phase, a conceptual view of the warehouse is first specified according to user requirements (DWH design). Secondly, the related data sources and the Extraction-Load- Transform (ETL) process (data acquisition) are determined. Finally, decisions about persistent storage of the warehouse using database technology and the various ways data will be accessed during analysis are made. After the initial load (the first load of the DW according to the configuration), during the operation phase, warehouse data must be regularly refreshed, i.e., modifications of operational data since the last DW refreshment must be propagated into the warehouse such that data stored in the data warehouse reflect the state of the underlying operational systems. (Başaran, 2005). All the above steps must be carefully applied to fragmented healthcare data to guarantee the efficient and reliable retrieval at the point of care.

In this research Enterprise Data Warehousing Architecture was identified as the most suitable architecture to be used for the deployment of the Clinical data warehouse.

In this thesis, widely accepted conceptual and logical design approaches in DWH design are discussed. In the conceptual design phase DF, starER, ME/R and OOMD design models were compared. OOMD model was chosen because it is significantly better than the other design approaches. OOMD supports conceptual design phase with a rich set of diagrams that enables the designer model all the business information and requirements using a case tool with UML. OOMD design model meets the following factors while the others lack one or more:

- Additivity of measures
- Many-to-many relationships with dimensions
- Derived measures
- Non-strict and complete classification hierarchies
- Categorization of dimensions (specialization/generalization)
- Graphic notation
- Specifying user requirements
- Case tool support

In the logical design phase star, fact constellation and snowflake schemas were discussed. These three logical schemas are the mostly used models commercially. These three models are compared in terms of efficiency, usability, reusability and flexibility quality factor. Considering these factors and the information requirements of a cardiac surgery unit, star schema was chosen as the most suitable logical model for the purpose.

As a major objective of this research 13 main obstacles for healthcare data integration were identified through literature review and documented in Chapter 5.

6.1 CONTRIBUTIONS OF THE THESIS

1. Documenting the comparisons of data warehousing architectures, logical and conceptual data warehousing models.
2. Deriving and documenting Cardiac Surgery data table structures using data manual published by Australasian Society of Cardiac and Thoracic Surgeons (ASCTS) under ASCTS National Cardiac Surgery Database Program (ASCTS, 2008).
3. Proposing a data warehousing model suitable for a Cardiac Surgery Unit.
4. Identifying and documenting 13 main obstacles for healthcare data integration.
5. During the one year research period presenting 3 conference papers among which 2 were included in conference proceeding publications. (Three papers are given in the Appendix – A)

6.2 RECOMENDATIONS AND FUTURE DIRECTIONS

Following listed are the future works that can be carried out based on this research to add more practical value.

1. Populate the table structures given in Chapter 4 by writing a program to simulate the data based on the data types and ranges for the fields given in the data manual published by Australasian Society of Cardiac and Thoracic Surgeons (ASCTS) under ASCTS National Cardiac Surgery Database Program (ASCTS, 2008).
2. Implementing a Data Warehouse based on the model proposed in this research using Data Warehousing and Integration software.
3. Populate the Data Warehouse using Extract, Transform and Load tools.
4. Run query and report generation operations against the data warehouse.
5. Measure and document the performance of the data ware house.
6. Develop a Data Integration framework which can be customised and adjusted according to the changing needs of the Cardiac Surgery Unit and the future developments in the technology.

Bibliography

Australian Society of Cardiothoracic Surgeons (ASCTS) (2008). ASCTS Data Definitions Manual. Retrieved from: http://www.ccretherapeutics.org.au/assets/images/ascts-datadefinitionsmanual-v3-aug2009.pdf

Australian Government, Final Report of the National Health and Hospitals Reform Commission (2009, June). *A Healthier Future for All Australians*. Retrieved from: http://www.health.gov.au/internet/nhhrc/publishing.nsf/Content/nhhrc-report

Başaran, B.P (2005). A comparison of data warehouse design models.(Master's thesis,Atilim University, 2005). Retrived from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.126.6863&rep=rep1&type=pdf

Beale, T. (2007). openEHR a primer[Presentation Slides]. Retrieved from: http://www.openehr.org/shared-resources/getting_started/getting_started.html

Blobel, B. (2006). Advanced and secure architectural EHR approaches. *International Journal of Medical Informatics* , 75, 185—190. doi:10.1016/j.ijmedinf.2005.07.017

Baosheng, Y., Zhijun, R., & Jinsong, X. (2007, 6-8 July 2007). A Data Integration Approach to E-Healthcare System. Paper presented at the Bioinformatics and Biomedical Engineering, 2007. ICBBE 2007.

Budgen, D., Rigby, M., Brereton, P., & Turner, M. (2007). A Data Integration Broker for Healthcare Systems. Computer, 40(4), 34-41.

Chenhui, Z., Huilong, D., & Xudong, Lu. (2008, May). An integration approach of healthcare information system. 2008 International Conference on BioMedical Engineering and Informatics. doi:10.1109/BMEI.2008.109

Chen, C., & Wang, H. (2010, 25-26 Dec. 2010). A Framework for Medical Information Integration Based on Ontology and Web Services. Paper presented at the Information Engineering and Computer Science (ICIECS), 2010.

Chiu, D. K. W., Kwok, B. W. C., Wong, R. L. S., Cheung, S. C., Kafeza, E., & Kafeza, M. (2004, 5-8 Jan. 2004). Alerts for healthcare process and data integration. Paper presented at the System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference.

Davis, N. & LaCour, M. (2007). Health Information Technology (2nd Ed). St. Louis, Missouri,USA: Saunders Elsevier Inc.

Demetriades, J. E., Kolodner, R. M., & Christopherson, G. A. (2005). Person Centered Health Records. Towards HealthePeople. USA: Health Informatics Series, Springer.

Frank, L., & Andersen, S.K. (2010, July). Evaluation of different database designs for integration of heterogeneous distributed electronic health records. The 2010 IEEE/ICME International Conference on Complex Medical Engineering, 204 – 209. Retrieved from: http://ieeexplore.ieee.org/Xplore/

Giordano, A. D. (2011). Data integration blueprint and modeling: techniques for a scalable and sustainable architecture. Boston: IBM Press Pearson plc.

Goldstein, D., Groen, P. J., Ponkshe, S., & Wine, M. (2007). Medical informatics 20/20: quality and electronic health records through collaboration, open Solutions, and innovation. Sudbury, Massachusetts, USA: Jones and Bartlett Publishers, Inc.

Gong, P., Feng, D., & Lim, Y. S. (2005). An Intelligent Middleware for Dynamic Integration of Heterogeneous Health Care Applications. Paper presented at the Multimedia Modelling Conference, 2005. MMM 2005. Proceedings of the 11th International.

Handler, T., & Holtmeier, R., & Metzger, J., Overhage, M., Taylor, S., & Underwood, C. (2003). HIMSS Electronic Health Record Definitional Model Version 1.0. Retrieved from: http://www.himss.org/content/files/EHRAttributes.pdf

Health Informatics Society of Australia (2007). A Vision for an Australian Healthcare System Transformed by Health Informatics. Retrieved from: http://www.hisa.org.au

Hersh, W. (2004). Health Care Information Technology: Progress and Barriers. JAMA. 2004; 292(18):2273-2274.

Inmon, W. H. (2002). Building the Data Warehouse. New York: Wiley Computer Publishing.

Khoumbati, K., Themistocleous, M., & Irani, Z. (2005, 03-06 Jan. 2005). Integration Technology Adoption in Healthcare Organisations: A Case for Enterprise Application Integration. Paper presented at the System Sciences, 2005. HICSS '05. Proceedings of the 38th Annual Hawaii International Conference.

Kerkri, E. M., Quantin, C., Allaert, F. A., Cottin, V., Charve, Ph., Jouanot, F., & Y´etongnon, K. (2001). An approach for integrating heterogeneous information sources in a medical data warehouse. Journal of Medical Systems, 25(3), 167-176. doi: 10.1023/A:1010728915998

Kimball, R., Ross, M., Thornthwaite, W., Mundy, J., & Becker, B. (2008). The data warehouse lifecycle toolkit (2nd ed.). Wiley Publishing, Inc.

Knaup, P., Bott, O., Kohl, C., Lovis, C., & Garde, S. (2007). Electronic patient records: moving from islands and bridges towards electronic health records for continuity of care. IMIA Yearbook of Medical Informatics. Methods Inf Med 2007; 46 Suppl 1: 34-46. Retrieved from: http://www.ncbi.nlm.nih.gov/pubmed/17700902

Leslie, H. (2007). openEHR - The World's Record. PulseIT, pp50-55, Issue 6. Retrieved from: www.openehr.org/301-OE.pdf

MacDougall, C., Percival, J., & McGregor, C. (2009, September). Integrating health information technology into clinical guidelines. 31st Annual International Conference of the IEEE EMBS. Retrieved from: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5332653

Maldonado, J. A., Robles, M., & Crespo, P. (2003, 26-27 June 2003). Integration of distributed healthcare records: publishing legacy data as XML documents compliant with CEN/TC251 ENV13606. Paper presented at the Computer-Based Medical Systems, 2003. Proceedings. 16th IEEE Symposium.

