Investigation of the ontology and information model of morbidity reporting in the electronic health record environment

Heather Llewelyn Grain

A thesis submitted for the degree of Masters of Health Informatics of Central Queensland University

March 2008

Declaration

Except where otherwise acknowledged in the text, this thesis represents my own original work.

Acknowledgements

The author wishes to express sincere appreciation to Professor Evelyn Hovenga and Dr Sebastian Garde for their continual support and assistance in the preparation of this manuscript. In addition, special thanks to my family for their untiring endurance of my pre-occupation. Thanks also to the members of the Australian health informatics community for their encouragement.

Abstract

The emergence of electronic health records in the Australian health environment is anticipated to improve the quality and availability of secondary data for health care service planning, public health and health research. Achievement of this goal requires in-depth knowledge about the business requirements and data collection format to enable secondary data collections to be automated whilst ensuring data access, accuracy and data integrity. The research objective was to understand the clinical admitted episode morbidity entity-relationship model in the context of ontological structures used to represent clinical data in EHRs. The applicability of the current national morbidity data collection format used to collect information on admitted patient episodes of care from Australian hospitals to collection automation and analysis is dependent upon the capacity of that data to be represented in formats that accurately transfer knowledge and retain meaning. The hypothesis that an ontological categorical structure exists within the data instructions and component definitions of the national morbidity data collection was tested by the adoption of ontology engineering. This methodology included 1) a review of existing data collection contents and formats, and 2) detailing the inclusion and exclusion instructions provided in ICD-10-AM, the classification system used to describe clinical information in the collection. This resulted in the identification of categories used for classification purposes and the relationships between these categories in the collection of these data from which an ontology of Australian morbidity data was developed. A significant outcome of this study is a clearer understanding of what types of knowledge are represented in these data collections and the relationships between the different knowledge components. It is recommended that this knowledge be used to inform the (EHR) system requirements suited to maximizing the efficiency of a future automated data collection process and optimise the usability of the data collected in this manner.

Table of Contents

Table of Figures

Table of Tables

Glossary

Admitted Episode	The period of admitted patient care between a formal or statistical *admission to hospital* and a formal or statistical *separation*, characterised by only one care type. Modified from{AIHW, 2006c}
Admitted Episode Morbidity System	The manual or computerised extraction of information from a patient's admitted episode for the purpose of morbidity reporting. This information comes both from the administrative data in the Patient Administration System and clinical information from the patient's health record. This information is classified and record according to the coding standards applicable at the time of the patient's discharge from the hospital.
Aggregate Terminology	A grouping of similar concepts, for particular purposes, using relationships that may e hierarchical and/ or uni or multi-dimensional.{Standards Australia, 2005}
Categorial Structure	Minimal set of domain constraints for representing concept systems in a subject field {ISO, 2007b}.
Classification	A classification is a system of categories to which morbid entities are assigned according to established criteria. Health classifications consist of hierarchical systems of codes for diseases, injuries and interventions as documented in health care services. (NCCH, 2006).
Clinical Coder	The person responsible for manual data extraction and for quality control of any automated data extraction for the purpose of morbidity reporting.
Clinical Morbidity Data	That part of the morbidity data that identifies disease, treatment and injury information and does not include the data collected about the demographics of the patient, the service provision organisation or the episode of care.
Clinical Terminology	The component of health language used at the point of care for the purpose of clinical management of the subject(s) of care.{Standards Australia, 2005 }
Clinician	The professional who provides direct care to the patient. In a hospital this might be a doctor, surgeon, anaesthetist, nurse, or allied health professional.
Coding	Coding is the translation of clinical data such as diseases, injuries and interventions from a patient record into an agreed coded format. Currently in Australia, diagnoses and procedures are assigned a series of numerical and/or alphanumerical codes using ICD-10-AM. This allows the comparison, analysis and interpretation of collected morbidity data. (NCCH, 2006)
Coding Standards	These standards indicate which clinical information from an episode of care should be included in the National morbidity data. These standards also provide some guidance on the structure of the data to be submitted. Each standard is uniquely numbered and relates to a specific clinical circumstance. National standards for the classification of morbidity data are reviewed every two

	years {Commonwealth of Australia, 2004}.
Australian Refined Diagnosis Related Group (AR-DRG)	A patient classification scheme which provides a means of relating the number and types of patients treated in a hospital to the resources required by the hospital, as represented by a code. (AIHW, 2006c)
Domain Constraint	Rule prescribing the set of sanctioned characteristics that are valid to specialise a concept representation in a certain subject field.
Electronic Health Record (EHR)	A repository of information regarding the health of a subject of care, in computer processable form (CEN, 2005)
Entity	In a database, anything about which information can be stored; for example, a person, concept, physical object or event. Typically refers to a record structure.{The Computer Language Company, 2007}
Admitted patient	Person who is the subject of an admitted episode.
Integrated Care Electronic Health Record (ICEHR)	A repository of information regarding the health of a subject of care in computer processable form, stored and transmitted securely, and accessible by multiple authorised users. The ICEHR has a standardised information model which is independent of EHR systems. Its primary purpose is the support of continuing, efficient and quality integrated healthcare and it contains information which is retrospective, concurrent and prospective. (CEN, 2005)
Medical Record	A written or electronic account of a patient's medical history, current illness, diagnosis, details of treatments, chronological progress notes, and discharge recommendations. This is a legal document the patient is entitled to read and is authenticated by the physician's signature. There are strict rules regarding the confidentiality of the medical record (VPA Web, 2006).
Metadata	Data that defines and describes other data.{ISO, 2004 }
METeOR – Value Domain	METeOR, the National Health and Welfare Metadata Registry describes a value domain as "*a set of permissible values by which a data element can be implemented*" {AIHW, 2005b}.
Model	Pictorial representation of concepts and the relationships between them.
Morbidity Data	Data that represents the incidence of disease in a society. This data is collected around individual people or instances of care and since 1958 in Australia has included: • demographic information about the person, • care provision information (e.g.: source of admission to hospital, place to which the person is discharged after care, or specialty or ward of service) • clinical information about the disease, health status, treatment and injuries of the patient – recorded using the International Classification of Disease. Modified from {AIHW, 2006d}
Morbidity Data Collection	Structured data collections identified by each State or National body. In Australia the state bodies are required to submit data to the National collection by the health care agreements. National Health Information Agreement and the Australian Health Care

	Agreement in 1958. The collection includes data on the causes, effects and nature of illness among Australians and the utilisation of health services {Health Data Standards and Systems Unit, 2003; AIHW, 2006b}. For the purpose of this research morbidity data collection means the data collected from hospitals on acute episodes of care.
Morbidity Data Extraction	The structure and rules established to extract data for inclusion in morbidity data collections from the medical record, the electronic health record and the hospital administration computer system/s.
Ontology	"*Ontology is a branch of metaphysics concerned with the nature and reality of a thing*" {Merriam-Webster, 2007}. In information science and ontology shows concepts within a domain and the relationships between those concepts Ontologies have been developed as a tool in information systems to identify the concepts of relevance of the domain, the semantic modelling of and relationships between these concepts{Gruber, 1993; Noy, 2001; George, 2005}
Patient Administration System	The systems used to administer patient services, in this case in a hospital environment. These systems include: Patient Master Index (PMI) – the system that holds details to identify individual patients and the information common to that patient in all sectors of the health care system. This includes information such as name, date of birth, address, country of birth. Admission, Discharge, Transfer (ADT) systems. These systems are used to coordinate and monitor admitted patient services in the hospital; they record the ward in which the patient resides, the doctor responsible for the patient, patient billing information and enquiry facilities.
(Semantic) Relationship	An association indicates that one entity works together with one or more other entities to create meaning that goes beyond the component entities. The association relationship encapsulates this chunk of meaning that extends beyond the boundaries of the individual entities. In object-technology, associations are usually thought of as the most generic kind of relationship. In psychology and in semantic model generally, semantic associations are often treated as a specialized kind of relationship on a peer level with other kinds of relationships such as a whole-part.{anon, 2003}
Sanctioned characteristic	"*Formal representation of a type of character*"{ISO, 2007b}. Example: "*CauseOfInflammation canset {bacteria, virus, parasite, autoimmune, chemical, physical where 'canBe'" is the semantic link*."{ISO, 2007b}
Semantic link	Formal representation of a directed associative relation or partitive relation between two concepts.{ISO, 2007b} Example: hasLocation (with inverse isLocationOf): is CauseOF (with inverse hasCause) A direct associative relationship is one where the association is complete, while a partitive relation is one where the relationship is

	not definitive but may be useful to partly define the concept required.
Terminology	In health informatics terminology usually refers to computable representations of language including classifications, and clinical terminologies. (Modified from {Standards Australia, 2005}).

Chapter 1 Introduction

Australian governments are implementing significant information system changes to information systems in hospitals as part of a National approach to improving information availability and quality for clinical decision support and communication. A by-product of this approach is the opportunity to obtain improved morbidity data to support health care planning and public health information.

Davidson in his review of the use of clinical information system identified that the value of these systems and the data in them are not being maximized {Davidson, 2007}. Though the data collected is used to monitor simple public health trends and to support financial reporting, the amount of detail information collected, including the many co-morbidities and procedural information are rarely used to support the development of knowledge either of the health of the community or of health care processes.

Data are described using a range of tools and terms. This research focuses on the clinical entities which are defined as 'anything about which information can be stored..'{The Computer Language Company, 2007}, in this case diagnoses and procedures related to people who have been admitted as patients in a hospital. Entities can be represented at different levels of detail, the less detailed, the more the data are grouped to represent common ideas in the data (categories).

Entities can be described using definitions and represented using codes or values. They also have collection instructions to guide the user in assigning values to the entity. When conveying meaning in language people use different semantic elements to express meaning. In the entity relationship an entity can be equated to a noun (a thing), while the relationships are used to express the way different entities interact, and thereby provide a greater specificity and clarity of knowledge. For example, the entity of John may have a relationship to Mary of "IsBrotherOf". This approach to the representation of data is based upon an understanding of the ontology (nature of reality) of morbidity data. This research centres on the ontology of one element of health information that used for morbidity reporting which is an abstraction of the total patient record. This can be expected to simplify the complexity required to represent all relationships and meaning in the full patient record.

This research investigates and identifies the entities and relationships between these entities within the field of clinical morbidity data reporting for admitted patient services in Australian healthcare. The reason for taking an ontological approach to the study of morbidity data is based upon the need to understand clearly what types of knowledge are represented in these data collections and the relationships between the entities in order to understand the system requirements suited to maximizing the usability of the data collected taking advantage of today's technology{Gruber, 1993; Chen, 1976; Endres, 2006; Abbott, 2001; Campbell, 2004; Collins, 2002; George, 2005; Safran, 2007}.

1.1 Morbidity Reporting

The Oxford Dictionary defines morbidity as:

> "*1 having* ***o****r showing an unhealthy interest in unpleasant subjects, especially death and disease.*
> *2 Medicine of the nature of or indicative of disease*{Weiner, 2006}".

The second definition is the one to which this research applies, this can also be considered as the relevant incidence of disease {Merriam-Webster, 2007; National Cancer Institute, 2007}.

The Regional Director of the Western Pacific for the World Health Organisation (WHO) indicated the importance of morbidity data as the source of information to identify health trends and issues in communities in the Western Pacific since the second world war {Report of the Regional Director to the Regional Committee for the Western Pacific, 1998}.

The collection of information about people treated as admitted patients in hospitals began in the late 1800's{World Health Organization} in the United Kingdom where registers were kept that identified people admitted to the hospital, their diagnosis and whether their treatment resulted in them leaving the institution alive or dead. The United Kingdom legislated the reporting for selected diseases in 1899{Gostin, 1998 }.

The data were often held in books or registers requiring users to access, read and count cases in order to report upon activity or undertake research. The data included were limited, and difficult to access, but it was also a time when the processes of evidence based medicine and health service planning were emerging and therefore the demands made upon the data were not great.

Collection of National hospital admitted episode morbidity data was instituted through the Australian Health Care Agreement in 1958, though information on communicable diseases had begun much earlier than this. These data were reported using paper forms for each episode of care. The forms were collected and forwarded directly to the Australian Bureau of Statistics (ABS).

Admitted episode morbidity data are collected for patients who have completed an episode of care for which they were formally admitted (an admitted episode). Admitted episode clinical morbidity data are the data collected for every person discharged from hospital. This information includes details of diseases, external causes of injury and procedures performed during the stay{Health Data Standards and Systems Unit, 2003; AIHW, undated}. Huffman, the original text book source for health information management, as it emerged as a profession in the 1960's and 70's stated that:

> "*Health care institutions need to be able to study patterns of illness and injuries treated for clinical, financial and administrative purposes. Comparing health care data between individual facilities within a defined area or country, or even among countries is vital to the growth of medical information around the world*" {Huffman,1994} p321.

By the 1970's local hospital registers were being coded using internationally recognised classification systems. For example: In 1973 Australian hospitals were using hospital International Classification of Disease Adaptation (HICDA) to collect local data. In the early 1970's Prince Henry's Hospital in Melbourne was submitting data to the electronic morbidity data collection system operated by the Commission on Professional and Hospital Activities in the USA. Early computerised systems for National reporting of morbidity data in Australia were introduced in 1987 through the requirement that public hospitals submit morbidity to the Australian Institute of Health and Welfare (AIHW){AIHW,2005a}.

Figure 1.1 represents the people and organisations (green lozenges), instructions and business rules (red lozenges), information systems (black lozenges) and data collections (blue lozenges) of Australia's system of morbidity data collection and users. The figure represents the relationships between these system components. Though the focus of this research is on the content of the actual data collection, rather than the system used to collect that data, the complexity of our system underpins the need for a solid and sound data structure. This diagram also highlights the points at which instructions are used and the actual different data collections. Appendix A outlines the different organisations involved in the collection of morbidity data.

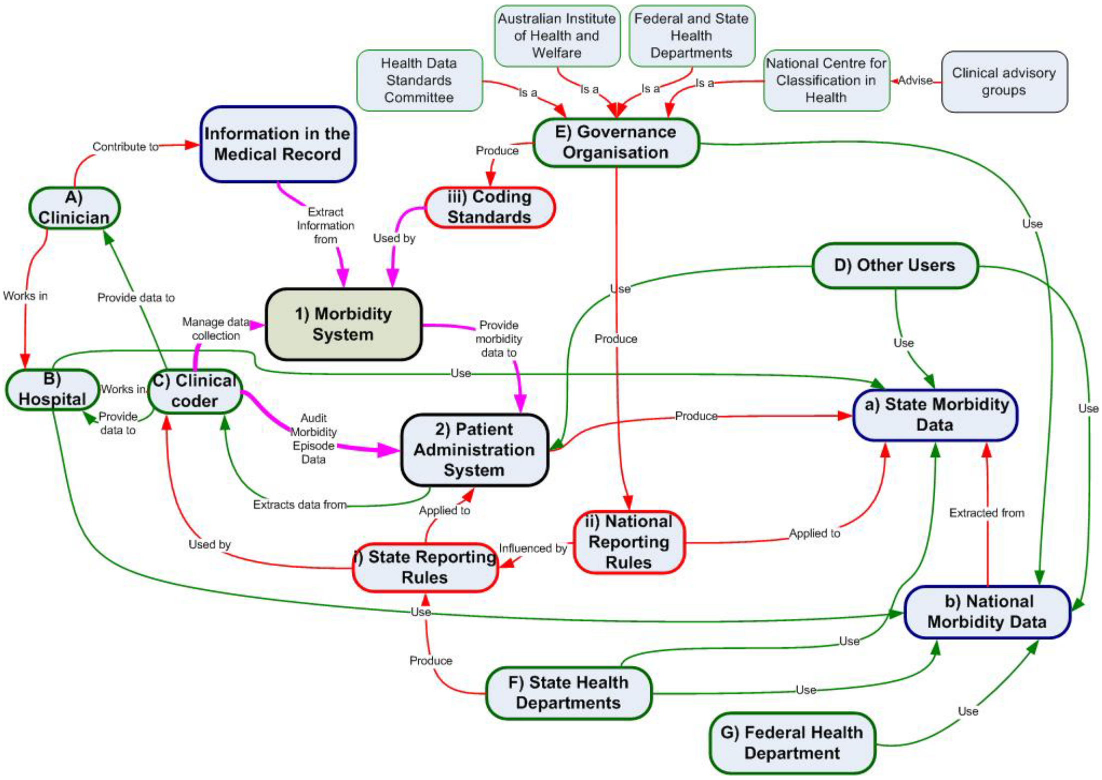

Figure 1-1 The morbidity system today

This research was limited to the study of the structures (entities) of the clinical information and of the relationships between these entities in the data collections. These data are coded using a classification system, the value domain of the entities in the collection. During the period of this research, 2005 – 2006 fiscal year, the morbidity data in Australia were collected using The Tenth Revision of the International Statistical Classification of Diseases and Related Health Problems, Australian Modification (ICD-10-AM) 4th Edition. This classification system is described as:

> *"…….. a system of categories to which morbid entities are assigned according to established criteria. There are many possible axes of classification and the one selected will depend upon the use to be made of the statistics to be compiled. A statistical classification of diseases must encompass the entire range of morbid conditions within a manageable number of categories*" {National Centre for Classification in Health, 2004}

The classification system documentation has 5 sections, each with a specific purpose and type of information. The first is a tabular list of disease, and related health problems that includes the hierarchy of the code system and notes on inclusion and exclusion and aggregation criteria. The second is an index used to direct the user to the code and instructions for use in the first section. Sections 3 and 4 include the "*Australian Classification of Health Interventions*" (ACHI) to classify and report procedures, treatments and other health care interventions, in both a tabular and index format. The 5th and last section is the "*Australian Coding Standards for ICD-10-AM and ACHI*" (Coding Standards). The coding standards are used to provide additional instructions on the coding of cases to meet the requirements of morbidity reporting. Each section supports the manual use of and specifies the business rules applicable to sound coding convention for secondary uses of admitted patient data collection {National Centre for Classification in Health, 2004}.