Mann, L. (2005). From “silos” to seamless healthcare: bringing hospitals and GPs back together again. Medical Journal of Australia, 182 (1), 34-37. Retrieved from: http://www.mja.com.au/public/issues/182_01_030105/man10274_fm.html

Martyn, T. (2004). “Reconsidering Multi-Dimensional Schemas”, SIGMOD Record, Vol. 33, No. 1.

Masseroli, M., Bonacina, S., & Pinciroli, F. (2004, September). Java-based browsing, visualization and processing of heterogeneous medical data from remote repositories. Proceedings of the 26th Annual International Conference of the IEEE EMBS, 5:3326-9.
Retrieved from: http://www.ncbi.nlm.nih.gov/pubmed/17270994

Meystre, S. M., Savova, G. K., Kipper-Schuler, K. C., & Hurdle, J. F. (2008). Extracting information from textual documents in the electronic health record: a review of recent research. IMIA Yearbook of Medical Informatics. Methods Inf Med 2008; 47 Suppl 1:128-44. Retrieved from: http://www.ncbi.nlm.nih.gov/pubmed/18660887

Mishra, D., Yazici, A., & Başaran, B. P. (2008). A Casestudy of Data Models in Data Warehousing. Design, 314-319.

Moner, D., Maldonado, J. A., Bosca, D., Fernandez, J. T., Angulo, C., Crespo, P., . . . Robles, M. (2006). Archetype-Based Semantic Integration and Standardization of Clinical Data. Paper presented at the Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE.

National Health Information Management Principal Committee (2007). Strategic Work Plan 2007-08 to 2012- 13. Retrieved from: http://www.ahmac.gov.au/NHIMPC_Strategic_Work_Plan.pdf

National E-Health Transition Authority Ltd (2010, August). DRAFT Concept of Operations PCEHR System. Retrieved from: http://www.yourhealth.gov.au/internet/yourhealth/publishing.nsf/Content/pcehr

National E-Health Transition Authority Ltd (2010, August). NEHTA Blueprint. Retrieved from: http://www.nehta.gov.au

Payne P.R., Embi P.J., & Sen C.K.(2009).Translational informatics: enabling high-throughput research paradigms.Physiol Genomics. 2009 Nov 6; 39(3):131-40. doi: 10.1152/physiolgenomics.00050.2009

Pearce, C., & Haikerwal, M. C. (2010). E-health in Australia: time to plunge into the 21st century. Medical Journal of Australia, 193(7), 397–398. Retrieved from : http://www.mja.com.au/public/issues/193_07_041010/pea10510_fm.html

Pedersen, T. B., & Jensen, C. S. (1998). Research issues in clinical data warehousing. Proceedings of the 10th International Conference on Scientific and Statistical Database Management. doi: 10.1109/SSDM.1998.688110

Phipps, C. & Davis, K. (2002). Automating Data Warehouse Conceptual Schema Design and Evaluation, DMDW'02, Toronto, Canada, 23-32.

Sahama, T. R., & Croll, P. R. (2007, January). A data warehouse architecture for clinical data warehousing. First Australasian Workshop on Health Knowledge Management and Discovery. Retrieved from: http://crpit.com/confpapers/CRPITV68Sahama.pdf

Sandsmark, F. (2008). Best Medicine: Five Healthcare IT Challenges and Cures. CIO Digest, April 2008, 12-15.

SAS® Data Integration. (2009). Achieving faster ROI from your information assets. SAS Institute Inc. Retrieved from: www.sas.com/reg/wp/corp/3522

SAS® for Health Care. (2010). Software and expertise for meeting your organization's goals. SAS Institute Inc. Retrieved from: www.sas.com/resources/brochure/sas-for-healthcare.pdf

Sen, A., & Sinha, A. P. (2005). A comparison of data warehousing methodologies. Communications of the ACM, March 2005/48(3), 79-84. Retrieved from: http://cacm.acm.org/maga http://cacm.acm.org/magazines/2005/3/6272-a-comparison-of-data-warehousing-methodologies/abstractzines/2005/3/6272-a-comparison-of-data-warehousing-methodologies/abstract

Shepherd, M. (2007). Challenges in health informatics. Proceedings of the 40th Hawaii International Conference on System Sciences. doi: 10.1109/HICSS.2007.123

Szirbik, N.B., Pelletier, C., & Chaussalet, T. (2006). Six methodological steps to build medical data warehouses for research. International Journal of Medical Informatics, 75, 683–691. Retrieved from: http://www.intl.elsevierhealth.com/journals/ijmi

The New Data Integration Landscape. (2007). Moving beyond ad-hoc ETL to an enterprise data integration strategy. SAS Institute Inc. Retrieved from: www.sas.com/resources/whitepaper/wp_3498.pdf

Tryfona, N., Busborg, F., & Christiansen, J. G. B. (1999). starER : A Conceptual Model for Data Warehouse Design. Computer.

Woolf S.H. (2007). The Meaning of Translational Research and Why It Matters. JAMA. 2008;299(2):211-213. doi:10.1001/jama.2007.26.

Unlock the Value of Clinical Data. (2010). Bring repeatability and automation to the data integration process with SAS® Clinical Data Integration. SAS Institute Inc. Retrieved from: http://www.sas.com/industry/pharma/cdi/index.html

Yong-Gang, G., & Xin, C. (2010, 5-7 July 2010). Healthcare information integration and shared platform based on service-oriented architectures. Paper presented at the Signal Processing Systems (ICSPS), 2010 2nd International Conference.

Zenios, S. (2005). Information Technology in Health Care Systems: Barriers to Adoption. Proceedings of the International Multi-Conference on Computing in the Global Information Technology (ICCGI'06), IEEE Computer Society.

Zhao, C., Duan, H., & Lu, X. (2008, 27-30 May 2008). An Integration Approach of Healthcare Information System. Paper presented at the BioMedical Engineering and Informatics, 2008. BMEI 2008. International Conference on.

Zheng, R., Jin, H., Zhang, H., Liu, Y., & Chu, P. (2008, May). Heterogeneous medical data Share and integration on grid. 2008 International Conference on BioMedical Engineering and Informatics. doi: 10.1109/BMEI.2008.185

Appendices

APPENDIX A: PUBLISHED AND PRESENTED PAPERS

Presented and Published in the Proceedings of *International Conference on Innovation and Management (IAM)*, Kuala Lumpur, Malaysia. ISBN: 978-986-84899-2-9

Barriers to the Adoption of Health Information Technology

Saliya Nugawela[1] and Tony Sahama[2]

Faculty of Science and Technology, Queensland University of Technology, Australia
saliya.nugawela@connect.qut.edu.au[1]
t.sahama@qut.edu.au[2]
Corresponding Author: saliya.nugawela@connect.qut.edu.au

Abstract

Information Technology (IT) is successfully applied in a diverse range of fields. Though, the field of Medical Informatics is more than three decades old, it shows a very slow progress compared to many other fields in which the application of IT is growing rapidly. The spending on IT in health care is shooting up but the road to successful use of IT in health care has not been easy. This paper discusses about the barriers to the successful adoption of information technology in clinical environments and outlines the different approaches used by various countries and organisations to tackle the issues successfully. Investing financial and other resources to overcome the barriers for successful adoption of HIT is highly important to realise the dream of a future healthcare system with each customer having secure, private Electronic Health Record (EHR) that is available whenever and wherever needed, enabling the highest degree of coordinated medical care based on the latest medical knowledge and evidence. Arguably, the paper reviews barriers to HIT from organisations' alignment in respect to the leadership; with their stated values when accepting or willingness to consider the HIT as a determinant factor on their decision-making processes. However, the review concludes that there are many aspects of the organisational accountability and readiness to agree to the technology implementation.

Keywords: Medical Informatics, Health Informatics, Health Information Technology, Health Information Systems

1. Introduction

Medical Informatics (or Health Informatics) is the applied science which interconnects the disciplines of medicine, business, patient centered care, and information technology, in order to significantly and measurably improve both healthcare quality and cost effectiveness. Medical informatics tools, technology and tactics include not only computers and information systems, but also clinical guidelines, formal medical languages, standards, interoperability, and communication systems. (Goldstein, et al., 2007, p.8)

As with many applications of IT, if applied effectively, health information technology can dramatically improve the healthcare service experience of both clinicians and patients.

'*Informatics can help physicians better incorporate into clinical practice one of the most underused resources in medicine, the patient, whose help is greatly enhanced through this new technology*.' (Hersh, 2009) However, far too little attention has been paid to the efficient adoption of IT in health care.

Medical informatics should engage patients to be active partners in their medical care. This will help in reducing deadly medical errors and reduces unnecessary costs (Goldstein, et al., 2007, p.4). Medical practitioners and researchers can harness the power of IT to elevate the medical capabilities and knowledge to new heights. An appropriate and accurate application of IT into healthcare can create better systems to save lives, improve quality and reduce costs. According to Goldstein, et al., (2007) tangible and intangible benefits such as increased revenue, cost reductions, improved productivity, improved patient satisfaction, reduced length of stay, improved quality of care, improved medication safety, enhanced compliance efforts and Utilisation of community based, evidence based best practices can be achieved by adopting HIT in clinical environments.