Figure 1-2 shows the flow of data collection for morbidity data in Australia today. At the hospital administrative episode data that defines the admission and discharge details (start and end of the episode) are entered into the hospital computer system at the time they occur. I Australia today the paper based record of clinical care is used as the source of information for manual allocation of codes to represent the clinical course of the episode of care. In some States it is a legislative requirement that only information documented in the clinical record can be used to inform morbidity coding for reporting of inpatient episodes of care. Australian hospitals are moving to electronic mechanisms for health recording, but this is still in the very early stages of development. The morbidity codes identified by the manual coding process are recorded in the hospital computer system. The hospital computer system periodically extracts and sends all episode details not previously submitted to the state data collection. The hospital also has access to the morbidity data for internal use. The state audits data received, and notifies the hospital of any corrections necessary while storing the accepted episode details into the State Morbidity Data Collection. Periodically this state collection has data extracted and sent to the National Morbidity Data Collection. At both the State and National level the data is used to support a range of epidemiology and administrative purposes.

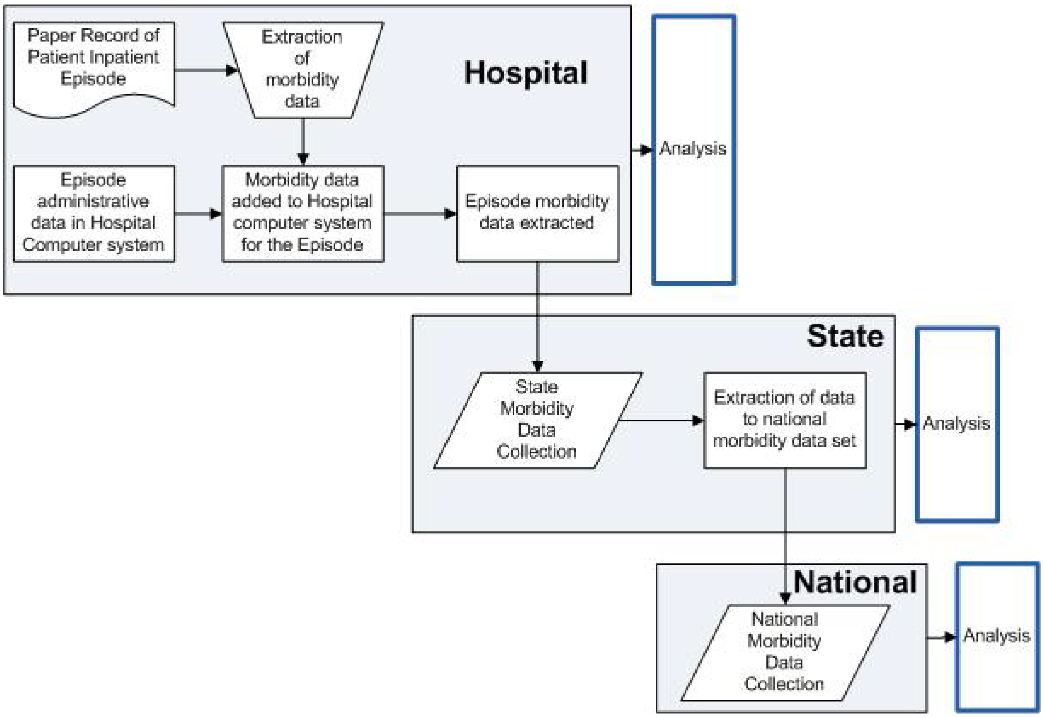

Figure 1-2 Flow of data collection for morbidity data in Australia

Figure 1-2 indicates the flow of information from the hospital data collection system, through the State to the National data collection system. At each stage there is the ability and requirement to extract and analyse data to meet the needs of the different user groups.

1.2 The Purpose of Morbidity Reporting

Formal collection of morbidity data began in the search for clinical and public health knowledge to solve specific health problems, particularly those of communicable disease in the 1800s {Brachman, 2006}. The AIHW indicate that morbidity data collection is used for

> "*developing effective health policies and programs, to ensuring quality provision of services, to coordinating treatment and care and empowering consumers*"{AIHW, 2006a}.

Today the uses of these data have extended well beyond those original boundaries, and now include service planning and fiscal control.

The National Centre for Classification in Health (NCCH) in their Strategic Plan 2000-2005 stated that classifications of morbidity provide:

> "*information about health and use of health services is a complex resource vital to a diverse range of consumers and providers of health services*" {National Centre for Classification in Health, 2000}.

The World Health Organisation have identified a range of stakeholders including health service managers at all levels, health promotion programme managers, health policy makers and strategist. To these we must add the health service and financial planning requirements and impact of casemix, a costing system for health care calculated from the codes in the National Morbidity Data Collection{National Centre for Classification in Health, 2004; Roberts, 2003}.

1.2.1 Epidemiology

In Australia, the Australian Institute of Health and Welfare uses the data in the National Morbidity Data Collection to support a wide range of epidemiological purposes, including injury and chronic disease investigations {Stephenson, 2003; AIHW, 2007a; Australian Centre for Asthma Monitoring, 2008 ; AIHW, 2007c; Harrison, 2007; Henley, 2007; Tong, 2007; AIHW, 2007b}.

The Australian Department of Health and Ageing (DoHA) has invested in making morbidity data available for independent researchers. During the 1990's DoHA supported the development of HealthWIZ. This software system combines National social and health data with computer tools to support extraction and statistical analysis of that data by health professionals. The 2007 version of the HealthWIZ data library includes Admitted patient morbidity data from every State and Territory's admitted patient morbidity data collection covering a period of 1992 – 2003{Prometheus, 2006}.

A search of the MEDLINE database of citations and abstracts of biomedical life sciences journal literature using the PubMed search engine to access publications using the criteria "*epidemiology admitted patient morbidity data*" for the calendar years 2005, 2006 and 2007 returned 1473 entries, an average of 491 per year. Though not all of these publications necessarily used National collections, nor was the search able to be restricted to exclude administrative research, a review of the titles of the papers returned showed a strong epidemiological, clinical bias to the topics and a strong use of morbidity data collections as a source of epidemiological research.

In recent years bio-surveillance has become an additional focus of epidemiological research using morbidity data, often through expansion of the data or linkage of data from multiple sources. For example: influenza surveillance data has expanded to include alternative sources such as emergency department data, absenteeism, reports, pharmaceutical sales, website access and health advice calls {Dailey, Watkins, Plant 2007}.

Criticisms have been made of morbidity data collections, and the inability to obtain meaningful data beyond the simple main condition or procedure, the reason for admission {Humphries, 2000; Campbell, 1997; Chute, 2000}. Other information systems have provided greater utility through the application of suitable structure to the system and the data in that system. {Endres, 2006}

The capacity of computer systems to collate, compare and link data quickly and to apply statistical probabilities to trends offers the attractive proposition of early warning of changes in health trends. Admitted patient morbidity data is a relevant data set for inclusion and expansion of these initiatives. This research supports the definition of the concepts within the morbidity data in a way that is computable and comparable with data from other systems, thereby supporting this objective of the collection.

1.2.2 Health Service Administration and Planning

The morbidity information collected in the local hospital system is used by management of the health institution to indicate the number and type of services provided and by medical staff to support their research activities{Huffman, 1994}.

Health service performance indicators represent an additional use of morbidity data. Statistical measures are used to indicate services provided. These measures include a review of the hospital's casemix data, patient length of stay comparisons between different hospitals and service areas, patient demographics. Casemix information is calculated using clinical morbidity data coded using ICD-10-AM and details of the individual episode, such as the person's age and type of service. The casemix calculations are used to group each individual admitted patient episode into an Australian Refined Diagnosis Related Group (AR-DRG) a classification system used to group cases of similar clinical and financial characteristics {Mooney, 2002 ; Roberts, 1998} {AIHW, 2005 }.

During the 1990's Australia extended the use of morbidity data collection to support direct fiscal re-imbursement and service agreements. This has added a level of tension to the collection and the rules of what is included / excluded. Decisions on what is included and excluded from the collection have been made not only on the clinical importance of the data included, but also on the difference the specific element may make to the financial category into which the patient's episode of care will fall. For example: Some hospitals interested in obtaining information about anaesthetic impact and nursing care requirements coded details of patients with post-operative vomiting. These cases were relevant to nursing impact and to investigation of anaesthetic events; the recording of the post-operative vomiting code also changed the casemix group into which the patient's episode of care fell into a more expensive AR-DRG. The coding standards currently include the instruction:

> "*Some transient conditions occur in the postprocedural period but are not regarded as postprocedural complications.*
> *Examples of such conditions are:*
>
> - *anaemia*
> - *cardiac arrhythmias*
> - *confussion*
> - *electrolyte disturbances*
> - *headache*
> - *hypertension*
> - *hypotension*
> - *nausea*
> - *paraesthesia*
> - *urinary retention*
> - *vomiting*
>
> ***Classification of transient conditions***
>
> *Transient conditions should not be coded as postprocedural complications when occurring in the postprocedural period, unless they are present at discharge or persist postprocedurally for at least seven days or there is documentation by a clinician that the condition is a complication of the procedure. If it cannot be determined whether a condition is transient or persistent, then the condition should not be coded as a postprocedural complication*" {National Centre for Classification in Health, 2002}.

There are a number of impacts to this decision. The simplest is that only those cases with ongoing or severe problems are included in the National health data, and in the calculations of AR-DRGs) upon which hospitals are funded (in Victoria this funding is directly related to the AR-DRG mix at the hospital. The funding weight for cases with a complication such as those on the list of conditions not to be included is higher than for those without such a complication. This decision also means that nursing staff or anaesthetics who were using this data to monitor performance and workload at a local level would no longer have this information available, and the ability to measure the impact or otherwise of these minor complications on length of stay, or the relationship to other intervention relationships is no longer a possibility.

No judgement is made here on the relevance or otherwise of the decision to restrict these codes inclusion in the morbidity data collection, however, it is clear that there are competing requirements for data inclusion. There is clearly a requirement to identify conditions that are transient and thereby exclude them from casemix algorithms or analytical purposes where these conditions are not of interest.

Chute and others indicate that the emphasis, in the USA at least, is now strongly fiscal rather than epidemiological, while the relationship to the financial side of health care organisations such as the Centre for Disease Control (CDC) in the USA use the data for administrative and epidemiological purposes as well {Health Statistics Branch - Health Information Centre, 2002; Chute, 1996; Chute, 2000; Hausam, 1996; Treloar, 2003}. This is particularly true in the USA as it relates to USA medicare reimbursement.

1.2.3 Purpose of the Classification System used to describe Clinical Information

The National Centre for Classification in Health state that:

> "*ICD-10-AM is built and maintained to meet the needs of all users. It is: clinically relevant, scientifically robust, a living classification, published in hard copy and electronic formats*" {National Centre for Classification in Health, 2004}.

ICD-10-AM and other variations of the International Classification of Diseases aggregate heterogeneous admitted patient episodes of care into broad categories according to standardised rules, clinical context and concepts. The ICD-10-AM coding standards indicate that the use of the coding system as a representation of clinical information is secondary to clinical care, meaning that it occurs after that care and is produced as a by-product of the information collected naturally though the administrative and clinical processes.

The uses at different points in the data collection continuum are also a confounding element. Users at an individual hospital level have clinical, direct administrative and planning uses for the local data and for comparison with wider data collections. The State has a governance role, and fiscal requirements according to the Australian Healthcare Agreements and use the data for service planning and public health monitoring and planning. This research seeks to identify the ontology of the facts held in the morbidity information system and to consider whether the current structure of data collection supports the needs defined of the data collection. The promise of well constructed and appropriately modelled information systems is that these systems provide improved re-use, flexibility and accuracy of the knowledge held in data {Pisanelli, 2003; Miguel E. Ruiz, 2005; Wand, 1996}.

Stakeholders also include those responsible for the development, maintenance, collection and use of the data. In Australia the governance of admitted episode classification and morbidity data collection rules are managed by the NCCH and The Australian Institute of Health and Welfare.

In summary, the purposes of morbidity reporting are diverse. The users of the data occur at all levels of the health system, within a clinical unit of a hospital, hospital administrators, State and National planners and epidemiologists.   Table 1-1 provides an overview of these uses and user communities.

Table 1-1 Purpose and users of admitted episode morbidity data

Purpose	Community of Users
Identification of health trends and monitoring of treatment protocols	Clinical research – at all levels, hospital department, hospital, regional, speciality, State and National.
	Researchers including epidemiologists
Public Health Surveillance (Biological Surveillance)	Researchers including epidemiologists
	Health policy and planning
	Health service governance
Health Service Performance Indicators	Health policy and planning, including indicators of population health (disease incidence rates) and health service performance (average length of stay).{McCormick, 1990; AIHW, 2004a}
Legislative reporting requirement	Hospitals have a legislative requirement to report to the State health authorities and the State Health Departments are required to provide data under the provision of the Australian Health Care Agreements.
Fiscal governance	Hospital administration, State and National Government

1.3 Development of Morbidity Reporting Systems

Each State has been responsible for the development of systems to collect morbidity data and to provide that data to the National data collection maintained by the AIHW.

The principle motivators for moving to electronic collection mechanisms were to improve the efficiency of collection by moving collection closer to the source of the data, and the timeliness and quality of the data. The data being entered into the National collection system in 1975 was for people treated in hospital in 1970 {Travaks, 1975 }, this time delay was detrimental to using these data for the management of the health care system and potentially to the health of Australians. The increasing use of computer technology for data collection in Australia reflects both the increased capacity of these systems to provide benefit, and the ever decreasing cost of the system. The move to computerized data capture moved the collection process closer to the source of data and significantly reduced the data collection time frame and workload, both at the hospital where duplicate entry into indices and report forms were stopped and at the data collection point where coding and data entry tasks no longer occurred.

Today the data for the 2005-2006 year were not only received but reported upon in the Australian Institute of Health and Welfare's publication on Australia's Hospital Statistics by

May of 2007 (within 10 months of the end of the fiscal reporting year){AIHW, 2007c}. Computerisation also supported data validation processes. The Victorian Admitted Episode DataSet User manual provides details of and guidance on the monthly submission of morbidity data required of each hospital. It incorporates an automated error management system of reporting and correction aimed at improving the quality of the data received.{Health Data Standards and Systems Unit, 2006}

From a systems perspective the introduction of a computerised approach to morbidity data collection has required the development of a set of different software modules. At the hospital there are two components:

1. the hospital computer systems that can collect the clinical morbidity data. This system is designed by each hospital system software provider to meet State specifications. These specifications are modified yearly to coincide with fiscal year data collections.{Data Services Unit, 2005; Health Data Standards and Systems Unit, 2006; NSW Health, 2004}
2. the data extraction module for extraction from hospital computer system to send data to the State collection. Each software vendor produces their own extraction mechanism, though most States mandate the use of the State's defined HL7 standards for transmission of messages.

The hospital systems are able to use their standard report generation facilities to provide analytical capacity on the data in their systems, including the morbidity data.

At the State health department three components are required:

- Data receipt module to collect, edit and check the data received to ensure that it is compliant. They produce reports of receipts and errors which are returned to the hospitals for action and confirmation.
- Data storage and analysis capacity
- Data extraction module to extract data for submission to AIHW to meet the State's admitted patient morbidity data legislative reporting requirement.

The State is also responsible for maintaining instructions and oversight on the morbidity collection system and the data within it.

Nationally

- Data receipt module to analyse the ‘clean’ data received with the capacity to identify and mark any anomalies in the data.
- Analysis tools.
- Governance tools. The AIHW maintain a detailed data dictionary of each data element required in the National admitted patient morbidity data collection.

These multi-level systems are expensive to maintain and the Health Data Standards Committee, the group responsible for advising modifications to the data collection, have been cognisant of the cost of any changes to these systems. Costs are both those associated with computer system changes and those associated with data degradation when changes are made that may affect the ability to compare data over time, and thereby degrade the value of that data. This research seeks to identify a generic approach to the structure and concepts held in morbidity data that will support a more flexible and standardised approach both to data extraction and collection, but also to data longevity.

In the 1970’s computerised admitted patient systems were introduced into Australian hospitals{Health Computing Services Victoria, 1979}. In Victoria the Patient Reporting System (PRS) was developed in the 1970s to collect details of admissions, transfers, discharges and morbidity for individual hospitals. The design of the morbidity data collection system was built to meet the structure and data components in the form used in the previous collection method. During the early 1970’s database technologies were just emerging and the science of system engineering was in its infancy. Neither of these technological approaches appears to have been considered when the morbidity data capture system was designed originally.

The data required to be included in morbidity data collections have been extensively defined over the years, both at State level and Nationally, originally through the National Health Data Dictionary, and today in the Australian Institute of Health and Welfare’s Metadata Online Registry (METeOR){AIHW, 2004b; Health Computing Services Victoria, 1979; Health Data Standards and Systems Unit, 2007; Department of Health, 2006; Data Services Unit, 2007}.