Though IT has been successfully applied in many diverse fields such as education, defense, science, business, etc., it has not been productively used in healthcare to realise the true potential due to various impediments. Whilst limited body of literature suggests HIT is a positive step forward, there are many body of knowledge highlights that negative effect of the HIT in particular for healthcare arena.

'*Out-dated and non-existent health information technology systems lead to high costs, poor quality, non-patient centric care, an epidemic of medical errors, and insufficient disaster preparedness*.' (Goldstein, et al., 2007, p.xxix)

The rest of this paper is organised as follows: proceeding section discusses the barriers to the adoption of health information technology in clinical environments. Section 3 presents the different approaches used by various countries and organisations to tackle the issues successfully. Finally the conclusion is presented in Section 4.

2. Obstacles to Apply Health Information Technology in Clinical Environments

Although the case for adoption of improved health care informatics appears quite compelling, significant barriers to its use remain (Hersh, 2009). Proceeding subsections explains these barriers (Hersh, 2009; Sandsmark, 2008; Gans, et al., 2005; Zenios, 2005; Reynolds & Wyatt, 2011; HFMA, 2006; Goldstein, et al., 2007).

2.1 Initial and Maintenance Cost of Health Information Systems (HIS)

Costs of implementing HIS falls into categories such as acquiring software licenses, training and installation, procuring computer hardware, IT staffing, short-term reduction in productivity due to learning effects, short-term loss of revenue due to billing, and system upgrading (Zenios, 2005). Certainly the biggest impediment is financial since most of the small scale healthcare organisations are unable to bear the above mentioned costs without the help of a funding agency.

2.2 Concerns about Privacy and Confidentiality of Data
'*Many healthcare organizations extend beyond hospitals to neighbourhood clinics, home-health providers, and off-site services such as radiology interpretation and medical transcribers. In this dispersed environment, EMRs are always on the move, and the security of critical infrastructure—networks, PCs, servers, databases, becomes more difficult.*' (Sandsmark, 2008)
Special cultural environment in a healthcare organisation demands security to be a fine-tuned balance between technologies, human elements, standard practices and procedures (Sandsmark, 2008). In many instances the same data set of a patient is accessed by administrative staff, physicians, nurses and laboratory in order to make decisions regarding patient's healthcare. Therefore all those who are involved in the process have shared responsibility and accountability to maintain the security and integrity of a patient record.

2.3 System Interoperability
Various clinical and administrative systems within and beyond a healthcare organization must work together in a smooth manner to give optimum performance. But this does not happen since most of the proprietary software systems by various vendors do not communicate with each other effectively.
Most health care data, whether on paper or electronic format, are trapped in multiple data silos in multiple vendor products. As a result, a patient may have a physician or health system with an advanced EHR, yet if that patient requires care elsewhere, there is little likelihood the information from that advanced system will be accessible when it is needed(Hersh, 2009).

2.4 Fragmented Clinical Data Silos in Heterogeneous Sources
'*All medical data are located in different hospitals or different departments of single hospital. Every unit may use different hardware platforms, different operating systems, different information management systems or different network protocols. Medical data is also in various formats. There are not only a tremendous volume of imaging files (unstructured data), but also many medical information such as medical records, diagnosis reports and cases with different definitions and structures in information system (structured data).*' (Zheng, et al., 2008).
Health care practitioners, providers and patients often make decisions about medical treatments without complete understanding of the circumstances. The main reason for this is that medical data are available in fragmented, disparate and heterogeneous data silos. Without a centralised data warehouse structure to integrate these data silos, it is highly unlikely and impractical for the users to get all the information required on time to make a correct decision. (Shepherd, 2007)

2.5 Lack of a Well-Trained Medical Informatics Workforce to Lead the Process
To maintain an efficient Health Information System, an organisation must have a well trained workforce with a clear understanding of the requirements of both the worlds of medicine and IT. They should be highly motivated to carry out operations and make innovations to support rapidly changing requirements of healthcare industry. This type of a workforce building is a highly time consuming and costly effort.

2.6 Data Storage Requirements
The need for data storage in healthcare grows rapidly. Systems such as Picture Archival and Communications Systems (PACS), which handle digital X-ray, CT, and MRI images use significant storage. These systems are frequently used by most of the healthcare organisations today. Improving the storage environment means more than simply adding better storage hardware. Centralised, standardised storage-management software, which are independent of hardware and are able to manage the diverse, heterogeneous environments that exist in real-world data centers, are important ingredients in the ideal storage prescription. This also should include maximising utilisation of existing storage, improving backup and recovery performance, and classifying structured and unstructured data to improve archiving and retrieval. This last benefit is of particular importance to

clinicians, who need the right information at the right time. (Sandsmark, 2008).

2.7 Problems Related to the Implementation of Electronic Health Records (EHRs)

According to Gans, et al., (2005) the major impediments in implementation of EHRs are lack of support from practice physicians, lack of capital resources to invest in an EHR system, concern about physicians' ability to input data into the EHR, concern about loss of productivity during transition to EHR, inability to easily input historic medical record data into EHR, available EHR software does not meet the practice's needs, insufficient return on investment from EHR system, lack of support from practice clinical staff, practice staff does not have skills or training to use EHR, and security and privacy concerns.

2.8 Medical Practitioners' Resistance to Health Care Software

'*Healthcare is very complex. The paper based and manual processes that have evolved over the last 100 years will not change easily. The need to involve clinicians in transforming processes from paper to electronic media is the reason that often 50% of the expenditures for implementation are not related to technology but to involving clinicians, educating them, obtaining their input, and in supporting them learning by doing. Any technology tool will require humans to run that tool, so involvement, training, careful process mapping and redesign are critical. In the end, medical informatics deployment is more likely to fail due to human factors rather than a failure of technology itself.*' (Goldstein, et al., 2007, p.20)

Physicians' resistance to health care software might occur due to reasons such as the time cost of learning something new, fear of lawsuits, risk of data breaches, fear of automation and deprofessionalisation, and poor track record of existing HIS. (Reynolds & Wyatt, 2011)

3. Approaches to Tackle the Barriers

According to Protti & Johansen, (2010) and Protti & Bowden, (2010) approaches to tackle some of the above mentioned barriers to adaptation of health information technology can be discovered by analysing the success stories of two of the world's leading countries in the use of health care technology, namely Denmark and New Zealand. In Denmark, virtually all primary care physicians have electronic medical records with full clinical functionality, while New Zealand's use of information technology (IT) in health care is among the highest in the world compared with other developed nations.

The findings about the successful approaches as stated in Protti & Johansen, (2010) and Protti & Bowden, (2010) are:

- Peer influence and collegial pressure also played a significant part in the movement to adopt technology in Danish primary care.
- Giving financial incentives for physicians for adoption of Electronic Medical Record (EMR) systems. Including faster reimbursement and additional fees for patient–doctor e-mail consultations.
- Payments for technical support by the government.
- Regular visits by Data Consultants to healthcare practices to train physicians and staff, help practices improve data quality and implement standards, and encourage use of the full functionality of EMRs.
- Peer pressure through public monitoring of participation has been a helpful factor in encouraging EMR uptake in Denmark.
- Placing high priority on the engagement of clinicians in determining the precise content of the EMRs and in setting standards for data.
- In New Zealand the government provides general practices in several regions a one-time grant of approximately NZ$5,000 (US$3,600) to purchase computers.
- New Zealand Government made it a requirement to submit patient disease information to registers and to file fee-for-service claims electronically in order to receive subsidies, combined

with financial incentives for primary care.
• Most health IT investment has been provided by the New Zealand government via its central agencies.
• The central government of New Zealand has made core investments into standards development, privacy and security frameworks, infrastructure such as the national health index, and associated systems.

Use of Free and Open Source Software (FOSS) is growing among healthcare oraganisations worldwide as a solution to high initial cost of acquiring a health information system. Open solutions is the next major wave in healthcare information systems. Open solutions was identified as one of nine key healthcare technology trends for 2004 by Healthcare Informatics magazine (Goldstein, et al., 2007).

4. Conclusion

'*EHRs and information technology are essential to solving the problems in medical field only if they are intelligently applied, which require the active participation of leadership, clinicians, patients and all those involved in healthcare*.' (Goldstein, et al., 2007, p.xxx)

Benefits of adopting Health Information Technology in clinical practices largely outweigh the efforts required to overcome the barriers. Some of the above mentioned successful approaches implemented by the governments of Denmark and New Zealand may not be appropriate or affordable for large countries with huge population. Further research and experimentation have to be carried out in order to come up with suitable strategies for overcoming the barriers to adaptation of health information technology in different social, cultural and political environments. While all Australians to have and own 16 digits unique identifier for health related processes and services by 2012, it is worthwhile experiment to revisit the factors effecting adoption of HIT in general (ANHHRC, 2010; National E-Health Transition Authority, 2011).