Morbidity data collection has grown in the number of items collected but changed little in core structure over the years {Health Data Standards and Systems Unit, 1999; Health Data Standards and Systems Unit, 2003; Health Data Standards and Systems Unit, 2006}. The National minimum data set shows that there have been changes to various fields, the inclusion/exclusion criteria and the value domain (codes used as defined in Meteor) each year {AIHW, 2006b; Health Data Standards and Systems Unit, 2004; Health Statistics Branch - Health Information Centre, 2002 ; NSW Health, 2003 }. There are many rules defined for Victoria in the Victorian Admitted Episode Database User Manual. Though many of the data collection systems include verification of the rules, there are no systems in place that automatically correct or apply these rules for classification of clinical concepts {Health Data Standards and Systems Unit, 2006}.

The collections include limited, de-identified demographic data, information on diseases, injuries and procedures. The number of disease, injuries and procedures that can be captured has increased but the structure, from a systems perspective is still a flat structure. The structure has three primary components,

- the Admission (collecting information including the person's age, sex, residential area, date of admission, source of admission referral)
- the Discharge which indicates the end of the admitted patient stay and includes information about whether the individual was discharged alive or dead, to home or other health care facility, and
- the morbidity entry, which includes diagnostic, injury and procedural information.

These component elements have not conceptually changed since their introduction in the 1970s. The structure of the data collection is flat with each data element stored in a specified sequence. Meaning is derived from the sequence of the data in the file. The format of the message used in Victoria as described in the user manual indicate that disease and procedure codes are received in a string, for example.

X1:123456:150383:1415:P123:P567:P765:::O321:O321::::

This file contains a representation of a morbidity entry. The first field - X1 "tells" the computer that the information is a diagnosis, the ":" indicates the end of the field and the start of the next field. The next field is the patient's record number, followed by a delimiter the ":", then the date of admission, time of admission, then the diagnosis codes. Note that there is no data between some of the colons. This file allows for up to 5 diagnoses and there are only 2, there are also two operations recorded and room for three more. The limitations of this type of file structure include the inability to identify which operation was performed for which procedure, and to record all diseases or procedures that might be relevant. These are limitations of early computer system design.

The Victorian ICD Coding Committee was formed in 1979 to support improvement in the consistency of morbidity coding in Victoria. The Victorian Annual Reporting Act 1983 required public hospitals to report morbidity data to the State. The decision was made at that time to use a revised PRS system to collect these data. The new system became known as PRS-2. During the 10 years since the initial development of PRS, computer technology had advanced significantly and was used more broadly in health care. However, during that same time period there had been sufficient system failures and cost overruns to make those in healthcare reluctant to change or significantly enhance any system that was functional. Though there was advice at the time that the data structure and design should be improved, the decision makers responsible for the collection of morbidity data were not experienced in computer technology, and did not have an understanding of best practice, or the potential offered by improved system design{Victorian Coding Committee, 1983}. For this reason, the structure and data components of the morbidity data collection were not changed in any significant way at that time.

In 1993 the National Health Information Agreement required State Health Departments to collect public hospital admitted episode morbidity data and provide it in electronic form to the Australian Government Health Department {Health Data Standards and Systems Unit, 2003}. This requirement continues today, though the introduction of EHRs and new information systems in hospitals make this an appropriate time to reconsider the design structure of the morbidity collection system.

1.4 Electronic Health Record Development

In his State of the Union Address in 2004 George W. Bush identified health computing as a major strategic direction for the USA. In this address he provided many reasons for this initiative, these objectives included establishing a system that would support the identification of bio-hazards in the community{Bush, 2004}. The use of data collections to inform communicable disease trends is not an issue that has disappeared. Outbreaks of Severe Acute Respiratory Syndrome (SARS) in 2002-3 generated a desire to improve and extend existing morbidity data collection systems using the Electronic Health Record as a basis for more timely information availability.

There are many initiatives in Australia aiming to develop electronic health records and information systems. The Health Online action plan {National Health Information Management Advisory Council, 2001} aimed to support the coordination of these efforts to improve health outcomes.{HealthConnect, 2003} These initiatives concentrate on the potential of new health information systems to collect data at source in order to meet clinical needs.

All States have committed to software products that will allow their hospitals to introduce clinical information systems over the next few years {Department of Human Services - Victoria, 2003; Bryan, 2004;}. These systems will gradually provide a computer based health record that will support improved clinical practice and communication but also provide a resource for extraction of data to meet the needs of morbidity reporting {HealthConnect, 2003; Information Communications Technology Standards Committee, 2006a}. Clinical information systems are also likely to support data extraction as a 'real-time process' {Roberts, 2003} allowing morbidity data to be produced more quickly. The hope is that these systems will also provide improved 'by-product' data through the automation of data extraction processes from the original source data {HealthConnect, 2003; Roberts, 2003}. The Australian Health Information Council have called for an update to this strategy in recognition of the developments made by the States and the availability of more of the functional components of the EHR {Australian Health Information Council, 2008}.

1.5 Opportunity for Improvement

Hovenga et al investigated the use of health data collections in Australia from a data warehousing perspective and identified issues with obtaining quality, reliable data from the collected data. There were a range of causes including the current inability to link data and the lack of a relationship of the data back to clinical source data {Hovenga, 2002a}. This is one of the issues that could be addressed by the capacity to obtain by-product data from electronic health record systems.

While the existing legacy systems used throughout healthcare represent a significant financial and infrastructure investment and are unlikely to be removed from service quickly, the progressive implementation of electronic health record systems throughout Australian health care offers an opportunity to review and improve the systems used to collect morbidity data. Wangler has identified that it is more efficient to solve problems of information transfer using IT based process changes {Wangler, 2003}.

1.6 The Place of Ontology in the Study

In computer science, the concept of an ontology describes the structure of a system, a mechanism for describing what exists in a system. Ontologies have been developed as a tool in information systems to identify the concepts of relevance of the domain, the semantic modelling of and relationships between these concepts {Gruber, 1993; Noy, 2001; George, 2005}. This process supports identification of the concepts in a system, identifying their relationships and mutual exclusivity in an effort to provide unambiguous meaning to data in a system.

Pisanelli, et al. state that an ontological approach can reduce the potential misunderstandings of the meaning of data, and provide a 'solid conceptual foundation {Pisanelli, 2002}' that supports data sharing and re-use as does conceptualizations of the concepts and their relationships going behind the terminology, or in this case a classification use case {Pisanelli, 1999; CEN TC251 WGII Terminology and Knowledge Bases, 2005;Pisanelli, 2003; Gangemi, 2006}.

Morbidity data are collected for a wide range of purposes and the ontological approach has the advantage of modelling the knowledge represented in the real world for a specific domain (clinical morbidity data) at a high level of abstraction which supports re-use of the information. It is acknowledged that re-use always has difficulties represented in the clash of purpose and context. This abstraction is represented in metadata models, a technique being used increasingly as a mechanism to support the consistent representation of health concepts internationally {CEN ECfs, 1997; ISO, 2007a; ISO, 2007b; ISO, 2003a; George, 2005}. "*Metadata is used to describe different types of data for different purposes. It includes the definitions of data, the format of the data and examples of the data*" {Standards Australia, 2005}. Metadata models are used to give an overarching view of data in a system and the relationships between these data. The use of visual, model based representations of data can be used to provide clearer information on the information structure than can be achieved using linguistic representations alone. Consideration of the variations between models and how these impact the usability of the data they represent is also an issue.

In healthcare, information models are being used both to understand the domain more clearly and to identify commonalities that have the potential to improve data quality and clarify, and to simplify data system development.

The HL7 Reference Information Model(RIM) is an example of the desire to develop a generic model to underpin healthcare. Development of the RIM began in 1995 and is now at version 1.1 {HL7, 2007}. An information model is a specific view of the world, representing the concepts of that world and the specific components of the architecture or design suited to a specific purpose. The HL7 RIM is an attempt towards a standardised representation of healthcare to support system interoperability. It is therefore a model of the healthcare world, though even the use of such a high level model is not a guarantee of success.

Models that relate directly to the information in morbidity data are those that represent the patient and the information about them. There are different levels and views of this data and context is critical to interpretation of the data. These issues are discussed in Chapter 3 – Patient data vs. Clinical Domains.

1.7 Aim of the Research

This research aimed to develop an entity-relationship model of the clinical concepts admitted patient morbidity reporting in Australia to support improved information re-use and clarity of content of that data in Australia.

1.8 Research Objectives

This study aimed to investigate the components and structure of disease and procedure morbidity data in admitted patient data collection by the:

- Identification of the structural components of the disease, injury and procedure components of hospital admitted episode morbidity data collections(SQ 1.1)
- Identification of the entities and meaning relationships indicated through coding instructions represented in ICD-10-AM (SQ1.2)
- Review of State data entry instructions to Identification of any additional categorial relationships indicated in data collection instructions relative to disease, injury and procedure data. This objective supports the development of a complete model of categories and relationships. (SQ 1.3)

Throughout this text RQ is used to identify a research question and SQ to identify a sub-question.

Another objective was to establish the ontological entities and the semantic relationships between these concepts identified and the categories of information represented in a model of the information in hospital admitted episode morbidity data collection (RQ2).

As the morbidity data collection is an abstraction of the clinical environment represented by ontologies such as SNOMED-CTand OpenEHR information model, a review of the high level ontological structures of these models needs to be undertaken to identify similarities and variations between these existing models and the morbidity data collection (RQ3).

1.9 Research Questions

The central hypothesis of this research is that:

An ontology model can be used to define the structure components of the clinical elements of admitted patient morbidity data collections.

The following questions (RQ) and sub-questions (SQ) need to be answered in order to develop the ontological model required to support the central hypothesis.

RQ1 What are the concepts and relationships in the current Admitted Episode Morbidity Data Set?

SQ 1.1 What data elements are required for the national collection, and for State morbidity collections for disease, injury and procedure information collections?

SQ 1.2 What meaning relationships are represented by the coding instructions held in ICD-10-AM?

SQ 1.3 What concepts and relationships are indicated by instructions for data collection from the State data collection specifications for disease, injury and procedure information?

SQ1.4 Do the variations in State collections represent additional or different concepts or relationships?

RQ2 What is the entity-relationship model and the semantic relationships that represents the data collection as defined by the answers to research question 1? (RQ1)

RQ3 Is the entity-relationship model consistent with the concepts in the real world morbidity data and in existing clinical ontologies.

1.10 Scope of the Research

Admitted patient morbidity data is an extracted summary of data from the patient's record of hospital care. It is made up of information about the admitted patient event and the individual (patient demographic information, admission and discharge) and clinical information (diseases and procedures) collected using ICD-10-AM.

The clinical information included in morbidity information can represent multiple concepts in single fields (many diseases and causes of injury are all collected in a single Diagnosis concept with multiple occurrences and procedures, anaesthetics, allied health interventions are all recorded in the single Procedure concept with multiple occurrences. In the existing information systems this approach has limited the usability and clarity of the information in those fields. The morbidity data collection includes patient details an information about the admission, such as source of advission as well as diagnosis and procedure fields. To focus on the issues of clinical information representation the research is limited to an investigation of the clinical component of morbidity data collection (Diagnosis and Procedure fields only).

This scope is also influenced by the long standing existence of systems that collect episode and individual data for administrative purposes, while the introduction of systems to collect and maintain clinical data is an emerging development in Australia. This research therefore sought to provide a basis for the development of clinical morbidity data in these hospital based clinical information systems and did not include episode and individual data in the research.

Relevant and reliable sources of data content and relationships were required to identify the entities within clinical data. These sources were limited to documents that describe data collection through National metadata, State collection instructions and coding instructions and standards, and confirmed using a small set of samples of real world data. The ability of the coding system (ICD-10-AM) used to collect the clinical concepts for Diagnosis and Procedure fields in morbidity data to represent the concepts found in the original source medical records was not considered in scope as these are the value domain of the work, rather than the structure of the data collected.

The data instructions change over time. This research has referenced instruction sets and metadata from 2005 to 2008 (January) and where relevant indicates modifications made during that time period. ICD-10-AM coding standards applicable during this time period were reviewed to confirm the data structure and rule based requirements and to mitigate against an individual year which may have had less inclusive structures or rules. State collection instructions were investigated from the States with the largest collections (Victoria, New South Wales and Queensland) and Western Australia. Western Australian requirements were used as this was the only State with a linked data warehouse into which the resulting morbidity data is collected. This State warehouse has developed and maintained "*a system of linkages connecting data about health events for individuals in WA"* {Information Management and Reporting Directorate, undated}.

The coding standards and instructions were investigated from ICD-10-AM 4th Edition and 5th Edition as these were the code sets valid during the chosen data collection period 2005 – 2008.

Existing models for electronic health records provided methodological input and were included in the capacity of identification of variations between these points of care clinical information models and morbidity collection. The research did not include an evaluation of the models themselves other than their suitability for the purpose of this research. Only those models with international standing were used.
Figure 1-4 identifies the scope of the research. The EHR models were used to inform the research but did not affect the scope.

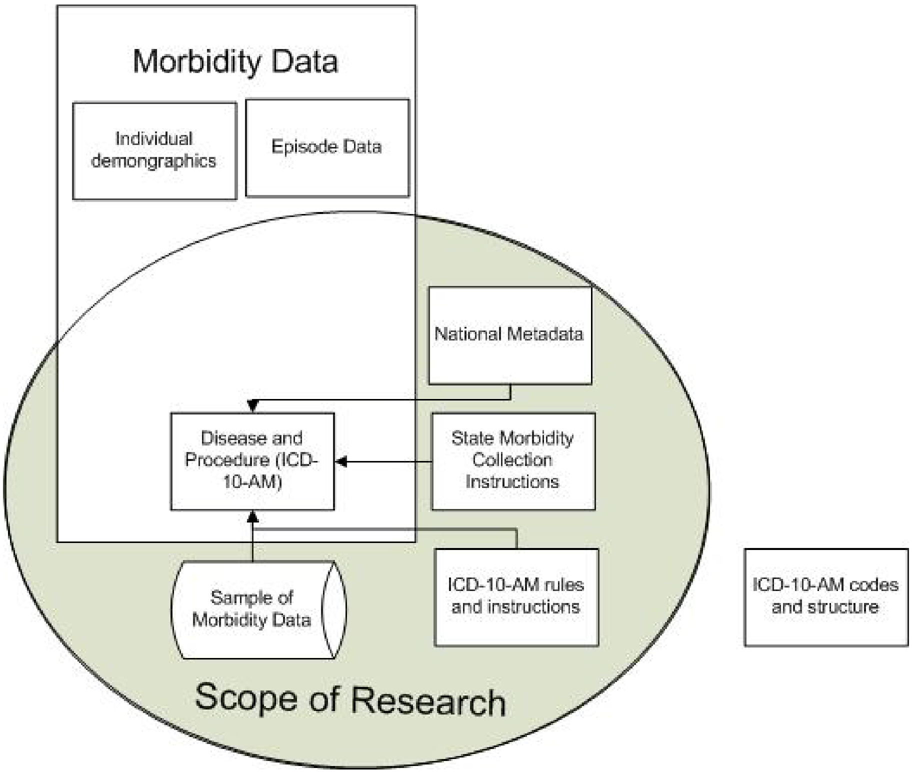

Figure 1-1 Scope of research

The research did not investigate the software used to collect, extract or transmit these data through the collection continuum as these are proprietary to each software vendor and are assumed to be compliant with the National and State collection rules.

1.11 Significance of the Research

The study will use ontological modelling techniques for the identification of entities and relationships of data collected in admitted patient morbidity data. Existing metadata standards define the structure of metadata clearly{ISO, 2004}. These metadata document the definition, source and value domains of data components and their position within the collection structure (eg: that age is an attribute of a person). Though morbidity data collection data elements have been defined for many years a model representing the concepts within the clinical components of these data and the relationships between them has not been developed. Morbidity data content has grown by the number of individual data elements and the quantity of data collected within some of these data elements but has not changed in the data representational schema or information model used to collect and analyse these data.{Chute, 2000}

This research was undertaken to:

- Develop an ontological model (using entity-relationship representation) for the collection system for morbidity data using the new technological approaches relevant to electronic health record systems, health information messaging, and the morbidity reporting systems in Australia. This model identifies the entities and relationships in the current Admitted Episode Morbidity Data Set (RQ1), and the concepts and relationships required to represent morbidity data as they compare to other clinical information models.
- Inform the emerging fields of health informatics and health concept representation of the potential requirements for representation of morbidity data. (RQ3)
- Have the potential to inform the development of National and local policy and practice in the fields of clinical information collection within an Electronic Health Record and morbidity data collection systems. (RQ3)
- Have the potential to inform the design and development of messaging requirements for morbidity data, specifically Health Level 7{21731, 2005}, the international standard for information interchange in healthcare, and the National E-Health Transition Authority{NEHTA, 2007} and could inform the development of a sustainable and re-usable data collection.(RQ3)
- Provide information that has the potential to support the development of automated approaches to the collection of morbidity coding out of the data held in electronic health record hospital environment. Further study and developments of software to meet this need and support the implementation of terminological clinical systems such as SNOMED-CT in the hospital environment in Australia could be undertaken using the conceptual model as a basis.(RQ3)
- influence further EHR standard developments.

1.12 Expected outcomes

The project is expected to provide:

- An ontology based entity-relationship model/s for admitted episode morbidity data in Australia that considers the semantic relationships between the clinical components of the inpatient morbidity data collection.
- Clearly defined issues and identify potential for changes to the hospital admitted episode morbidity data collection system through identification of the relationships and ontological requirements to obtain meaning from an electronic health record environment to support morbidity data extraction.
- Knowledge of the potential processes and structures needed for hospital acute episode morbidity data collection automation in an electronic health record hospital environment, to support the development of business cases and terminology implementation in the Australian health care system through better understanding of the ontological relationships between morbidity data and ontologies such as SNOMED-CT and OpenEHR archetypes. SNOMED-CT and OpenEHR ontologies cover different domains at different levels of granularity. Some components of SNOMED-CT fit with some archetypes. and their impact on data usage

1.13 Use of results

This research is intended to produce publishable results in the fields of Health Informatics, specifically morbidity data collection metadata, health terminology, health concept representation and data aggregation.