References

Australian National Health and Hospital Reform Commission (2010). A healthier future for all Australians - Final Report June 2000 Retrieved from http://www.health.gov.au/internet/nhhrc/publishing.nsf/Content/nhhrc-report

Gans, D., Kraiewski, J., Hammons, T. and Dowd, B. (2005). *Medical Groups' Adoption Of Electronic Health Records And Information Systems*. Health Affairs, Vol 24: No 5.

Goldstein, D., Groen, P. J., Ponkshe, S., and Wine, M. (2007). *Medical Informatics 20/20: Quality and Electronic Health Records through Collaboration, Open Solutions, and Innovation*. Jones and Bartlett Publishers, Inc. Sudbury, Massachusetts, USA.

Healthcare Financial Management Association (HFMA). (2006). *Overcoming Barriers to Electronic Health Record Adoption*. Retrieved from http://www.providersedge.com/ehdocs/ehr_articles/Overcoming_Barriers_to_EHR_Adoption.pdf

Hersh, W. (2004). *Health Care Information Technology: Progress and Barriers*. *JAMA*. 2004; 292(18):2273-2274.

National E-Health Transition Authority. (2011, Febuary 16). *The National E-Health Transaction Authority Strategic Plan*. Retrieved from http://www.nehta.gov.au/about-us/strategy.

Protti, D. and Johansen, I. (2010). Widespread Adoption of Information Technology in Primary Care Physician Offices in Denmark: A Case Study. Issues in International Health Policy, Commonwealth Fund pub. 1379: Vol. 80

Protti, D. and Bowden, T. (2010). Electronic Medical Record Adoption in New Zealand Primary Care Physician Offices. Issues in International Health Policy, Commonwealth Fund pub. 1434 Vol. 96

Reynolds, C. J. and Wyatt, J. C. (2011). Open Source, Open Standards, and Health Care Information Systems. Journal of Medical Internet Research. Retrieved from http://www.jmir.org/2011/1/e24/

Sandsmark, F. (2008). *Best Medicine: Five Healthcare IT Challenges and Cures*. CIO Digest, April 2008, 12-15.
Shepherd, M. (2007). *Challenges in Health Informatics*. IEEE. Proceedings of the 40th Hawaii International Conference on System Sciences.
Zenios, S. (2005). *Information Technology in Health Care Systems: Barriers to Adoption*. Proceedings of the International Multi-Conference on Computing in the Global Information Technology (ICCGI'06), IEEE Computer Society.
Zheng, R., Jin, H., Zhang, Q., Liu, Y., & Chu, P. (2008). Heterogeneous Medical Data Share and Integration on Grid. IEEE. International Conference on BioMedical Engineering and Informatics.

Presented and Published in the Proceedings of the Conference *Statistical Concepts and Methods for the Modern World*. Colombo, Sri Lanka. ISBN: 978-955-0056-01-9

Internet Usage Trends in Medical Informatics

Saliya Nugawela and Tony Sahama

Computer Science Discipline, Faculty of Science and Technology
Queensland University of Technology (QUT)
Brisbane, Australia
Saliya.nugawela@connect.qut.edu.au
t.sahama@qut.edu.au

ABSTRACT

Technological growth in the 21st century is exponential. Simultaneously, developments of the associated risk, uncertainty and user acceptance are scattered. This required appropriate study to establish people accepting controversial technology (PACT). The Internet and services around it, such as World Wide Web, e-mail, instant messaging and social networking are increasingly becoming important in many aspects of our lives.

Information related to medical and personal health information sharing using the Internet is controversial and demand validity, usability and acceptance. Whilst literature suggest, Internet enhances patients and physicians' positive interactions, some studies establish opposite of such interaction in particular the associated risk. In recent years Internet has attracted considerable attention as a means to improve health and health care delivery. However, it is not clear how widespread the use of Internet for health care really is or what impact it has on health care utilisation. Estimated impact of Internet usage varies widely from the locations, locally and globally. As a result, an estimate (or predication) of Internet use and their effects in Medical Informatics related decision-making is impractical. This opens up research issues on validating and accepting Internet usage when designing and developing appropriate policy and processes activities for Medical Informatics, Health Informatics and/or e-Health related protocols.

Access and/or availability of data on Internet usage for Medical Informatics related activities are unfeasible. This paper presents a trend analysis of the growth of Internet usage in medical informatics related activities in USA. This study is an initiation of Internet usage in developing countries in particular healthcare services related indirect measures. In order to perform the analysis, data was extracted from surveys carried out by Pew Internet and American Life Project. Internet health information usage trends and their influence to the field of medical informatics is reviewed and discussed. The study clearly indicates a trend of people becoming active consumers of health information rather than passive recipients.

Keywords: Health Information, Internet Use for Healthcare, Trend Analysis

INTRODUCTION

Easy and widespread access to the Internet has transformed worldwide communication and delivery of all types of information including information related to healthcare (Ayantunde et al., 2007). The Internet is emerging as a means to propagate information about health and health care, improve communication, and facilitate various interactions between patients and the health care delivery process (Baker et al., 2003).

The Internet presents opportunities for combining great reach as a mass medium of communication with good effectiveness for supporting health behaviour change through computer adaptation, and possibilities for increasing availability of social support (Wangberg et al., 2009). In contrast, Internet may perhaps aggravate existing socio-economic differences in health (Korp, 2006; Wangberg et al., 2008), spread faulty health information (Sillence et al., 2007; Ipser et al., 2007), contribute to medicalization (Korp, 2006) and overwhelming responsibility for own illness (Pitts, 2004). Furthermore, Internet is a medium with unlimited possibilities, and that it is the users who put meaning into it through their use and their creation (Henwood et al., 2003). It is observed that, individuals are taking on greater responsibilities in managing their own health (Gianchandani, 2011).

The Internet is already an important source of providing health information, which will further increase in significance in the future. This study highlights the potential of Internet use for health related purposes. The aim of this research paper is to study the trends in the use and projected use of Internet for health purposes in United States of America. This includes more detailed analyses of various aspects related to the use of Internet for healthcare.

In USA, the use of Internet for health purposes has been monitored since 2000 by Pew Internet and American Life Project (Pew Internet, n.d.). This study looks closer at the trends in the USA population's use of the Internet for health purposes, and pursues five research questions.

1. On the basis of the present data, what can we predict about the future health-related use of the Internet?
2. How the Internet is used for health related activities by different age groups?
3. What kind of health-related Internet activities appear more important?
4. What are the trends in health related information search using Internet for people under various employment statuses?
5. What are the trends in health related information search using Internet by gender?

METHODS

The data collected by Pew Internet and American Life Project through survey sampling were used. The study was conducted through telephone interviews with a nationally representative sample of adults living in continental United States telephone households. The survey sampling was obtained during years 2004, 2006, 2008 and 2010 respectively.

Sample Design

'*The sample was designed to represent all continental U.S. telephone households. The telephone sample was provided by Survey Sampling International, LLC (SSI) according to PSRAI specifications. The sample was drawn using standard list-assisted random digit dialing (RDD) methodology. Active blocks of telephone numbers (area code + exchange + two-digit block number) that contained three or more residential directory listings were selected with probabilities in proportion to their share of listed telephone households; after selection two more digits were added randomly to complete the number. This method guarantees coverage of every assigned phone number regardless of whether that number is directory listed, purposely unlisted, or too new to be listed. After selection, the numbers were compared against business directories and matching numbers purged.*' (Topline et al., 2004)

Measures

There were 1124, 2928, 2253, 3001 participants for the surveys conducted for the years 2004, 2006, 2008 and 2010 respectively. All the survey questionnaires contained general and specific questions related to the use of Internet. We filtered out the data relevant to the questions about age, employment status and use of Internet for healthcare purposes.

All four surveys consisted of questions starting form general demographic questions, background questions on Internet usage in general to Internet usage for health related activities.

Response alternatives for Internet use for health related activities questions were, ‘Yes, have done this’, ‘No, have not done this’, ‘Don’t know’ and ‘Refused’. ‘Don’t know’ and ‘Refused’ alternatives were re-coded and added in to the category of ‘No, have not done this’. For a particular respondent, if at least one of the questions with respect to Internet usage on health related activities were answered with ‘Yes, have done this’ response alternative, the respondent was classified as an Internet health resource user.

Employment status of respondents was registered as ‘Employed full-time’, ‘Employed part-time’, ‘Retired’, ‘Not employed for pay’, ‘Disabled’, ‘Student’, ‘Other’ and ‘Refused’. ‘Retired’ and ‘Disabled’ response categories were re-coded in to one category ‘Retired or Disabled’ and , ‘Not employed for pay’, ‘Student’, ‘Other’ and ‘Refused’ response categories were re-coded as ‘Others including students and not employed’.

Respondent’s age was recorded directly, and hence grouped in to categories; ‘Less than 30’, ’31 to 45’, ‘46 to 60’and ‘Above 60’.