This research contributes to the current reviews of morbidity collection content and processes related to the development of electronic health records and clinical metadata in Australia and around the world. Publication of these results informs these developments and enables the provision of specific information to hospitals and government agencies in determining priorities for morbidity collection modification, electronic health record and terminological developments. Results will be provided to the Australian Institute of Health and Welfare as custodians for the National morbidity data collection and the agency responsible for governance of that collection.

This study is expected to produce useful information to inform National and State policy regarding the design and development of admitted episode morbidity data collection processes in an EHR environment. In the short term this information may inform the design components of morbidity data collection, while in the long term it advances the understanding of the need to be able to represent meaningful information in a consistent and computer processable manner within morbidity reporting systems.

Chapter 2 Literature Review

A review of the literature was undertaken which focused on journals, conference proceedings and other published materials in the fields of health information management, morbidity data governance, public health, ontological science, and health informatics. When searching and reviewing the Australian and international literature the following topics were actively investigated: morbidity and health data governance and collection in Australia, data structure representation, data modelling, ontology, EHR ontology, EHR modelling, system engineering, semantic relationships as well as on methodological data structure modelling, Specific literature searching web based tools used included the PubMed search engine accessing MedLine and Google Scholar in addition to accessing specific professional organizations' web sites and publications.

The Health Information Management literature was investigated as this is the profession most closely associated with morbidity data collection in Australia and internationally. Members of this profession administer the data collection systems in Australian hospitals and contribute to the development of rules governing the use of ICD-10-AM used for each morbidity data collection, through the National Centre for Classification in Health.

The scientific community in the field of information system engineering has developed the use of ontological modelling of reality as a mechanism to support more efficient and effective information and information system development and as a mechanism proven to improve information availability and usability{Gruber, 1993; Gangemi, 2006; Gangemi, 2004; Gómez-Pérez, 2002}. This literature has been included to identify any generic (non health) ontological approaches to information or knowledge modelling methodologies appropriate to the investigation of the reality of the concepts included in morbidity data.

The field of health informatics is the natural discipline at the centre of this research activity. The National Library of Medicine in the USA defines Health Informatics as:

> *"….. the field of information science concerned with the analysis and dissemination of medical data through the application of computers to various aspects of health care and medicine. —Medical Subject Heading"*{National Library of Medicine, 2004}.

Health informaticians have researched the application of ontology to represent health concepts in order to develop clinical information systems. Ontological approaches are used in two different ways to represent clinical information. Firstly as a representation of clinical knowledge, the clinical terminology used to describe health concepts, in this case the internationally adopted Standardised Nomenclature of Medicine (SNOMED), an essential underpinning for concept representation in EHRs{ISO, 2007a ; Elkin, 2004; HL7, 2007; Lacey, 2002; , 2008; Brown, 2006; National Electronic Decision Support Taskforce, 2002; HL7, 2004; HL7 Australia, 2007}. Secondly an ontological approach has also being adopted to the development of the next version of ICD, ICD-11, currently under development by WHO to make it appropriate for use in EHRs. This next version of ICD is expected to be submitted to the World Health Assembly in 2014 {Ustun, 2007}.

Electronic Health Record(EHR) system literature was reviewed, particularly that which related to the structure of the data in these systems to identify whether there have been investigations into these systems' potential relationship to morbidity data collection. This literature also provided information about appropriate methodological approaches to the representation of the concepts and relationships identified in the morbidity data collection review and to evaluate the potential of EHR systems to support morbidity data collection.

Documents included in the review include those published by government organisations and reports available to the author in her employment with the developer of morbidity reporting software for the Department of Health in Victoria during the 1970's and 1980's and more recently in her position as a representative on National and international committees responsible for morbidity data governance, content and definition.

2.1 Morbidity Data Governance

Admitted episode morbidity systems include the manual or computerised extraction of clinical and administrative information, its classification and recording in accordance with the National coding standards {National Centre for Classification in Health, 2006}.

Rules are established by National and State authorities which are applied through software products and by the human application of these standards; Figure 2-1 indicates the areas where rules are developed and applied.

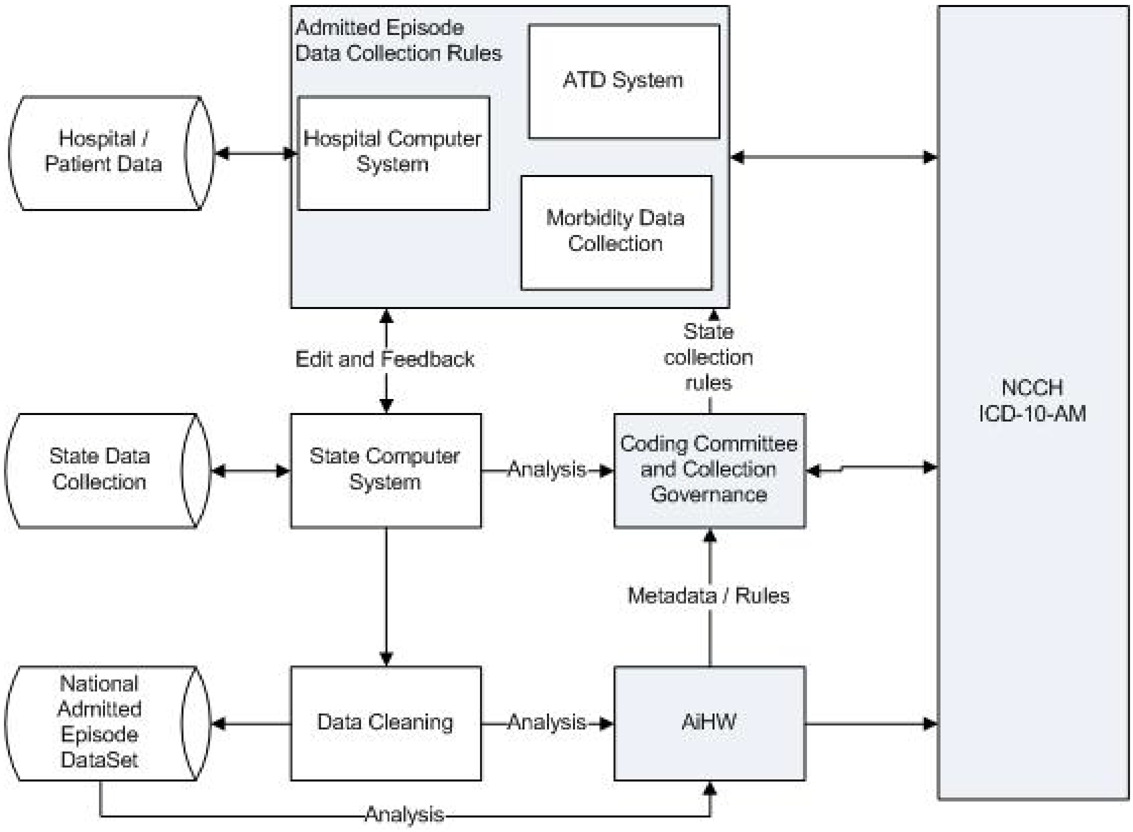

Figure 2-1 Morbidity governance rules

Figure 2-1 also identifies the source of governance rules for morbidity data collection. At the National level governance is represented by the Australian Institute of Health and Welfare (AIHW) a statutory authority established by the Commonwealth of Australia in 1987 {AIHW, 2006b}; the National Centre for Classification in Health (NCCH), the organization responsible for the development and maintenance of the classification used to represent clinical information in the collection, ICD-10-AM {National Centre for Classification in Health, 2004}

Each State takes the National requirements and may extend them to specify additional information needed at the State level, these reporting requirements are specified in detail through yearly published manuals{Health Data Standards and Systems Unit, 2006; NSW Health, 2004; Department of Health, 2006; Data Services Unit, 2005; AIHW, 2006d}. These State and National rules are implemented but not modified by the hospital systems.

2.2 National Morbidity Metadata

The AIHW is responsible for the collection and governance of the National Morbidity Data Collection. The mission of the AIHW is:

> "*To improve the health and well-being of Australians, we inform community discussion and decision making through national leadership in developing and providing health and welfare statistics and information*"{AIHW, 2006b}.

AIHW maintains detailed specifications for all data elements included in the National data collections. These details include definitions of scope, content data elements and collection instructions for the Admitted Patient National Minimum Data Set. These details are available online through their METeOR. The specifications have recently changed, indicating modifications that will come into effect on 1st July 2008. The collection will be titled Admitted Patient Care National Minimum Dataset 2008-2009{METeOR, 2008}. This indicates the scope of the data collection to be:

> "*The purpose of this National Minimum Data Set is to collect information about care provided to admitted patients in Australian hospitals. The scope is episodes of care for admitted patients in all public and private acute and psychiatric hospitals, free standing day hospital facilities and alcohol and drug treatment centres in Australia. Hospitals operated by the Australian Defence Force, corrections authorities and in Australia's off-shore territories may also be included. Hospitals specialising in dental, ophthalmic aids and other specialised acute medical or surgical care are included. Hospital boarders and still births are not included as they are not admitted to hospital. Posthumous organ procurement episodes are also not included"* {METeOR, 2008}.

This scope statement was the same as that of the previous versions of the data collection, in all versions held in METeOR since its inception in 2005. The Minimum Data Set includes Diagnosis and Procedure details (the clinical morbidity components).

METeOR provides definitions of these components as:

> *"A diagnosis is the decision reached, after assessment, of the nature and identity of the disease or condition of a patient or recipient of residential care (resident)"* {METeOR,2005}.

and declares the purpose of this information to be:

> "*Diagnostic information provides the basis for analysis of health service usage, epidemiological studies and monitoring of specific disease entities" {METeOR, 2005}.*

The diagnosis has two components, the Principal Diagnosis and Additional Diagnoses. METeOR defines these as:

Principal Diagnosis:

> *"The diagnosis established after study to be chiefly responsible for occasioning an episode of admitted patient care, an episode of residential care or an attendance at the health care establishment, as represented by a code"* {METeOR, 2005}.

The metadata instructions indicate that the principal diagnosis is required for each episode of care and "*must be determined in accordance with the Australian Coding Standards*". It "*can include a disease, condition, injury, poisoning, sign, symptom, abnormal finding,* *complaint, or other factor influencing health status*", but may not be a cause of injury of morphology {METeOR, 2005}.

Additional Diagnoses:

> *"A condition or complaint either coexisting with the principal diagnosis or arising during the episode of admitted patient care, episode of residential care or attendance at a health care establishment, as represented by a code."* {METeOR, 2005}.

These diagnoses must also be recorded in accordance with the Australian Coding Standards. Instructions are provided about injury codes included in additional diagnosis:

> *"Generally, external cause, place of occurrence and activity codes will be included in the string of additional diagnosis codes. In some data collections these codes may also be copied into specific fields"* {NCCH,2004}.

The additional "*diagnosis can include a disease, condition, injury, poisoning, sign, symptom, abnormal finding, complaint, or other factor influencing health status" {NCCH,2004).* These diagnoses represent conditions that affected patient management in that they required: treatment, investigations and/or used resources during the episode of care and are used to support casemix allocation. If a condition does not meet these criteria it must not be recorded as an additional diagnosis.

Additional diagnoses are "*significant for the allocation of Australian Refined Diagnosis Related Groups*" and they "*should be recorded when relevant to the patient's episode of care*" {METeOR, 2005}.

These instructions, provided in such a prominent position stress the fiscal purpose of the data collection over other uses. The instruction to exclude conditions not actively managed during the stay is given as additional diagnoses affect the calculation of AR-DRGs, to the extent that they may change the category of the episode of care into one that is deemed to be more costly than the case actually was if the additional condition was recorded, but not treated.

These instructions indicate that a condition of epidemiological or clinical research interest cannot be included in the data collection unless it is actively treated, investigated or used resources. This instruction has the potential to exclude information which may indicate risk factors or other diagnostic linkages between conditions,

If the data collection is to meet all of the purposes of morbidity data there is a need for the system to be able to differentiate between additional diagnoses which meet the conditions for inclusion in AR-DRG calculations (casemix) and which do not. Additional diagnoses also include external causes of injury codes. These codes,

> "*although not diagnosis of condition codes, should be sequenced together with the additional diagnosis codes so that meaning is given to the data for use in injury surveillance and other monitoring activities*" {METeOR, 2005}.

This instruction does not make it clear how the sequencing is to be used. Best practice in clinical coding indicates that where an injury code is in the principal diagnosis or additional diagnosis the cause code must follow 'somewhere in the string of codes'. Convention has the cause code sequenced immediately after the injury code or codes to which it applies. For example: A person who has a broken arm and leg from a car accident would have:

Principal Diagnosis:	Fractured Leg
Additional Diagnosis:	Fractured Arm
Additional Diagnosis:	Injured in car accident

Procedure:

> "A *clinical intervention represented by a code that*:
> - *is surgical in nature, and/or*
> - *carries a procedural risk, and/or*
> - *carries an anaesthetic risk, and/or*
> - *requires specialised training, and/or*
> - *requires special facilities or equipment only available in an acute care setting"* {METeOR,2005}.

The collection methods section of the metadata standard indicates that all documented procedures meeting the characteristics above that were undertaken during the episode of care should be included {METeOR, 2005}.

2.3 NCCH Coding System Governance

While the data fields are governed through AIHW METeOR the coding system and coding standards are governed by the National Centre for Classification in Health (NCCH) which has units in Sydney and Brisbane and *"….. is funded by the Casemix Program, Australian Government Department of Health and Ageing (DHA). The National Centre for Classification in Health (Brisbane)….. is funded by the Casemix Program, the Australian Institute of Health and Welfare, the Australian Bureau of Statistics and the Queensland University of Technology"* {National Centre for Classification in Health, 2004}. The NCCH Annual Report indicated that

> *"ICD-10-AM is built and maintained to meet the needs of all users. It is: clinically relevant, scientifically robust, a living classification, published in hard copy and electronic formats"* {National Centre for Classification in Health, 2004}.

Though the claim is made that the coding system is maintained to meet the needs of all users, the governance process described includes representatives from the clinical and coding environments but the funding mechanism would appear to give undue pressure to meet the needs of casemix. Of relevance to this work is the capacity of systems collecting morbidity data to accurately represent the characteristics of the world in which the individual elements of the classification (codes) are used such as case morbidity reporting.

The NCCH web site provides a portal to accept queries and suggestions from those using the classification and these questions are reviewed and are either answered to inform the next version of the coding system or the standards used to define the use of the coding system; questions of clarity that do not require changes to the coding system or standards are communicated back to the coding community through the NCCH coding newsletter 'Coding Matters' which is a quarterly production. New versions of the coding system and standards are produced every two years {AIHW, 2004}.

2.4 State Collection Requirements

State rules for information to be included and reported in the admitted episode morbidity collection are represented in each State's morbidity data collection system user manual. Of relevance to this work are the content and business rules for 2005-2008 clinical data elements of the data collection for the three largest States, NSW, Victoria and Queensland.

Western Australian requirements are also relevant for inclusion as this is the only State with a linked data warehouse into which the resulting morbidity data are collected. This State warehouse has developed and maintained "*a system of linkages connecting data about health events for individuals in WA" {Information Management and Reporting Directorate, undated}*.

The rules for State morbidity collection are published each time they change, often yearly. They indicate the data elements, value domains and business rules applicable to the data collection. Each hospital in the State is required to provide these data for every person admitted to that hospital. Reporting of clinical information occurs regularly after the patient has been discharged from the hospital {Health Data Standards and Systems Unit, 2007; NSW Health, 2004; Data Services Unit, 2007; Department of Health, 2006}.

The States are required to provide data that meets the National coding rules for reporting of hospital Admitted Episode Data by the requirements of the National Health Agreements {Australian Health Care Agreement Reference Groups, 2002} as published in the National Health Data Dictionary (now METeOR) {AIHW, 2005b}. The States add to this minimum data set to create their own collections.

The general process of collection of the morbidity data at the hospital was reviewed to identify concepts or issues with collection that may impact upon the morbidity model requirements.

The State collection systems have adopted transaction based reporting that is undertaken periodically. Hospitals collect data through the course of the patient's care. The clinical component of Admitted Patient Morbidity Data is added to the existing details of the admitted patient episode of care.