Analyses

The analysis was carried out in three main sections. First, testing for the differences in proportions of users with respect to age, gender and employment categories throughout four survey years. Second, testing for the differences in proportions of users from 2008 to 2010, based on a wide range of user activities, such as seeking information on a specific health related area, posting information on the Internet, signing up to receive information, consulting online rankings of facilities or health care professionals …etc. (Table 4). Third, projecting the proportion of users in the population in 2015. First and second tasks were carried out using the Chi-Squared test, while the multiple column proportion comparisons in the first task were carried out using Bonferroni adjusted z-test procedure in SPSS 19. The third task, projection was carried out using a logistic regression model.

RESULTS

Table 1: Age group classification of proportion using Internet for healthcare purposes (2004-2010)

		Age Group				Test values for overall difference in age groups	
		less than 30	31-45	46-60	above 60	Chi-Square	p-value
Proportion using the Internet for health care purposes	2010	73.4%$_a$	66.7%$_b$	60.0%$_c$	31.6%$_d$	304.898	0.000
	2008	72.8%$_a$	73.8%$_a$	69.6%$_a$	36.3%$_b$	258.17	0.000
	2006	67.6%$_a$	67.9%$_a$	58.6%$_b$	28.8%$_c$	368.929	0.000
	2004	53.0%$_a$	60.1%$_a$	55.2%$_a$	21.9%$_b$	93.318	0.000

Each subscript letter denotes a subset of age whose column proportions do not differ significantly from each other at the .05 level.

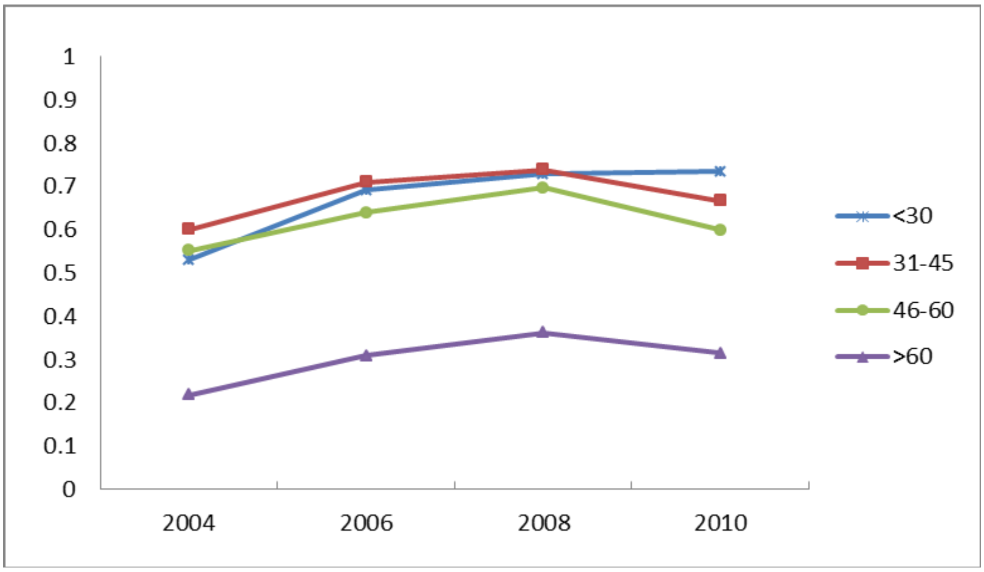

Figure 01: Proportion of people using the Internet for health purposes by age group

Age wise comparisons of users reveal that there has been a significant difference in proportions of users below 60 and above 60 up to year 2008. However, this difference in proportions is evident across all age groups in 2010. It is further evident that more than 70% of the youngest age group are Internet health resource users, while almost 70% of the oldest age group are non-users.

Table 2: Employment category classifications & use of Internet for healthcare purposes (2004-2010)

		Employment Category				Test values for overall difference in employment	
		Employed full-time	Employed part-time	Retired or Disabled	Others including students and not employed	Chi-Square	p-value
Proportion using the Internet for health care purposes	2010	71.8%$_a$	66.1%$_b$	36.0%$_c$	51.1%$_d$	277.042	0.000
	2008	74.5%$_a$	68.3%$_b$	36.7%$_c$	58.0%$_d$	237.932	0.000
	2006	67.8%$_a$	65.1%$_a$	30.3%$_b$	49.8%$_c$	325.223	0.000
	2004	58.1%$_a$	55.3%$_a$	21.1%$_b$	38.3%$_c$	88.379	0.000
Each subscript letter denotes a subset of employment whose column proportions do not differ significantly from each other at the .05 level.							

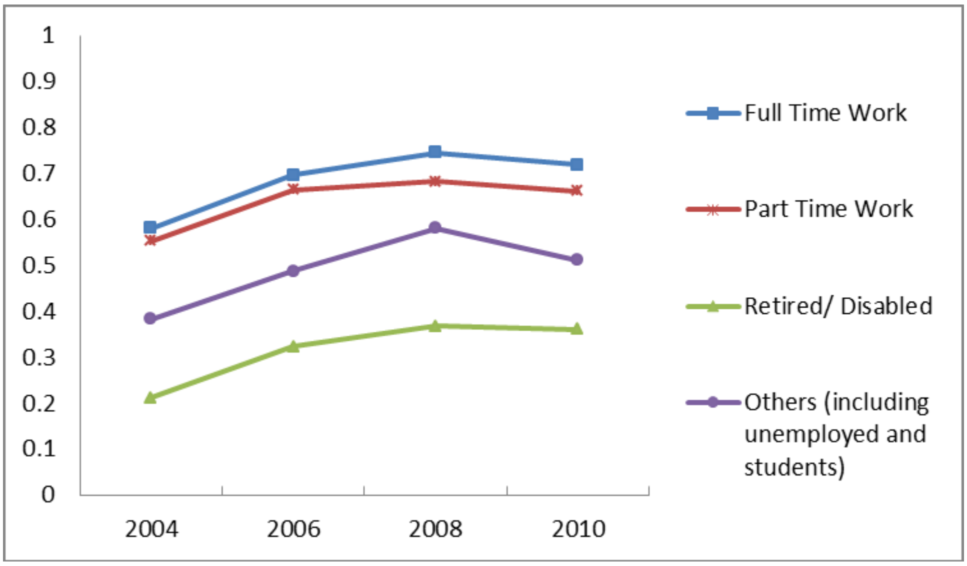

Figure 02: Proportion of people using the Internet for health purposes by employment category

Comparisons in employment categories reveal that up to 2006, there has not been a significant difference among proportions of users in fulltime and part time employees. But there has been a significant difference between proportions of employed users (full time and part time taken as one group), retired/ disabled users and other users (including students and not employed). After 2008, the difference in proportions is significant among all employment categories. In general, employed (full time and part time) persons are dominant users of Internet for health related resources.

Table 3: Gender wise classification of proportion using Internet for healthcare purposes (2004-2010)

		Gender		Test values for overall difference in gender	
		Male	Female	Chi-Square	p-value
Proportion using the Internet for health care purposes	2010	55.82%$_a$	57.83%$_a$	1.197	0.274
	2008	59.02%$_a$	61.95%$_a$	2.015	0.156
	2006	54.89%$_a$	54.12%$_a$	0.170	0.680
	2004	$_a$45.27%	$_a$46.57%	0.156	0.693

Each subscript letter denotes a subset of gender whose column proportions do not differ significantly from each other at the .05 level.

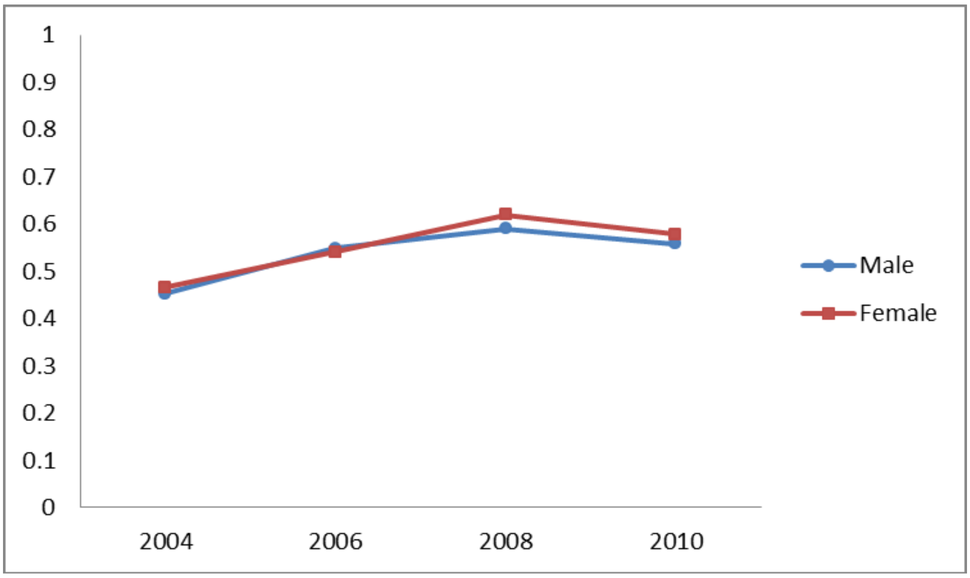

Figure 03: Proportion of people using the Internet for health purposes by gender

Gender group comparisons do not reveal any significant differences in proportions of male and female users. This implies that men and women are equally participating in the use of Internet for health purposes.

Second part of the study was focused on Internet health resource users in 2008 and 2010. The analysis was looking in to specific activities they were engaged using the Internet, with the objective of assessing the changes in proportions of users engaged in each activity from 2008 to 2010.

Other than activities listed in Table 4, there were considerable percentages of people engaged in activities; seeking information on a specific disease or medical problem, certain medical treatment or procedure, doctors or other health professionals and hospitals or other medical facilities recorded during 2008 and 2010. However, the changes in percentage of population seeking such information on the Internet, between years 2008 and 2010 were not statistically significant for these activities. All such activities can be categorised in to one broad category named 'Seeking health related information'. Evidence from Table 4 indicate that people are not merely seeking information on the Internet on diseases or treatments, but they are actively participating in discussion groups, posting information on social networking sites, consulting, posting reviews on health care practitioners on the web. In summary people are becoming active participants of Internet related health activities than being passive recipients of Internet health information.

Table 4: Changes in proportions of Internet health resource users from 2008 to 2010

Activity	2008			2010			Significance Level	
	%	95% CI		%	95% CI		Chi-Square	p-value
		Lower Bound	Upper Bound		Lower Bound	Upper Bound		
Seeking information related to health insurance, including private insurance, Medicare or Medicaid	31.74	29.27	34.22	40.62	38.29	42.95	25.7	0.000
Posted information related to health on a social networking site such as Facebook, MySpace or Linked In	4.91	3.77	6.06	31.15	28.96	33.35	1715.7	0.000
Posted information related to health on Twitter or another status update site	1.03	0.49	1.56	21.33	19.39	23.27	2611.3	0.000
Signed up to receive email updates or alerts about health or medical issues	18.26	16.21	20.31	19.05	17.19	20.91	5.9	0.054
Read someone else's commentary or experience about health or medical issues on an online news group, website or blog	39.22	36.63	41.81	40.62	38.29	42.95	5.0	0.080
Watched/ listened to an online video/audio about health or medical issues	11.66	9.95	13.36	31.15	28.96	33.35	171.6	0.000
Consulted online rankings or reviews of doctors or other providers	23.97	21.71	26.24	18.82	16.97	20.67	22.5	0.000
Consulted online rankings or reviews of hospitals or other medical facilities	22.80	20.57	25.03	17.42	15.62	19.21	23.0	0.000
Posted an online review of a doctor	4.77	3.63	5.90	5.32	4.26	6.38	10.5	0.005
Posted an online review of a hospital	4.11	3.05	5.16	3.62	2.74	4.51	11.9	0.003

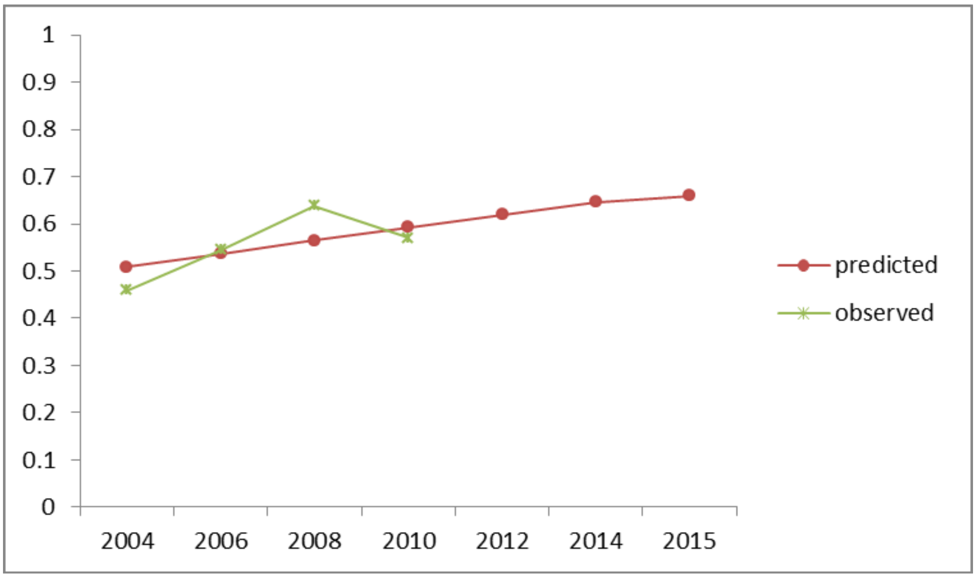

Figure 04: Logistic Regression of proportion of population using Internet for health purposes in American population

The logistic regression (Chi-square goodness of fit 30.751, p-value 0.000) results indicate that the proportion of Internet health resource users will increase but will gradually slow down. In 2010, around 57% of the population has used Internet for health purposes, while only 45% recorded in 2004. If the current trend continues, by 2015, the proportion of population engaged in Internet health related activities will increase and reach 66%.

DISCUSSION

If the observed increase in use of the Internet for health purposes in USA continues, we have estimated that by the year 2015, 66% of the American population will use the Internet for health purposes. Initially, the uptake is slow while the technology is new and is gradually increasing with the time. A similar pattern had been observed in a survey conducted in Norway by Wangberg et al., 2009.

There has been a significant increase (8.88 percentage points in proportion) of people who are seeking information related to health insurance, including private insurance, Medicare or Medicaid from 2008 to 2010. For the same period there was a strong increase of (26.24 percentage points) people who posted information related to health on a social networking site such as Facebook, MySpace or LinkedIn. Similar trend was observed in people who posted information related to health on Twitter or another similar status update site. There has been a substantial increase of (19.49 percentage points) people who watched/listened to an online video/audio about health or medical issues from 2008 to 2010. These increases reflect that video sharing, social networking and status updating activities are not merely activities of leisure, but Internet users have converted such resources on the net to productive health information dissemination.

There has been a significant decrease in activities such as consulted online rankings or reviews of doctors or other providers, consulted online rankings or reviews of hospitals or other medical facilities and posting an online review of a hospital for the period between 2008 to 2010.

CONCLUSIONS

Use of the Internet for health purposes continues to grow in the American population. Apparently, such mode of communications becomes an importan source of health information exchange. Moreover, the study shows a trend towards a more positive attitude of using the Internet for various healthcare related activities. This approach is a practicable realization if country specific data is available. The trend analysis performed, confirms that the potential of using the Internet for healthcare promotion is following similar pattern in comparison to Technological maturity around the globe. The Internet clearly is a key communication protocol with a potential to increase information propagation and possibly to improve health care delivery and outcomes. Professional, social and policy wise obstacles are inevitable however; "*the Internet has rapidly become indispensable to Medical Education. And, the Internet has transformed the patient-Physician relationship by empowering patients with information*" (Rajendran, 2001). Further research useful to identify the efforts that required to maximise the potential of this tool in particular Internet protocols, which could have great value for both patients and clinicians relationship. Such development may have greater impact on healthcare cost and service deliveries.

REFERENCES

Ayantunde, A. A., Welch, N. T. and Parsons, S. L. (2007). A survey of patient satisfaction and use of the internet for health information. International Journal of Clinical Practice, 61, 3, 458–462. doi: 10.1111/j.1742-1241.2006.01094.x

Baker, L., Wagner, T. H., Singer, S., and Bundorf, M. K. (2003). Use of the Internet and e-mail for health care information: results from a national survey. JAMA : the journal of the American Medical Association, 289(18), 2400-6. doi:10.1001/jama.289.18.2400

Gianchandani, E. P. (2011). Toward smarter health and well-being: an implicit role for networking and information technology. Journal of Information Technology, 26(2), 120-128. Nature Publishing Group. doi:10.1057/jit.2011.5

Henwood, F., Wyatt, S., Hart, A., and Smith, J. (2003). “Ignorance is bliss sometimes”: constraints on the emergence of the “informed patient” in the changing landscapes of health information. Sociology of health & illness, 25(6), 589-607. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12919447

Ipser, J., S. Dewing and Stein, D. (2007). A systematic review of the quality of information on the treatment of anxiety disorders on the internet. Current Psychiatry Reports 9(4): 303-309.

Korp, P.(2006). Health on the Internet: implications for health promotion. Health education research, 21(1), pp.78-86. Retrieved from: http://www.ncbi.nlm.nih.gov/pubmed/15994845

Pew Internet. (n.d.).Pew Internet and American Life Project. Retrieved May 10, 2011 from http://www.pewinternet.org/Data-Tools/Download-Data/Data-Sets.aspx?topicFilter=fa735cd6-5d6b-4ee6-b09e-a7330010e8a1

Pitts, V. (2004). Illness and Internet empowerment: writing and reading breast cancer in cyberspace. Health (London, England : 1997), 8(1), 33-59. doi:10.1177/1363459304038794

Rajendran, P. R. (2001). The Internet: Ushering in a New Era of Medicine. Journal of American Medical Association, 285(6), 804—809.