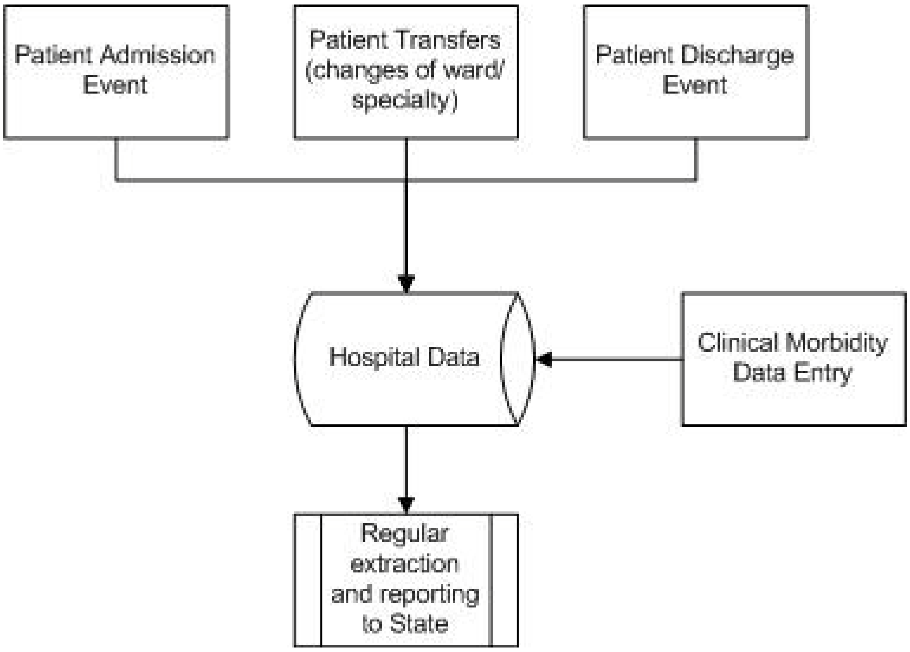

Figure 2-1 Building admitted patient morbidity data at the hospital

Figure 2-2 indicates the different stages of data collection for morbidity reporting. In this instance the hospital has a weekly batch extraction of data from the hospital system to the State collection. Though there are variations around Australia this reporting process can occur weekly, fortnightly or monthly from individual hospitals. When the weekly extraction occurs the hospital system sends whatever event details are available. For example: if the patient has been admitted but not yet discharged the system forwards admission details, and any transfer details, but no discharge information is sent to the State system. In the next weekly transmission the hospital system will send the discharge, if this has occurred. Morbidity data is usually entered after the completion of the admitted patient event (after discharge) and forwarded to the State collection in the next extraction run following that data entry.

In some cases State collection rules take precedence over the National rules {Information Collection and Management, 2006}. Investigation of State coding instructions must be included in the review of materials that identify content and structure of the clinical information collected.

Clinical data are often referred to as the diagnoses section, though it contains details of diagnoses and procedures along with other related clinical information {NSW Health, 2004; Health Data Standards and Systems Unit, 2006; Data Services Unit, 2005; Department of Health, 2004}. In Victoria, the Patient Reporting System of the late 1970's collected up to 5 diagnoses and 5 procedures for each episode of care. The diagnosis code section included information on the cause of injury. The researcher was responsible for the development of the user manual for this system. The instruction provided for the capture of injury information required the entry of an injury code or group of injury codes t be followed immediately by the cause of injury code thus representing the relationship between the injury and the cause of the injury{Grain, 1979}. This convention has not changed since that time, except for the current requirement to also record the place of occurrence and activity being undertaken at time of the injury{National Centre for Classification in Health, 2004}.

When developing a data model for morbidity data, knowledge of the volatility of the data is useful. A review of historical information on the structure of data collected in admitted episode morbidity collections was undertaken to give a reference of stability. Has the structure of the data changed in any major way over the years? Components of clinical data included in Diagnosis in the early 1980's, as represented by the Victorian data collection are detailed in Table 2-1.

Table 2-1 Data Elements in the Victorian Clinical Morbidity 1980 {Grain, 1980}

Data Element	Description and limitations
Principal Diagnosis	Mandatory and used to represent the most important condition treated during the admitted patient stay. Each State defined this concept according to their own requirements, but as the data was largely used within individual hospitals and there was no process of monitoring the quality of State data collected, the information recorded was often difficult to compare outside an individual collecting organisation.
Additional Diagnoses	Up to 4. If the principal diagnosis is an injury, then the additional diagnosis must include the cause of the injury. Convention at that time had the cause of injury immediately follow the last injury code to which it related. Eg: A patient admitted with a fractured leg, and arm from a car accident would have a principal diagnosis of Fractured leg, followed by fractured arm, followed by the cause code for car accident. If the person had other disease conditions relevant to the stay these would follow after the cause code. This could include recording of the patient's diabetic status.
Principal Procedure	In early data collections this concept was ill defined. It was often used to record the clinically most significant procedure during the admitted patient stay.
Additional Procedures	Up to 4 additional procedures were recorded if desired by the hospital.

Table 2-2 shows the structure for messages indicating clinical morbidity data sent to the Victorian State collection in 1980.

Table 2-2 Structure of clinical morbidity data in Victoria in 1980{Grain, 1980}

ID:Patient_Id:Episode_ID:Principal_Diag:Diag2:Diag3:Diag4:Diag5:Principal_Proc:Proc2:Proc3:Proc4:Proc5

This type of structure, where there is no relationship between the components of data other than their existence and position within the transactional data collection is referred to as a flat structure.

In 1982 Victoria established a 'Clinical Coding Committee' to provide support to the Department of Human Services and to clinical coders to assist in the re-design of the data collection system for morbidity data and provide advice to clinical coders to improve coding quality and consistency. As a member of this committee at the time the author was party to these design discussions. Though by this time there were well established principles of data base design established {Chen, 1976}, the use and structure of the information in the morbidity data were not modified, other than the addition of additional diagnosis and procedure fields. The clinical data to the Victorian morbidity collection in 1993 was called the Diagnosis Record. Table 2-3 shows the structure used for these data at that time.

Table 2-3 Diagnosis Record: file structure {Acute Health Division, 1993}

	Field name	Maximum characters	Alpha/ numeric	Format/Values	Clinical Component
M	Transaction Type	2	A/N	X1	
M	Unique Key	6	A/N		
*	Diagnosis code x 12 - each code	8 (8 x 12)	A/N	)Each)left justified)and)with)trailing spaces	Yes
¶	Procedure code x 12 - each code	8 (8 x 12)	A/N		Yes
#	Admission weight: infant <365 days of age	4	N	In grams, or spaces	
M	Intention to re-admit <28 days	1	N	0,1,2,3,4,9	
	User field	1	A/N	Optional field, free content	
†	Duration of Stay in ICU	4	A/N	0000 to 9999 or spaces	
M	Duration of Mechanical Ventilation	4	A/N	0000 to 9999 or spaces	
+	AR-DRG V3.1 - hospital generated	3	A/N	NNN or spaces	
‡	Duration of Stay in CCU	4	A/N	0000 to 9999 or spaces	
"	Reason for Critical Care Transfer	1	A/N	X, E, J, W or spaces	
	Filler	18	A/N	Spaces	
		240			

The diagnosis and procedure sections of this collection were still unstructured. They are a list of up to 12 diagnosis and/or procedure codes. This situation is consistent with morbidity data collection in each State. By 2006 the number of codes had increased but the data elements had not changed and the coded clinical information continues to be unstructured. This research focuses on the content within two clinical 'fields', diagnoses and procedures. Table 2.4 shows the changes in the number of diagnosis and procedure data items collected in the Victorian Admitted Episodes Database (VAED).

Table 2-4 Number of clinical concepts in Victorian admitted episode morbidity collection over time {Health Computing Services Victoria, 1979; Health Data Standards and Systems Unit, 2003; Health Data Standards and Systems Unit, 2006}

Time Period	**Number of Diagnoses (one or)**	**Number of Procedures (none or)**
Early Registers	1	1
1970's	5	5
1980's	12	12
2000	25	25
2006	40	40

This structure reflects the form based origin of the data collection. This structure, or lack of it, presents issues when analysing the data. For example: though complications of procedures are recorded amongst the many diagnoses that can be included in the data collection, there is no facility to relate the diagnostic description of the complication to the procedure recorded. Though the complication code used can indicate that the source of the complication was a procedure and give some details of the type of procedure, such as 'other reconstructive surgery' but it is not possible to identify whether the procedure was in this admission or a previous one, nor, if the person had more than one reconstructive surgery during the current episode, or which of these procedures generated the complication. The classification system and the sequential nature of the code entry are used to imply relationships but the computer system does not support ease or accuracy of analysis nor quality representation of knowledge.

When computer systems were introduced the computer design was a direct copy of the form data collection system. Minor variations in the way that States structure the data collected exist, though these systems are all able to reformat their data to meet the reporting requirements of the National data collection. Table 2-4 previously presented shows some of these variations.

Table 2-5 shows the clinical data items used within different State and National admitted episode morbidity data collections. Derived data components such as AR-DRGs are not included. Where the data component names are different but the definitions of the data element represent the same 'idea' they have been represented together.

All States collect principal diagnosis. Though their collections of additional diagnostic data vary widely in their representation, the effect of these variations is minimal. All of these structures must be able to extract data to meet the National requirements.

The instructions for State data collections did not indicate any reasons for variations between the States.

Table 2-5 Variations in clinical morbidity data structure by collection jurisdictions

Clinical data items	National	NSW	QLD	VIC	WA
Source	{AIHW, 2006d}	{NSW Health, 2004}	{Data Services Unit, 2005}	{Health Data Standards and Systems Unit, 2003; Health Data Standards and Systems Unit, 2006}	{Health Statistics Branch - Health Information Centre, 2002 Department of Health, 2004 }
Principal Diagnosis	✓	✓	✓	✓	✓
Co-Diagnosis					✓
Additional Diagnoses	✓	✓	✓	✓	✓
Diagnosis Type			✓	✓	
Procedures	✓	✓		✓	
Additional Procedures					✓
External causes of injury or poisoning	✓	✓	✓		✓
External cause related to associated diagnosis			✓		
External cause associated with the complication			✓		
Places of occurrence of external cause	✓				✓
Activity when injured	✓				✓
Morphology			✓		✓

Table 2-5 also represents issues that need to be considered in the model. The Queensland User Manual indicates that unlimited numbers of external causes can be included, there is no capacity to relate these to a specific injury or disease when more than one injury or cause of injury applies, other than through the sequence of data entry {Data Services Unit, 2005; Department of Health, 2004}.

Causes of injury are submitted in the diagnosis section of the collection and are distinguished by the codes used {AIHW,2007b}.

Figure 2-3 shows the diagnosis related data items collected by the Victorian and Western Australian morbidity collection systems {Health Data Standards and Systems Unit, 2003; Health Statistics Branch - Health Information Centre, 2002}. Each State collects details of the diagnosis relevant to the episode of care, the causes of injury, place of occurrence of each injury and the activity that occurred when the person was injured. If the diagnosis was of cancer then the morphology of the cancer is also coded.

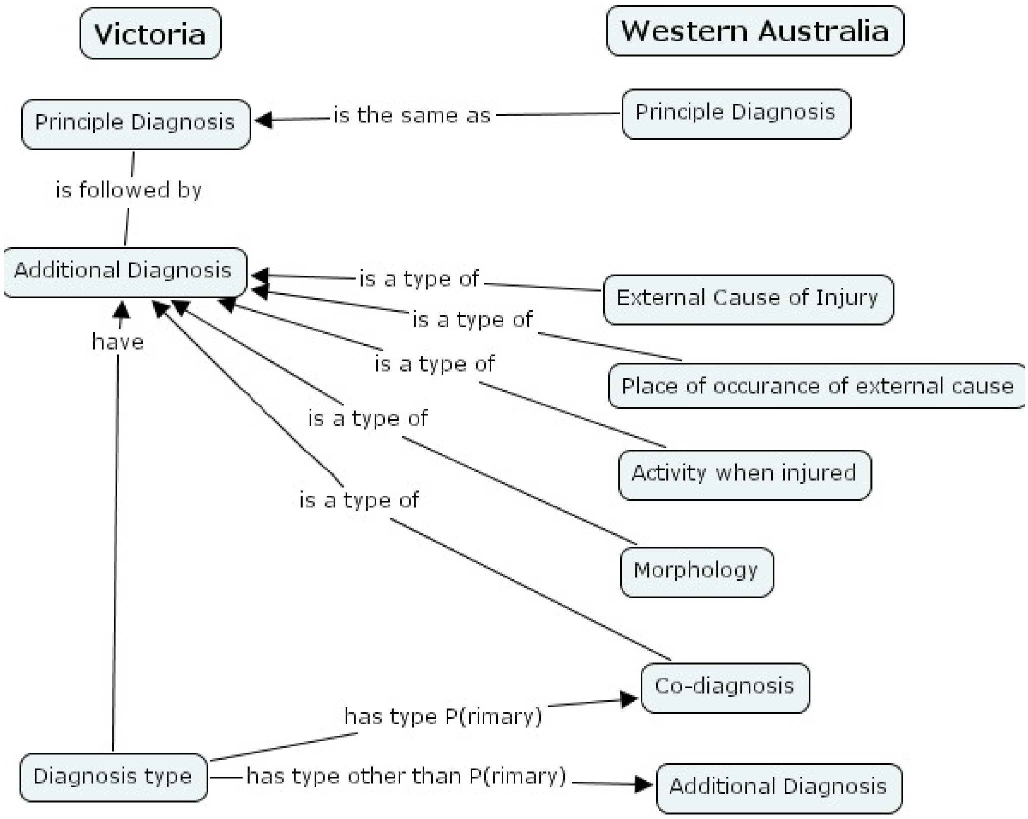

Figure 2-2 Comparison of diagnosis data items between Victoria and Western Australia

When the relationships are drawn between each of the data items it is clear that all concepts included in the Victorian data collection are also available in the Western Australian variation. This figure shows that though the structure of the data collected in each State is different the actual concepts represented are included in both collections.

In Victoria the conditions are entered in the principal diagnosis and additional diagnosis categories with the ability to indicate the type of diagnosis (primary, associated or complication) in a sequence that implies the relationship between items. For example the codes for a patient admitted with a broken arm and a cut leg caused by a car accident on the way to work would appear in the fields and sequence shown in Table 2-6.

Table 2-6 Example of morbidity data in Victorian and Western Australian data structures

Field	Victoria	Western Australia
Principal Diagnosis	Broken Arm (type P)	Broken Arm
Co-Diagnosis		Cut Leg
Additional Diagnosis	Cut Leg (type P)	
Additional Diagnosis	Caused by car accident (Type P)	
Additional Diagnosis	Occurred on public roadway (Type P)	
Additional Diagnosis	While working (Type P)	
External Cause of Injury		Cause by car accident
Place of Occurrence of Injury		Occurred on public roadway
Activity when injured		While working
Additional Diagnosis	Diabetes (Type Associated)	Diabetes

Each system records the same data, but the Western Australian system has provided a structure to make identification of some elements within the collection easier. However the Western Australian system only allows for one external cause of injury and there is an assumption that the cause shown relates to the principal diagnosis, this may not be the case. For example, if the patient above also has a procedure to suture the cut leg and an infection resulted from that surgery the structure of that information in each system is shown in Table 2-7.

Table 2-7: Comparison of data collected in Victoria and Western Australia for a multiple injury case

Field	Victoria	Western Australia
Principal Diagnosis	Broken Arm (type P)	Broken Arm
Co-Diagnosis		Cut Leg
Additional Diagnosis	Cut Leg (type P)	
Additional Diagnosis	Caused by car accident (Type P)	
Additional Diagnosis	Occurred on public roadway (Type P)	
Additional Diagnosis	While working (Type P)	
External Cause of Injury		Cause by car accident
Place of Occurrence of Injury		Occurred on public roadway
Activity when injured		While working
Additional Diagnosis	Diabetes (Type Associated)	Diabetes
Additional Diagnosis	Infected surgical wound (Type C)	Infected surgical wound
Additional Diagnosis	Caused by repair procedure (Type C)	Caused by repair procedure
Procedure	Suture of cut leg	Suture of cut leg

The literature reviewed showed that there has been little evaluation of the suitability of morbidity data structure and content to meet the needs of the users of the data. The National Centre for Classification in Health (NCCH) maintains quality assurance programs and standards for morbidity classification {National Centre for Classification in Health, 2002} which have been developed to improve the accuracy of the code allocation and extraction process. Hovenga identified problems with obtaining and confidently using data from existing data collections {Hovenga, 2002}.

Olsen considers that the quality of data can be measured from the perspectives of "*point of collection*" – the ability of the data to be collected accurately "*relevance to purpose*" - the relevance of the data to specific user needs "*raw versus processed data*" – the potential for errors to be introduced through the modification, extraction and aggregation of the data from the raw original to the processed version. {Olson, 2002}

Quality measures currently undertaken by the NCCH are of the "*raw versus processed data*" type and to a lesser degree commentary on the quality of original data collection at source "*point of collection*" {National Centre for Classification in Health, 2004}. Though this research is not evaluating the quality of ICD-10-AM to purpose it seems clear that the removal of context, such as the relationships between the concepts included in the collection is a significant degradation of the meaning clear when the data was originally collected. The potential of an ontologically sound data representation requires the inclusion of both data entities and the relationships between these entities to gain the advantages of re-usability and improved systems.
Investigation of the morbidity system documentation from each of the State morbidity data collection systems has not provided any indication of a systematic, systems based review of the system requirements for the clinical component of the morbidity collection. In Western Australia in 1997 the hospital morbidity data system was re-engineered to use relational database technology {Information Collection and Management, 2006}. This redesign created relationships between components of identification, event and morbidity data, but did not apply these design techniques to the structure of clinical morbidity data within the collection.

The data content representational forms have changed over the years. In the early definitions of the data collected fields were briefly described and the field size and rules were documented {Grain, 1980}. Today more sophisticated metadata registries are available {AIHW, 2004}. These changes indicate a desire to represent the data accurately and to provide information to support software developers and information users in the use of the data. However, a search of each of the four State and National collection systems provide no evidence of an ontological or systems approach to the structure and content of the clinical data in the collection. The researcher's personal experience of the morbidity developments in Australia since 1979 and in the ongoing data element focused development of morbidity collection since that time, rather than consideration of information structure and ontological impact upon data quality and meaning may be an indication of the governance communities a lack of understanding of the capacity of computer systems to represent these concepts more effectively than the current structure, or a fear of the risks and costs of change.

Having been present at many National discussions on simple data element changes over the last 4 years as a member of the Health Data Standards Committee of the AIHW and as a member of the Victorian Coding Committee for the first 5 years of its existence (1979 – 1984) it appears to me that the resistance to change are based upon, lack of understanding of the problems in the usability of the data collection, lack of knowledge of the potential approaches information science and systems could offer, and a fear of system change, without a realisation that improved ontology based structure of the data is likely to reduce the cost of system maintenance.

Development of an ontological model of the content of the data collection may serve to reduce fear, and encourage greater understanding amongst data managers of the potential of the system, and of the software development community of the opportunity for standardisation.

Though this study does not propose to undertake research into the data collection systems, it uses ontological methodologies to advance the understanding of the morbidity system through greater understanding of the data used to create this system.

2.5 Information System Theory

Information systems theory identifies that successful systems, those that survive implementation, support quality data capture and ease of information use, are those systems designed and built to meet user requirements.{Farre, 2001; Abbott, 2001; Bray, 2002; Alexander, 2002} In 1994 Glass presented research on causes of software cost and time overruns. This research indicated that a principal cause of failure and overrun was the lack of sufficient, considered and user focused system requirements, or clear definition of purpose and processes required of the system {Glass, 1994}. Several years later his paper "*Frequently forgotten fundamental facts about software engineering*" presented even further evidence to support this concept, now often called Glass' Law {Endres, 2003}.

The literature supports the concept of detailed analysis both of user requirements and of the information system and content to support those requirements as a mechanism for quality system development {Farre, 2001}. The evidence particularly supports this approach in the design of data intensive systems which can lead to the development not only of more cost effective systems, but also to systems that are more flexible and cheaper to maintain {Freitas, 2002; Olson, 2002; Harte-Hanks Trillium Software, 2006}.

Computer science has used the concept of models to assist in the understanding of healthcare systems since the 1980's. Frean explains that modelling is used to visualise, specify the structure of, provide a template for development of, and to document decisions made about information systems {Frean, 2006}.

In 1999 Booch et al developed four principles of modelling. These are:

- "*The choice of what models to create has a profound influence on how a problem is attacked and how a solution is shaped.*
- *Every model may be expressed at different levels of precision*
- *The best models are connected to reality*
- *No single model is sufficient*…."{Booch G. et al, 1999}.

Satzinger, Jackson and Burd demonstrate that "*the process of creating a model helps an analyst clarify and refine requirements and design details*" {Satzinger, 2005}. There are different types of models used for different purposes. Information models and data models are forms concerned with modelling data {Conrick, 2006; NHIMG, 2003}, as is required in this research.

The relationship between data and systems is often shown in architecture maps, including those based upon ontological concepts. This is critical as the system and the data are dependent upon each other to support collection and use of the information they support {Beale, 2001}. Architectures describe the global view of how the 'building' and its components such as electricity, water, bricks, and doors will fit together. The type of model concerned in this research takes a high level perspective on a detailed field of data.

Australia's National Health Information Model provides an organisational structure for National metadata. This research looks at a single component of that metadata, the clinical information of the morbidity data collection. Figure 2-4 shows the event section of the information model. This diagram shows that events occur to people and that there are different types of events. The modelling convention showing a 'box within a box' indicates that the inside boxes hold components or types of the outer box.

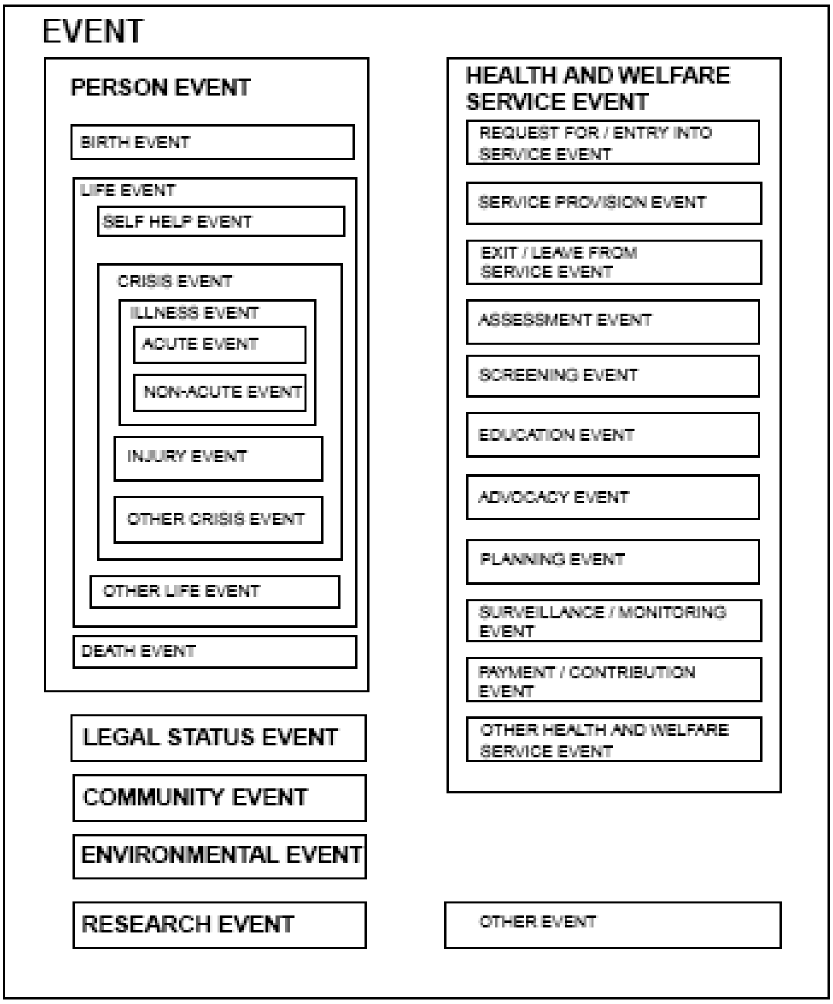

Figure 2-3 Event entities of the National Health Information Model

Models are a purpose specific representation of a view of the real world. Where the addition of the relationships between the concepts in that real world are included you offer an ontological perspective to that representation within the model. An ontological view is more robust and offers greater utility in system development and data re-usability {Campbell, 1998; Chen, 2004; Gangemi, 2006; Roman, 2006; Szirbik, 2006}. The use of a model to represent the entities and relationships in morbidity data is consistent with the use of models in ontology, computer science and health informatics, the literature review proceeded to investigate ontology based methodologies for the development of a data model.

2.6 Ontological Sciences

Literature on the development of ontological models shows that they have been used to represent concepts and the relationships between them. Wand and Wang proposed that data quality needs to be addressed in the context of system design and that ontological concepts that model the reality of the data being captured and improve the capacity to re-use and interpret that information {Wand, 1996} In 2006 Gengemi et al indicated that '*there is a need for evaluation methodologies in the field of ontology development*' {Gangemi, 2006} as there is no accepted international approach to this issue. Ontological approaches to information systems have developed significantly as a mechanism for knowledge engineering, particularly in healthcare and the internet. It is even described as having a similarity to domain modelling {Gangemi, 2006}.

Ontologies can be as simple as a model to describe a hierarchy of concepts related by subsumption relationships. Subsumption incorporates a concept under a more general category and defines how that concept differs from the general category using constraints {Guarino, 2001}. For example: Morbidity includes concepts called diagnoses, which include concepts such as injury.

Beale has described ontologies as a mechanism to support:

- *"computer-based reasoning on facts….*
- *Aggregation, search and retrieval of data from diverse original source systems, which necessitates rationalization of the original data"* {Beale, 2007}

High level ontologies such as the concepts incorporated in admitted patient morbidity reporting can be represented using Entity-Relationship Models (ERM). This modelling approach originated in the 1976's when Chen proposed a model to incorporate "semantic *information about the real world*" using a diagrammatic technique to improve data integrity, information retrieval and data manipulation {Chen, 1976}. This approach is still "*one of the most widely used software engineering techniques*" {Jimenez, 2006}.

The steps in ontology development used in the ontology development tool Protégé have been described by Noy and McGuinnes as:

- *"Determine the domain and scope of the ontology…*
- *Consider re-using existing ontologies…*
- *Enumerate important terms in the ontology…*
- *Define the classes and the class hierarchy. …*
    - *A top down development process starts with the definition of the most general concepts in the domain and subsequent specialisation of the concepts.*
    - *A bottom up development process starts with the definition of the most specific classes, the leaves of the hierarchy, with subsequent grouping of these classes into more general concepts.*
    - *A combination development process is a combination of the top-down and bottom-up approaches." Where more salient concepts are described first and then generalized and specialized as appropriate*
- *"Define the properties of the classes……*
- *"Define the facets of the classes*"{Noy, 2001}.

This methodology forms the basis of Protégé, a common ontology tool used in healthcare and in other environments.

The findings of this literature review combined with the need to meet the requirements of this project resulted in the adoption of an Entity Relationship approach as this was found to be the most appropriate although the development process itself needs to be influenced by the ontology development process used in ontology tools. This approach was further informed by a review of the methodologies used to represent health data.

2.7 Ontologies in Healthcare

This literature review has investigated ontologies developed in healthcare to identify use and methodological development. Despite investigating a wide range of approaches to this technique it was clear that there is little standardisation in the terms used to describe the process of ontology development, nor in the methodologies used. The literature was then investigated to identify existing health ontologies and methodologies that could be used to inform the methodological process for development of morbidity data.

2.8 Electronic Health Records Structures and Ontologies

The Australian electronic health record initiative, HealthConnect implementation strategy identified a wide range of potential secondary uses of electronic health record data {HealthConnect, 2005; Australian Health Information Council, 2008; NHIMAC, 2001}

Achievement of the desired outcomes for EHR systems requires a standard approach to both the information architecture and the terminology used to store information in that architecture. Without these components the data in systems are not able to be computer processed, or safely and consistently communicated to different health care professionals. This is the requirement for interoperability {The National Office for the Information Economy, 2003; Information Communications Technology Standards Committee, 2004; Beale, 2006}.

To achieve semantic interoperability, ontological models have been developed to represent different components of the EHR infrastructure. Review of existing ontologically based models was undertaken to identify potential bases for the development of an ontology model. International health informatics literature, particularly the Journal of the American Medical Informatics Association and the e-Journal of Health Informatics were reviewed in addition to using the PubMed search engine and MedLine as sources of models of health data that are at a high level and deal with clinical data concepts. The literature review also included health data standards produced through public and expert review by the peak health informatics standards organisations HL7, ISO and CEN. To ensure that E-Health initiatives were included websites representing government e-Health and EHR initiatives, specifically the UK, Canada, USA and Australia were included. This review highlighted four principal ontological approaches. These were:

- HL7 Clinical Document Architecture Release 2 (CDA2)– is based upon HL7 Reference Information Model, the CDA is an exchange model for clinical documents{HL7, 2007; HL7 Australia, 2007}. Though the definition of documents has proven useful, it is not used widely in the implementation of EHR systems at this time. Australia has been investigating the use of this approach to support e-referral documentation. Though there are similar concepts communicated in these documents the extent is built around the document alone. This approach was relevant to morbidity data collection to support its transmission from hospitals to State authorities and to national bodies, but more importantly concepts such as CDA can be used to retrieve information from data collections using standardised query formats.

- OpenEHR Archetypes that link to the OpenEHR information model as detailed previously (Section 1-6)- representing the generic structure of the clinical record{OpenEHR, 2007} supporting the maintenance of clinical meaning and information structure over time. The use of this ontological model is gaining credence, having been trialled to support implementation of EHR functions in the UK it is increasingly being seen as a component of the interoperability solution for implementation of EHRs.
- SNOMED-CT– a clinical terminology representing medical knowledge and concepts{IHTSDO, 2007}. This is Australia's chosen terminology for healthcare.
- ISO/TC215 – Conceptual framework for patient findings and clinical practice.{ISO, 2007b} an ontology based categorial model of clinical information for direct patient care. This work was selected for review as the concepts incorporated in this model are generic, being built to fit the concepts of clinical care from the broadest perspective. The concepts included are similar in many ways to those included in the morbidity data collections.

I have chosen not to investigate CDA as a potential ontology for the purpose of representing morbidity data. CDA was rejected for this research as it is difficult to retrieve the required concepts from CDA and is not designed for use in generic situates such as this one.

2.9 OpenEHR Archetypes

OpenEHR Archetypes are based upon an ontologically sound information model for electronic health records. They support computerisation and have been developed to represent data at various levels within the clinical environment and offer a computer implementable infrastructure of metadata. OpenEHR Archetypes provide a computable highly generic structure designed to represent clinical data requirements specified over more than 15 years. The origin of these archetypes was in representation of day to day clinical data to support direct patient care, but has extended, through the broader requirements of health care to support public infrastructure, even veterinary science {Beale, 2007}.

The methodology for development of OpenEHRs has been strongly grounded in clinical knowledge representation and in defined user requirements and had significant review by experts in the field as it represents the underpinnings of the ISO standard EHR reference model.{ISO, 2007a}.

This ontology represents the domain of the individual patient to support the care of that individual. The structure of this ontology is based upon 8 high level archetype structures, these represent the information model for clinical practice. These are:

- *"The Composition, which includes administrative information as well as clinically relevant components such as dated, signed 'event' information, that may represent concepts such as 'discharge', 'encounter', 'past history', 'report'. This functions as a grouping mechanism, to which any of the other components may be linked.*
- *The Section, which includes summaries of health information such as operation notes, problem lists, appointments, clinical findings and conclusions*
- *Structure, grouped information representing concepts such as 'patient demographics', imaging data, laboratory test request, Medication description*
- *Cluster, information 'sets' which have meaning when expressed together, but not apart such as menstrual cycle, gait, relative size, symptoms*
- *Action, actions that occur to data such as referral, procedure undertaken, medication action*
- *Evaluation, represents clinical opinion such as, differential diagnosis, medical directive, adverse reaction or alert, goal, injury*
- *Instruction to another to take action, including laboratory request, monitoring, procedure instruction,*
- *Observations made, including laboratory result, Glasgow Coma scale, heart rate, height"* {OpenEHR Foundation, 2007}

These groups and their defined composite elements will be reviewed as part of the methodology of this research to identify their relevance and potential relationship to morbidity ontology. The scope of this model is much broader than that of morbidity reporting and may include the structures and concepts necessary to represent that data clearly.

2.10 SNOMED-CT

The Systematised Nomenclature of Medicine – clinical terms (SNOMED-CT) has evolved as an ontology of healthcare generated by collaboration between the College of American Pathologists and the Read Codes from the UK. SNOMED-CT aims to represent all health concepts in a manner that supports computerisation of these concepts {Walker, 2004}. SNOMED-CT has been developed using an ontological approach where all elements of the clinical world have been categorised into a top level hierarchy.

SNOMED-CT represents the domain of clinical medicine as a discipline of knowledge. These concepts can be applied to describe clinical information about an individual. Each concept in SNOMED-CT is represented by a unique identifier and may have one or more descriptions associated with that identifier.

Each concept may also have relationships that define and describe the concept. There are 18 hierarchies of concepts. These are shown in Table 2-8. All concepts in SNOMED-CT fall into one of these categories.

Table 2-8 SNOMED-CT hierarchy of Concepts{IHTSDO, 2007}

Hierarchy of Concepts
substance
linkage concept
specimen
body structure
situation with explicit context
staging and scales
physical object
event
environment or geographical location
qualifier value
organism
special concept
pharmaceutical / biological product
clinical finding
record artefact
social context
procedure
physical force

Defining semantic relationships are used to indicate relationships between concepts in different hierarchical branches to define these concepts in computer processable ways. Every SNOMED-CT concept has at least one IS_A relationship that defines that concept with a place in the hierarchy

For example:

Cough IS_A

functional finding of respiratory tract

IS_A respiratory finding

IS_A finding by site

IS_A clinical finding {IHTSDO, 2007}

The last entry in this structure is one of the 18 main hierarchy structures of the terminology.

Other relationships include:

Attribute relationships are relationships that are constant characteristics of the concept. For example:

"Lumbar discitis (disorder) (a concept in the Clinical finding hierarchy) is related to concepts in the Body structure hierarchy through two attributes: FINDING SITE and ASSOCIATED MORPHOLOGY.

Lumbar discitis (disorder)
FINDING SITE Structure of lumbar intervertebral disc (body structure)
ASSOCIATED MORPHOLOGY Inflammation (morphologic abnormality)

The two attributes FINDING SITE and ASSOCIATED MORPHOLOGY and their assigned values provide definition for the concept Lumbar discitis (disorder)" {IHTSDO, 2007}.

The morbidity data collection is particularly concerned with diseases. Diseases are represented in SNOMED-CT in the 'Clinical Finding' In order to more specifically analyse that component ontology of morbidity data the clinical finding hierarchy type has been subdivided to the next level down in that hierarchy (Table 2-9).

Table 2-9 SNOMED-CT hierarchy Level 2 for clinical findings

Description
clinical finding - stage finding
clinical finding - administrative status
clinical finding - adverse incident outcome
clinical finding - clinical history and observation
clinical finding – deformity
clinical finding – disease
clinical finding - drug action
clinical finding - effect of exposure to physical force
clinical finding - finding by method
clinical finding - finding by site
clinical finding - finding of grade
clinical finding - finding related to physiologic substance
clinical finding - finding reported by subject or history provider
clinical finding - general clinical state finding
clinical finding - neurological find
clinical finding – oedema
clinical finding - prognosis/outlook finding
clinical finding - sequelae of external causes and disorders
clinical finding – wound finding

2.11 Conceptual Framework for Patient Findings and Problems

The third structure for patient care based clinical information representation is the Conceptual Framework for Patient Findings and Problems. This document specifies a categorial structure for patient findings and problems, a high level patient domain ontology for broad use within clinical care in computerised systems. The methodology used to develop this framework is not clearly defined in the document. However the diagrammatic representation uses Entity Relationship Model representations. The document includes a description of each concept in the framework, and semantic links between the concepts. The resultant framework allows any clinical finding to have the following semantic relationships:

HasAbnormalAnatomicalLocation
HasAnatomicalSite
HasStructuralEmbryologicalDefect
HasOnset
HasEpisodicity
HasCourse
HasOccurrence
HasMorphology
HasCausativeAgent
HasSeverity
HasStage
HasPathologicalProcess
HasExternalCause
HasPlaceOfOccurrence
HasSubjectOfInformation
HasQuantity
HasPsychosocialAspect
HasAssociatedFinding
HasAssociatedFunction
HasInterpretation
HasAssociatedTest{ISO, 2007b}

Though, like SNOMED-CT and OpenEHR these concepts are built for the clinical care domain they offer a set of relationships and concepts that have similarities with the morbidity data collection and were examined for suitability when developing the morbidity ontological model.