Sillence, E., Briggs, P., Harris, P., and Fishwick, L. (2007). Going online for health advice: Changes in usage and trust practices over the last five years. Interacting with Computers,19(3), 397-406. doi:10.1016/j.intcom.2006.10.002

Topline, F., Survey, P., and Project, A. L. (2004). November 2004 Activity Tracking Survey. Retrieved from : http://www.pewinternet.org/Reports/2005/Health-Information-Online/6-Methodology/November-2004-Activity-Tracking-Survey.aspx.

Wangberg, S., Andreassen, H., Kummervold, P., Wynn, R. and Sørensen, T .(2009). Use of the internet for health purposes: trends in Norway 2000-2010. Scandinavian journal of caring sciences, 23(4), pp.691-6. Retrieved from: http://www.ncbi.nlm.nih.gov/pubmed/19473317

Wangberg, S. C., Andreassen, H. K., Prokosch, H.-U., Santana, S. M. V., Sørensen, T., and Chronaki, C. E. (2008). Relations between Internet use, socio-economic status (SES), social support and subjective health. Health promotion international, 23(1), 70-7. doi:10.1093/heapro/dam039

West, D. M. and Miller, E.A. (2009). Digital Medicine: Health Care in the Internet Era. Washington, D.C.: Brookings Institution press.

Presented at *HIC 2011 Conference By Health Informatics Society of Australia (HISA)*. Brisbane, Australia.

Clinical Data Integration Approach Using SAS Clinical Data Integration Server (CDI) Tools

Saliya Nugawela,[1] Tony Sahama
Computer Science Discipline, Faculty of Science and Technology
Queensland University of Technology (QUT)
Brisbane, Australia

Abstract. The decisions people make about medical treatments have a great impact on their lives. Health care practitioners, providers and patients often make decisions about medical treatments without complete understanding of the circumstances. The main reason for this is that medical data are available in fragmented, disparate and heterogeneous data silos. Without a centralised data warehouse structure to integrate these data silos, it is highly unlikely and impractical for the users to get all the information required on time to make a correct decision. In this research paper, a clinical data integration approach using SAS Clinical Data Integration Server tools is presented.

Keywords. Clinical data integration, clinical data warehouse, SAS data integration.

Introduction

Informed decision making in health care is vital to provide timely, accurate and relevant advice to the right person, to reduce the cost of health care and to improve the overall quality of health care services. Since medical decisions are very complex, making choices about medical decision-making processes, procedures and treatments can be overwhelming [1].

One of the major Information Technology (IT) challenge in clinical practice is how to integrate several disparate, standalone information repositories into a single logical repository to create a single version of truth for all users [2]. A massive amount of health records, related documents and medical images created by clinical diagnostic equipments are generated daily [3]. Medical documents are owned by different hospitals, departments, doctors, technicians, nurses and patients. These valuable data are stored in various medical information systems such as HIS (Hospital Information System), RIS (Radiology Information System), PACS (Picture Archiving and Communications System) and etc., in various hospitals, departments and laboratories being primary locations [3]. These medical information systems are distributed and heterogeneous (utlising various software and hardware platforms including several configurations). Such processes and data flows have been reported by Zheng *et al*., (2008) [3].

"*All medical data are located in different hospitals or different departments of single hospital. Every unit may use different hardware platforms, different operating systems, different information management systems or different network protocols. Medical data is also in various formats. There are not only a tremendous volume of imaging files (unstructured data), but also many medical information such as medical records, diagnosis reports and cases with different definitions and structures in information system (structured data).* "[3].

This causes Clinical Data Stores (CDS) with isolated information islands across various hospitals, departments, laboratories and related administrative processes, which are time consuming and demanding reliable integration [4]. Data required to make informed medical decisions are trapped within fragmented, disparate and heterogeneous clinical and administrative systems that are not properly integrated or fully utilised. Ultimately, health care begins to suffer because medical practitioners and health care providers are unable to access and use this information to perform activities such as diagnostics, prognostics and treatment optimisation to improve patient care.

In this paper, the use of the latest data integration tools of one of the leading business analytics and business intelligence software to integrate disparate and heterogeneous medical resources are presented.

---

[1] Corresponding Author.

The rest of this paper is organised as follows: proceeding section introduces the related works on integrating data into a Clinical Data Warehouse (CDW). Section 2 presents use of SAS Enterprise Guide 4.3 to create a CDW. How the latest SAS Clinical Data Integration 2.2 and SAS Clinical Data Standards Toolkit 1.3 can be of a help in creating a data warehouse structure for medical resources are described in Section 3. Finally the conclusion is given in Section 4.

Integrating Data into a Clinical Data Warehouse (CDW)

"*Given its sensitive nature, diverse storage formats, and inherent privacy issues, healthcare data can benefit greatly from a data warehouse that integrates the data and ensures its correctness*." [5].

Data integration deals with incorporating all types of organisational data that are scattered throughout into a unified data warehouse. According to the SAS, New Data Integration Landscape [6], a comprehensive universal data integration solution should successfully complete following different programs or business initiatives:

- **Data cleansing and enrichment:** A data integration solution should have the capability to cleanse and enrich the data in order to ensure the completeness and accuracy of data.
- **Data warehousing/marts (ETL):** Should provide the capability to build and maintain data warehouses/data marts via the ETL (extract, transform and load) process.
- **Cross system data consistency (data synchronisation):** Should reflect changes made between systems across the enterprise.
- **Data migration/consolidation:** Should provide the capability to migrate data from multiple existing systems to one or more new or existing systems.
- **Master Data Management (MDM):** MDM is the practice of creating a single "perceived" truth by mapping multiple disparate definitions of items, such as names of patients and medicines, which are held in various systems.

According to SAS Solution Overview [7], SAS for health care, the main challenges in integrating clinical data into a Clinical Data Warehouse are as follows:

Siloed departments: Different departments within health care organisations use different clinical and administrative systems that do not communicate with each other.

Data overload: Health care organisations have huge volumes of patient and organisational data, but no way to make sense of it all.

Manual processes: The manual processes often involved with accessing disparate data sources are time-consuming and error-prone, taking up valuable time that could be better spent doing actual analysis.

Compliance mandates: The adoption of the electronic medical record (EMR) has been a slow transition, with data volumes growing exponentially at a time when compliance, regulations and increasing types of media-rich files are creating added pressure on organisations.

Building a CDW Using SAS Enterprise Guide 4.3

In order to demonstrate the use of SAS Enterprise Guide to create a CDW, a hypothetical data scenario for diabetes treatment was constructed. Figure 1 shows how the different clinical data silos related to diabetes treatment exist in various departments within the healthcare system. Table structures for the data scenario were adopted using the tables available on BioGrid Australia [8]. The tables were populated with hypothetical data and using sample data files of PASW Statistics software.

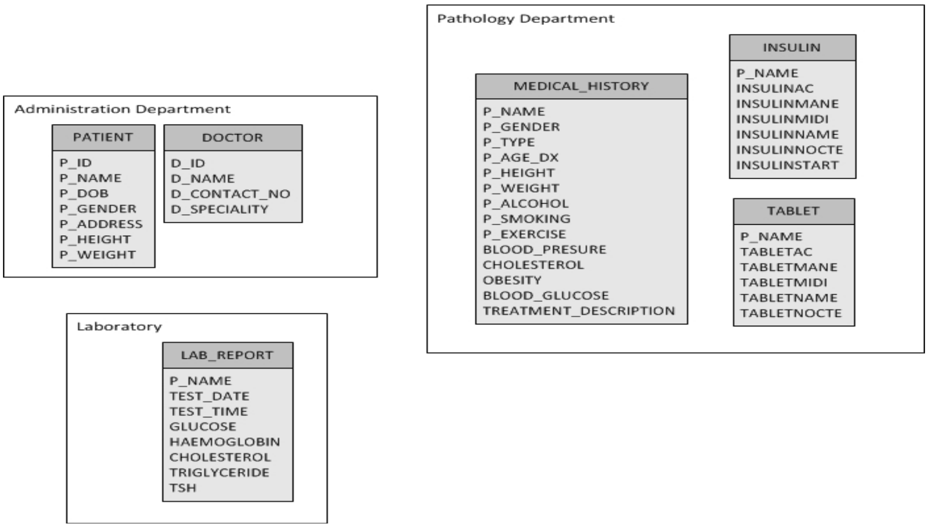

Figure 1. Clinical Data Silos Related to Diabetes Treatment (Source: BioGrid, 2010).

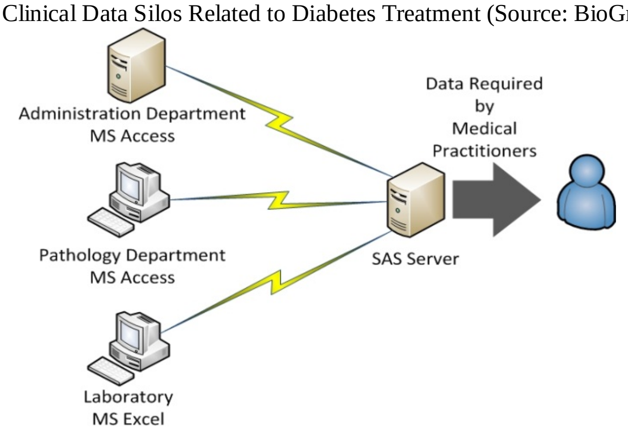

Figure 2. Integrating Fragmented Data Silos with SAS Server.