2.12 Literature Review Conclusions

This literature review has highlighted the fact that a morbidity data structure has not been analysed to identify ontological structures that would support more effective development of data capture and representation mechanisms for this valuable data. The instructions provided in the State and National metadata and user manuals provide information on the business rules and information content requirements for morbidity data that will be useful in the development of an ontological model.

These development techniques lead to systems where data and systems are re-usable, flexible and cheaper to maintain. In the environment where Australia is moving to EHRs there is the opportunity for change in the data collection structure that could realise the benefits that a sound ontology based model can provide.

The literature review indicated that an ontologically sound model for morbidity data has the potential to optimise data retrieval and usage and support the development of more flexible information systems. Entity-relationship representations offer a sound methodology for the development of an ontological model. The existing clinical information ontology model for patient findings and problems was found to be a sound source upon which to develop and define the morbidity data structure.

The documentation of morbidity data collections indicates that there has been little change in the structure or content of morbidity data over time and that there has not been an ontological approach to the collection structure of this data, though ontological review of the classification system used to record the data is occurring at an international level.

Chapter 3 Patient Data vs. Clinical Domains

The literature review identified different ontologies and the importance of understanding the domain being described by an ontology. The recognition that each model is a picture of the world looking through a different window (for a different purpose) is a critical issue that affects re-use of the data represented by that model.

Morbidity data represents a view of an individual person's health status and care at a given point in time. This data has the patient and their episode of care as the proximal domain. ICD-10-AM which is used to describe that person's health status and care is a representation of the medical clinical domain in general and only gains meaning in morbidity when used in relation to an individual. ICD-10-AM includes the representation of procedures an interventions, one of the features that differentiate it from ICD-10 as produced by WHO.

Figure 3-1 shows the resources used in this research and their relationship to the different ontological domains relevant to morbidity data collection. This diagram shows the OpenEHR archetypes and SNOMED-CT used in the domain of direct patient care, to support the EHR and the deliverables expected of the EHR, while the morbidity data collection is an abstraction of that data for an individual in a similar way to the concept of ICD-10-AM is a purpose specific aggregation to represent the general domain of medical/clinical knowledge. The resources used to develop the model in this research are demonstrated to indicate where they provide governance to both the patient based morbidity data and the domain of concept representation provided by ICD-10-AM. The Categorial Framework for Patient Findings is a model that represents patient data but at a higher level than the OpenEHR archetypes.

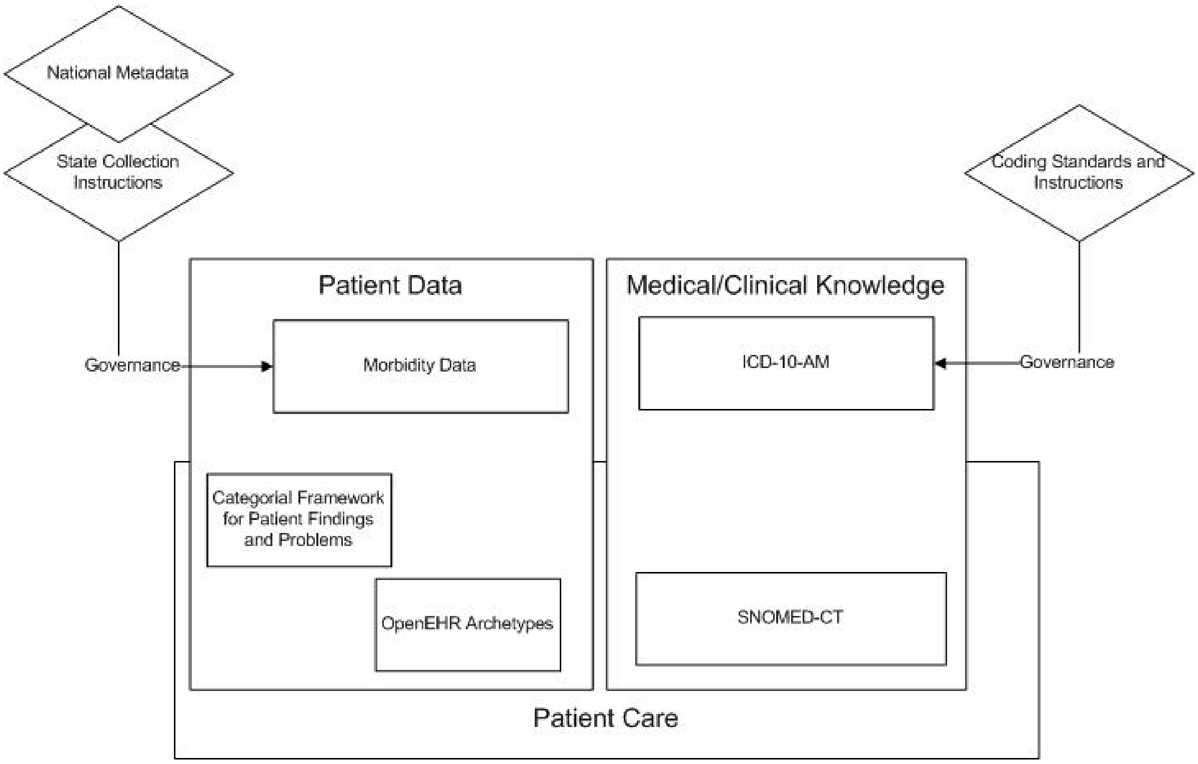

Figure 3-1 The ontology domains of this research

3.1 OpenEHR

Beale and Heard developed an ontology-based model of clinical information "*designed to make health information systems properly interoperable and safely computable"* {Beale, 2007}. This model identifies the ontological components of clinical information to be observation, opinion, instruction, action and administrative event. Though these are different to the reporting ontology of morbidity data, they are the logical source of that data. The methodology is based upon the identification of high level categories for the information collected and recorded in a clinical record.

The **open**EHR RM formalises the Electronic Health Record in terms of:

"Information model (RM)

This is the model that describes the health record itself - not the clinical data that are contained within it. The reference model deals with containers such as Folders and Compositions. Compositions are a broader concept than documents - but include documents. Examples of Compositions are an ECG report, a progress note, a laboratory report and a referral. The Composition is the minimum unit of communication and committal to the EHR. The ***open****EHR Reference Model specifications are available from; .*

Archetype Model (AM)

Archetypes are descriptions of valid Entries, Sections and Compositions. These are expressed in a formal manner which enables them to be shared between systems. A blood pressure archetype represents a description of all the information a clinician might want to report about a blood pressure measurement, and may include some aspects which are mandatory. A 'SOAP' archetype describes the sections of a problem oriented health note and which entries are valid under each section - for example only diagnoses may be allowed under the 'A' section. Templates are logical models of user forms - and are described in terms of choices of archetypes whose data are captured on a particular form. See the . The ***open****EHR Archetype Model specifications are available from ; .*

Service Model (SM)

This is the computational viewpoint of the ***open****EHR architecture. The service model consists of service definitions for the major services in the EHR computing environment. These are largely derived from existing work in OMG Corbamed, CEN HISA and implementation experience.* {OpenEHR, 2007}

The OpenEHR Archetype is an ontologically sound, implementable representation of this model, providing a standardised, computable EHR structure {Beale, 2006; OpenEHR Foundation, 2007}. This structure provides context, but also business rules to govern the terminological or classification based representation of that context. These archetypes offered a base for comparison and testing of the morbidity ontology.

Figure 1-2 represents the OpenEHR as the central component from which extracts can be made to suit a range of other information models built for other purposes. The domain of admitted patient morbidity clinical data does not have a similar model to inform system development. However morbidity data is extracted from the patient's record, and is therefore dependent upon the content and structure of that record. The OpenEHR approach provides a sound ontological model for the record and as such this research investigated the capacity of this ontological model to provide a methodological and ontological base for morbidity data structure.

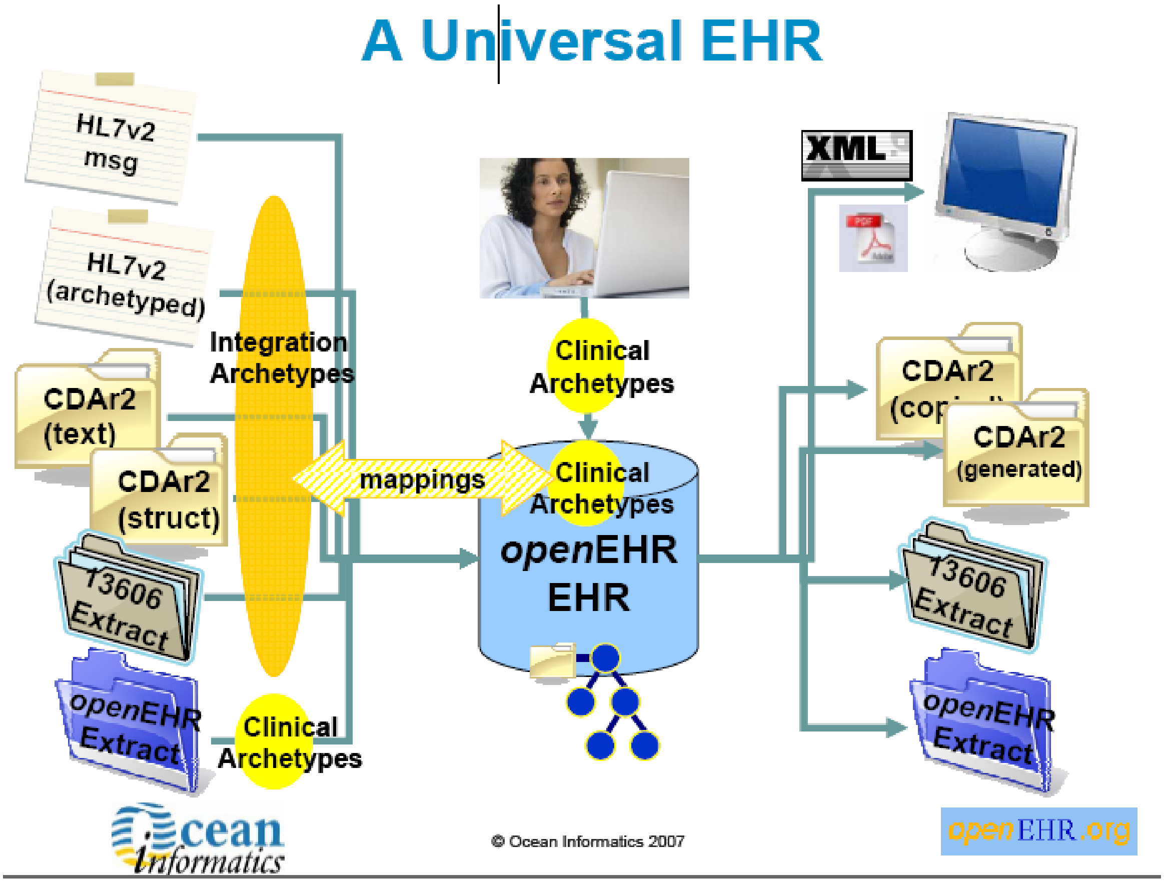

Figure 3-2 OpenEHR Archetypes in the EHR universe(Source www.openEHR.org)

3.2 SNOMED-CT

The other significant ontology used in health care to describe clinical concepts including diseases and procedures (representing reality) is the Standardised Nomenclature of Medicine (SNOMED). SNOMED-CT is considered the terminology most suited to represent clinical concepts for patient care. SNOMED-CT is governed by the International Health

Terminology Standards Development Organisation (IHTSDO) established in 2006, an organisation of which Australia is a charter member. SNOMED-CT content is structured to represent individual concepts (things) so that each thing fits into the categories to which it 'belongs'. In this way a computer is able to identify that the procedure Appendicectomy:

"*is an operation on the appendix"*

"Is a partial excision of the large intestine"

"Has a method of excision on"

"Procedure site - appendix structure"{IHTSDO, 2007}

These concepts are structured in two ways, a hierarchical representation (categories or grouping) and attributes (relationship types). Together this allows each individual concept to appear in more than one hierarchy. For example appendicectomy is related to the body structure (appendix structure) as well as being a procedure. These categories of SNOMED-CT hierarchies are listed in table 2-8 and 2-9 in Chapter 2.

Though SNOMED-CT is designed to represent the domain of clinical information, and morbidity data is a specific abstraction of that purpose this structure was considered as a mechanism to represent the entities and relationships required for reporting of morbidity data.

The electronic health record environment requires standardised clinical terminologies such as SNOMED-CT to represent clinical ideas (reality). They also require an ontologically sound, standardised structure to the information represented by the terminologies, for example a structural context such as personal history of 'term (asthma)' or current problem of 'term (asthma)'.

These developments represent an ongoing progression of information technologies capacity to represent information more accurately and in a manner which is more flexible and suited to the needs of healthcare.

3.3 *The Categorial Framework for Patient Findings and Problems*

This ontology represents the patient domain of healthcare at an abstraction of clinical information. The representation of patient findings and problems can be expressed using SNOMED-CT to define, for example the content of the clinical finding component. SNOMED-CT relationships include many similar concepts to those shown in this ontology.

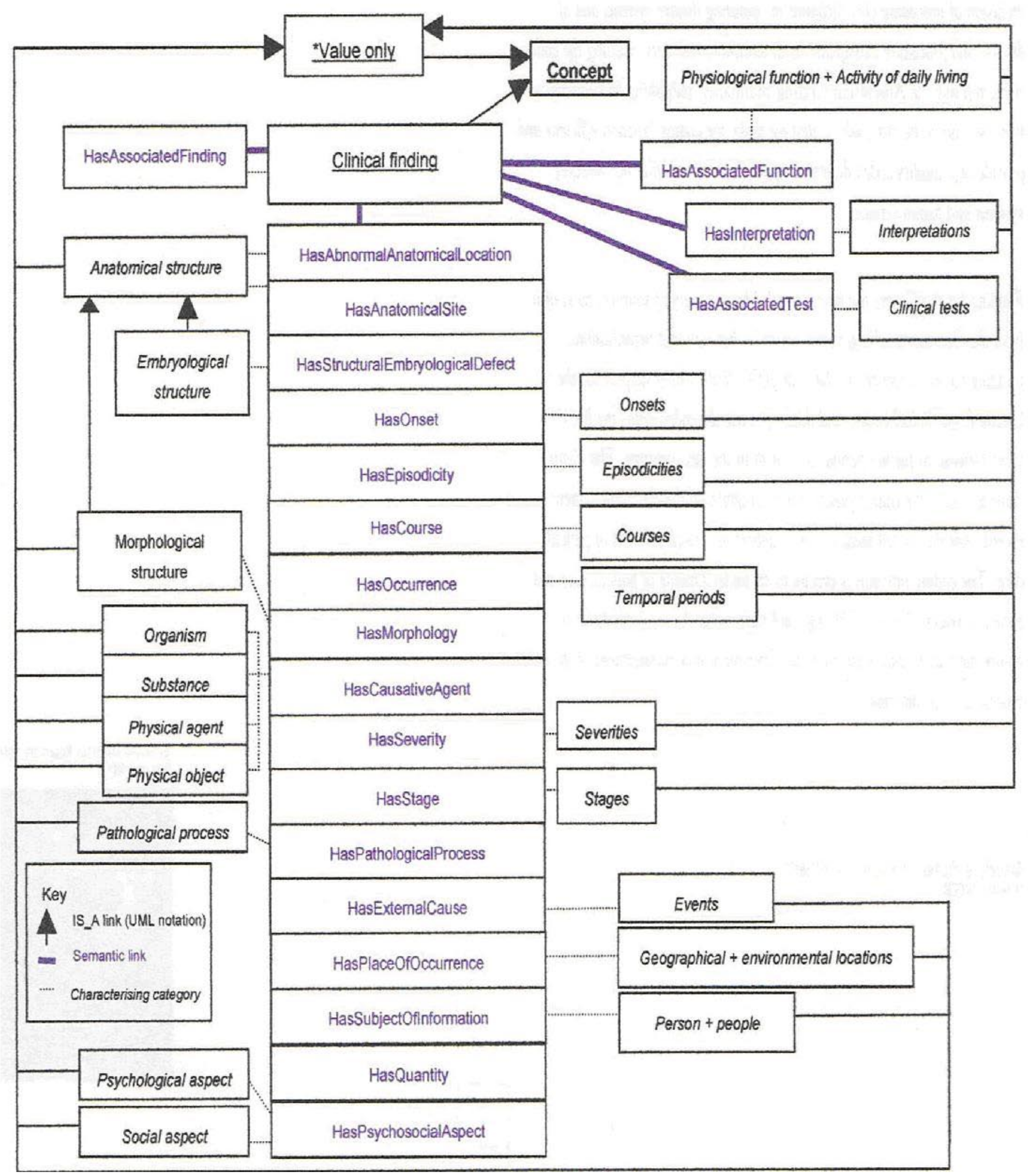

Figure 3-3 conceptual framework for patient findings and problems (ISO, 2007b)

Figure 3.3 shows the ontological representation of the concepts included in patient findings and problems developed by ISO {ISO 2007b}. Each of the models examined indicate relationships and concepts used to describe patient situation and care. Each represent the data differently as defined by the purpose for which the model has been constructed. This model represents clinical data for clinical use. Many of the components found in the diagnosis or procedures included in morbidity data reporting are represented here. Clinical findings may have a context that defines them as a principal diagnosis and some morphological structures are also included in these data collections. The issue of interest is that there is no directly mappable route from one form of viewing the data (clinical practice) to another (morbidity reporting).