Each department database was implemented under separate Database Management System or a Spread Sheets application. Administration Department and Pathology Department data in separate Microsoft Access databases and Laboratory data in Microsoft Excel.

After experimenting with all the data warehousing architectures shown in Figure 3, we finally selected the distributed data warehouse architecture due to its high suitability to the chosen data scenario. The data from three data sources were extracted, transformed and loaded into SAS Server to create a distributed data warehouse structure.

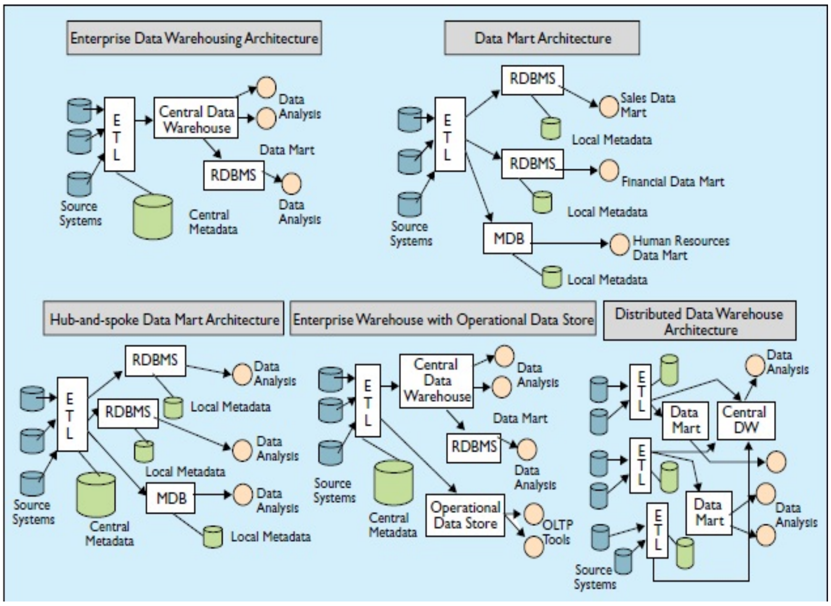

Figure 3. Different Types of DW Architectures (source: Sen & Sinha, 2005).

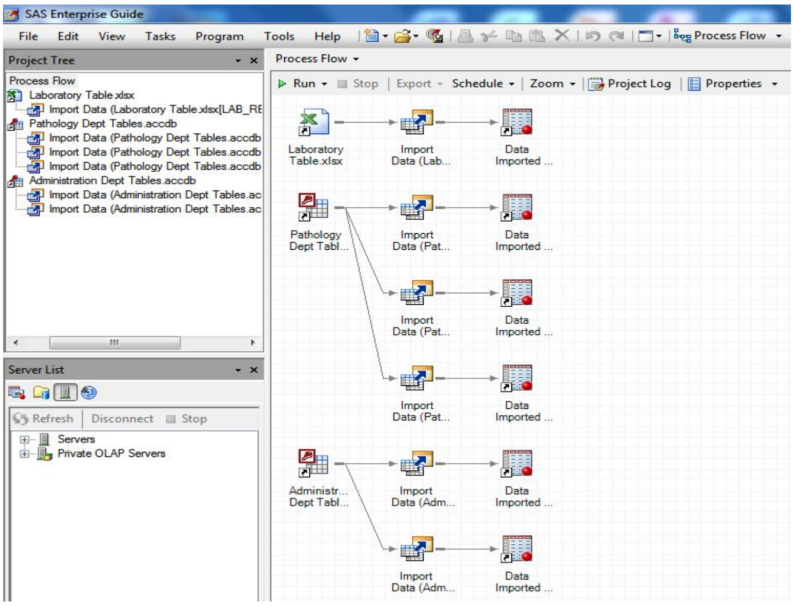

Figure 4. SAS Enterprise Guide Process Flow after ETL Process.

During the data cleansing and enrichment process additional data fields to establish relationships among tables and computed fields (e.g.: computed BMI using Weight and Height) were added. Figure 5 shows the logical ERD of the table structure in SAS server.

After constructing the CDW, SAS Enterprise Guide can be used to query data, apply data analytical functions, generate reports and share data over a network.

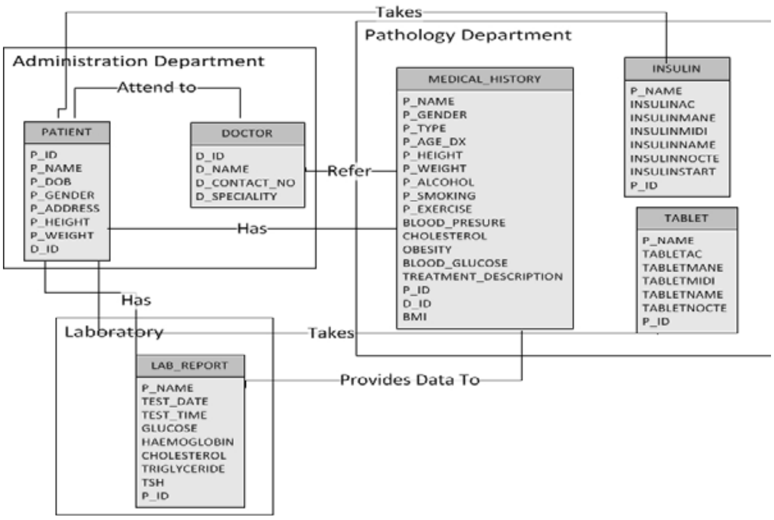

Figure 5. Logical ERD of the Diabetes Treatment Data Scenario.

Uses of SAS Clinical Data Integration when Building a CDW

SAS Clinical Data Integration mainly consists of SAS Clinical Data Integration 2.2 and SAS Clinical Data Standards Toolkit 1.3. SAS Clinical Data Integration is mainly used by clinical research professionals to improve efficiency, quality and speed in collecting, managing, analysing, reporting and assessing data from clinical trials.

SAS Clinical Data Integration implements CDISC (Clinical Data Interchange Standards Consortium) standards. This is helpful in data cleansing, enrichment process and to maintain the consistency of data. SAS can be used to gain both speed and efficiency by automating repeatable clinical data integration tasks and to deliver cleaner, more standard data for analysis.

Conclusion

Health care information is complex and must be accessed by health care practitioners, providers, researchers and patients. Some professionals involved in this processes are with minimal medical or information technology related training (or seldom no training). For all the health care information users, accessing medical data from fragmented, disparate and heterogeneous data silos is time consuming and expensive. Therefore medical data must be available in a centralised data warehouses equipped with proper tools and mechanisms to integrate data timely approach. As the amount of data and the number of systems involved increases rapidly, efficient and accurate data integration approach is required to create a single version of truth for all users.

References

Demetriades, J. E., Kolodner, R. M., & Christopherson, G. A. (2005). *Person Centered Health Records. Towards HealthePeople*. USA: Health Informatics Series, Springer.

Shepherd, M. (2007). *Challenges in Health Informatics*. IEEE. Proceedings of the 40th Hawaii International Conference on System Sciences.

Zheng, R., Jin, H., Zhang, Q., Liu, Y., & Chu, P. (2008). *Heterogeneous Medical Data Share and Integration on Grid*. IEEE. International Conference on BioMedical Engineering and Informatics.

Sahama, T. R., & Croll, P.R. (2007). *A Data Warehouse Architecture for Clinical Data Warehousing*. First Australasian Workshop on Health Knowledge Management and Discovery (pp. 227-232), Ballarat, Australia.

Berndt, D. J., Fisher, J.W., Hevner, A.R., & Studnicki, J. (2001). *Healthcare Data Warehousing and Quality Assurance*. IEEE.

SAS®: *The New Data Integration Landscape: Moving beyond ad-hoc ETL to an enterprise data integration strategy*, SAS Institute Inc., USA.

SAS®: *Solution Overview: SAS® for Health Care*, SAS Institute Inc., USA.

BioGrid Australia. Retrieved February 15, 2011, from http://www.biogrid.org.au/wps/portal

SAS Clinical Data Integration. Retrieved February 18, 2011, from http://www.sas.com/industry/pharma/cdi/index.html

SAS®: *Clinical Data Integration 2.2 User's Guide*, SAS Institute Inc., USA.

Sen, A. and Sinha, A. P. (2005): *A Comparison of Data warehousing Methodologies*, Communication of the ACM, 48(3), 79-84.

McDaniel, S., & Hemedinger, C. (2007). *SAS® For Dummies®*. Indianapolis, Indiana: Wiley Publishing, Inc.

Pedersen, T. B., & Jensen, C. S. (1998). *Research Issues in Clinical Data Warehousing*. Proceeding SSDBM '98 Proceedings of the 10th International Conference on Scientific and Statistical Database Management, IEEE Computer Society Washington, DC, USA.