Chapter 4 Methodology

Ontological engineering is the identification of concepts in the reality of a domain and the relationships between those concepts {Gasevic, 2005}. This approach was chosen to be used to investigate and identify the entities and relationships required to represent inpatient clinical morbidity data and provide a sound foundation for system development and information retrieval flexibility and offers an approach consistent with other data modelling undertaken in health care systems.

Ontology engineering was found to be the most appropriate methodology as this was seen as the best possible way to enable the stated research objectives to be met. This methodology included 1) a review of existing data collection contents and formats, and 2) detailing the inclusion and exclusion instructions provided in ICD-10-AM, the classification system used to describe clinical information in the collection. An entity modelling technique was then chosen to develop an information model. The data analysed included 12 months of Victorian Morbidity data consisting of 1,311,678 cases. The majority of cases in the data collection have 1 diagnosis (328,923 cases (25% of all cases)), or 1 diagnosis with 1 procedure(105,750 cases (8% of all cases). These simple cases do not require relationships as there are single concepts involved in the data collected. In order to identify cases that represent more complex relationships. The data collection was examined to find patterns that represent this complexity. 5 cases represent the most complex patterns of data recorded in the data for the year. Complexity was defined as where there are a large number of diagnosis and/or procedures represented, the assumption being that the larger number of concepts included would offer more potential to identify relationships present between these concepts. The literature review identified that there is no standardised approach to the methodology of ontology development. However the process offers a well established and suitable mechanism that, when combined with the information collection processes identified by Noy and McGuinness{Noy, 2001} has the advantage of using existing clinical ontology based models to inform the entity-relationship model.

Fig 4-1 outlines the methodology, including the relationships to existing methodologies. This diagram highlights the iterative process of testing which needs to be adopted to update the representation of entities in the model.

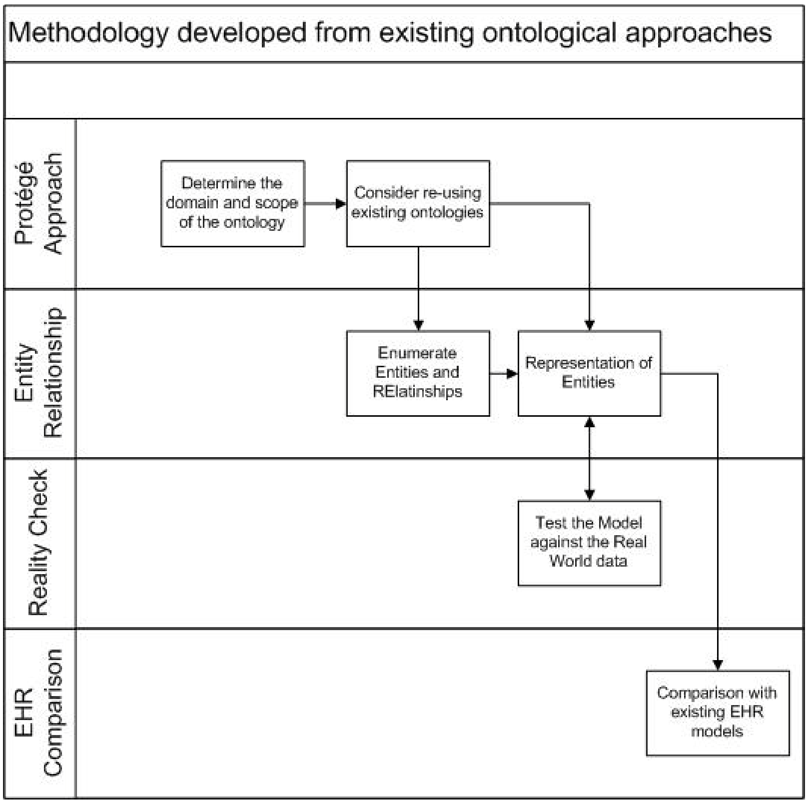

Figure 4-1 Methodology outline

Entity-relationship development methodology requires a thorough investigation of existing and potential entities in the data collection and the identification of relationships. Ontology development requires a precise and consistent approach to the representation of entities and relationships within a specified domain. The Protégé approach to ontology development offered a sound methodology requiring the a model built upon existing knowledge, rather than in a vacuum recognising only the single domain of inpatient morbidity data collection, rather than existing knowledge available about the domain from which the data is extracted, the clinical domain. For this reason the Protégé approach to clearly define the scope of the ontology and to consider existing entities and relationships was undertaken as the first step in this research.

The Protégé methodology suggested that there were alternative approaches to this activity. The enumeration could begin at the top, with general concepts in the domain and build increasing specificity; or where significant details are available the ontology can be developed from the bottom up, or a mixture of these approaches can be used.

This research took a top down approach. The reason for this decision was based upon the literature review's identification of high level concepts within the existing morbidity data collection and the fact that some of these components were also present in The Categorial Framework for Patient Findings and Problems{ISO, 2007b}. The only concepts present in both the Framework and the morbidity data collection are listed in Table 4-1 and were used as the starting point for the analysis. It should be noted that the data element Activity is not included in this list, as it is not present in The Framework used as the starting point. This research identified additional entities such as this one and, using relationships the appropriate place in the entity-relationship model was identified for each entity.

Table 4-1 High Level Entities from the Framework

Entity
Clinical Finding
Cause of Injury
Place of Occurrence

The purpose of this model was to clearly identify the component entities within the clinical morbidity data collection and the relationships between them, with a view to providing the health data and morbidity data user community with a greater understanding of the data in the collection to support further development of the systems involved in the collection and use of the data. This objective requires a representational form that is meaningful to this broad community. The research did not require the building of an actual system, but the development of a generic, high level representation. The use of entity-relationship models are proven suitable methodologies to meet these objectives, while modelling techniques such as UML modelling are more detailed than required for this purpose.
Though the concepts identified in the documentation were considered a sound basis for the development of the entity-relationship model, ontology requires that the model represent

reality. The reality of morbidity data collection is the data actually recorded in a morbidity data collection's clinical data fields. A reality check component was therefore included in the methodology, using 5 use cases designed to obtain a range of different patterns of data from the real world. Though it is acknowledged that this is not a thorough test of the real world, it offered a mechanism to identify any issues with the model based upon a comparison to the real world. At the conclusion of this stage of the methodology the entity-relationship model was considered to be complete.

The research objective was to understand the clinical admitted episode morbidity entity-relationship model in the context of ontological structures used to represent clinical data in EHRs. A comparison of the entities and relationships in the SNOMED-CT, OpenEHR Information Model and the Framework for Patient Findings and Problems was undertaken. Models outside the domain of clinical care such as the HL7 RIM, were not included.

4.1 Determine the domain and scope of the ontology

Ontology development requires a careful determination of the scope of the domain concerned, including the identification of the level of detail (granularity) to be included. This research investigated the domain of entities and their relationships in the clinical components of morbidity data reporting of admitted episodes of care in Australia. Clinical information is recorded in the disease and procedure data elements of the existing data collection. These are the data elements represented using ICD-10-AM {METeOR, 2007}. The ontology represents any concepts (entities and relationships) that can be recorded in the clinical components of morbidity data collection as they are identified in National metadata, system manuals and ICD-10-AM code books and coding standards.

This research considered the generic or high level categories of information included in the 2005-2007 morbidity data collection, as demonstrated by National Metadata, State collection instructions and coding instructions applicable from 2005 to the present.

The literature review identified that some of the collections divide the disease concept into different data elements:

- Principal Disease

- Additional Disease/s
- Cause of Injury
- Place of Occurrence
- Activity when injured

These components represent the entities used to start the research, providing a basis for the structure of the data.

4.2 Consider re-using existing ontologies

Ontological development theory suggests that the use of an existing ontological model can improve the resultant product depending on the purpose and granularity of the data. The existing ontology needed to have similarities to the domain of the investigation or to the purpose of the system involved. The literature review identified three potential approaches;

- the OpenEHR Reference Model is the ontology within which archetypes occur. Each archetype is a constraint model about a specific knowledge object where an individual object needs to be represented by a value domain or 'labelled' using unique concepts from a terminology. This ontological model offers a consistent structure for the EHR, particularly the clinical data, and was therefore considered appropriate for comparison.
- SNOMED-CT, a terminological ontology representing the details of clinical knowledge as they relate to a given individual's record, this model is significantly more detailed than that required for the purpose of this research, though the high level structures did provide a useful comparison and
- the Conceptual Framework for Patient Findings and Problems{ISO, 2007b} (in future referred to as **the Framework**), a high level conceptual model of clinical information. This model includes some of the entities and relationships included in morbidity data that were identified in the literature review. The content is therefore similar in scope and level of detail. This information is presented as an entity-relationship diagram.

For example: The Framework identifies morphological structures, rather than the specific type of morphology such as 'basal cell carcinoma', where the concept 'morphology' is a high level concept and the representation ' basal cell carcinoma' is represented by the specific coded value recorded in the system (which could be represented by a code in SNOMED-CT or as a code in a specific archetype in the OpenEHR model.

The Framework provides clear definitions of the components of an entity-relationships model, using ontological terms and providing a set of entities and relationships as a starting point for development of the morbidity model. This ISO Technical Specification also provides the methodology for the representation of relationships that has been used in this study.

This approach therefore supports the scope of the research in its intent to produce a generic approach to the content and structure of morbidity reporting.

4.3 Enumerate entities and relationships

Having established the domain scope and a starting set of entities and relationships the methodology requires the analysis of the governance instructions for acute episode clinical information for morbidity data in Australia in order to confirm the entities and relationships in these data.

Governance instructions were found at the National level as represented by METeOR (the National metadata registry) and at the State level through user manuals for admitted episode clinical morbidity data for collections in Victoria, New South Wales, Queensland and Western Australia. Instructions provided at the point of capture are represented by the instructions in the ICD-10-AM coding system.

A detailed review of the metadata and governance instructions for data collection was undertaken to identify entities and the semantic relationships between those entities in order to develop a model of the information in hospital admitted episode morbidity data collection (RQ2).

Important terms required in the ontology were identified through a detailed analysis of the categories of information included in disease and procedure data elements. A spreadsheet of each entity was established and an associated spreadsheet of relationships was used to collect the information, and to record an example of the use of each entity or relationship from the governance or real data. The Testing the Ontology Model section provides samples of these tools.

These are constrained and identified by:

- Identification of the data elements required for the National collection for clinical data through online search of METeOR for the Admitted Episode Minimum Data Set Specification. This provided details of all data elements included in the National collection. The detailed entry for each of these data elements was reviewed and where the value domain was listed as ICD-10-AM the data element was considered to represent an entity in the data collection. The instructions provided to guide the collection of each of these data elements were investigated to identify any rules or instructions that indicate relationships. As this analysis was undertaken any additions or modifications required to accurately represent the entities and relationships found were made to the list of entities and relationships identified from The Framework.
- Identification of data elements and relationships indicated by instructions for data collection from the State data collection specifications for disease, injury and procedure information. Each of the Victorian, New South Wales, Queensland and Western Australia admitted episode user manuals were searched and each data component collected using ICD-10-AM were identified and their definitions compared. Where there was more than one definition during the time period 2005 – 2008 these were each included and compared to identify whether the variation represented any variation in entities or relationships. This process supports the identification of entities and relationships to further develop the entity-relationship model. Additional or modified entities or relationships were included in the list of entities and relationships.
- Identification of the relationships/structures that are represented by the coding instructions held in ICD-10-AM. The ICD-10-AM tabular lists have coding instructions under the “includes” and “excludes” instructions throughout the text. The e-book of ICD-10-AM (NCCH, 2006) version 4 has a search facility which was used to find every entry in the tabular lists that indicated inclusion or exclusion instructions. These

instructions were recorded. Additions or modifications to entities were made, and relationships, identified through instructions in the code system were classified using the ontological approach based upon The Framework.

During this process of identifying entities, variations between the State and National collections, either in data content, definitions or instructions were documented for later discussion answering RQ1.4.

The concepts discovered were then analysed using an example of each entity/ relationship to represent the knowledge or rule in the real world. Where the concept could be represented in the real world of morbidity reporting as an individual concept it was retained or added to identify the entities and their relationships needed to accurately represent the information in the morbidity data collection and to support and enhance the usability of the information collected.

4.4 Representation of the Entities

The representation of the entities to be included in the morbidity ontology model was the core activity within the application of the chosen methodology for this research. Having enumerated the entities and relationships these were then analysed whilst considering semantic differences and variations, using concepts in the Framework where possible, and adding / modifying these to meet the requirements identified through an enumeration process.

The resultant entities and relationships were then developed into an entity relationship model that represents all unique concepts and relationships found in the enumeration process. The resultant model (Figure 4-1) and supporting documentation represent the entities and relationships identified by the analysis undertaken to this point.

This stage of the methodology implementation provided the information model sought to represent admitted patient clinical morbidity data and thereby answered the second research question:

- What is the entity relationship model and the semantic relationships that represents the data collection.

4.5 Sample

All instructions for the coding of diseases and procedures were reviewed through the tabular section of ICD-10-AM {National Centre for Classification in Health, 2004}.

State instruction manuals were reviewed for the three largest States, Victoria, New South Wales, and Queensland and for the only State with a history of health data warehousing – Western Australia. The State morbidity system instruction manuals were publicly available documents, ACT and Northern Territory have smaller populations and their system user manuals were not available.

A full year of de-identified actual admitted patient morbidity data publicly available from the State of Victoria was used to test the model. These data were for the year 2005-2006 as the most recent year of such data available at the time of starting the analysis process.

4.6 Test the model against real world data

An analysis of the full admitted patient morbidity data collection for Victoria in 2005 – 2006 was undertaken to test the newly developed morbidity ontology model against the real world collection. This data included 1,311,678 cases for the 12 month period. The data collection has unstructured diagnosis and procedure fields each of which can include up to 40 codes. Due to this it is not possible to confirm all concepts identified and used for the model development through a review of the real data. The data collection represents patterns of diseases/procedures likely to include less common relationships. An examination of the data reveals that where there is a single diagnosis, there are no relationships to be represented, while a single diagnosis and a single procedure have an implied relationship. The objective being to identify the more complex relationships, a selection of more complex entries in the data collection where chosen to identify whether there are any concepts that are not represented in the model that are present in the real data. The intention was to seek the cases with unusual patterns and complexity in the data collected in order to seek the cases that might elicit a clearer picture of the semantic relationships and content of the morbidity data collection.

To achieve this, the dataset was imported into MS Access and queries were developed to find cases with a variety of patterns of data to give an initial indication of the concepts within the collection. The criteria used are described in Table 4-1.

Table 4-1 Testing rationale and criteria

Case Number	Rationale	Query Criteria
1	A case with a large number of diseases and no procedures	Retrieve all records with an entry in the 20^{th} disease field and no entry in the 1^{st} procedure field.
2	A case with few diagnoses and a large number of procedures	Retrieve all records with no entry in the second diagnosis field and an entry in field 10 of the procedure fields.
3	Injury Case – a case where the reason for admission is an injury and no procedure	Retrieve all records where the principal diagnosis field contains a code that begins with an S or a T and where the 1^{st} procedure field is blank.
4	Injury case – a case where the reason for admission is an injury and there are at least 10 procedures	Retrieve all records where the principal diagnosis field contains a code that begins with an S or a T and where there is an entry in the 10^{th} procedure field
5	A complication – a case where there the reason for admission is not an injury and there is a code that indicates a complication of a procedure in the additional diagnoses. This case is used to analyse the content of a case that was admitted for a non injury and where a complication occurred during the patient's stay.	Retrieve all records where the principle diagnosis field does not start with an S or a T and where there is a diagnosis code indicating an injury (starting with an S or a T) in any of the other diagnoses.

When the data were provided from the inpatient morbidity data collection, each record was allocated an arbitrary record number by the MS Access system using the 'Autonumber' facility. The retrieved cases were sorted by this number and the record from the cases returned with the lowest autonumber was analysed. Each code in the disease and procedure fields was searched to determine its meaning (by searching the ICD-10-AM e-book for the code allocated) and the sequence and meaning of each code was recorded. Each code was allocated to an entity type and potential relationships considered to see if the model represents the situation described by the sample case. The content of each case was also reviewed to see if there were any entities or relationships concepts that were not present in the model, and the model was updated to reflect any deficits.

4.7 Comparison to existing models

The outcome of this final stage of the methodology was to evaluate the conceptual framework for admitted patient morbidity data.

A comparison of the identified morbidity data model to the ontological structures relevant for EHR implementation, i.e. OpenEHR RM & Archetypes, the Conceptual Framework for Patient Findings and Problems{ISO, 2007b} and SNOMED-CT high level hierarchy was undertaken to refine the morbidity structure and to provide an insight into the potential relationships between morbidity and EHR systems. This review entailed identification of entity gaps and overlaps.

This process tested the model to identify gaps and overlaps with the concepts in existing health ontologies used in the Electronic Health Record environment to understand where variations are true variations representing different concepts in the real world and where there are consistencies to answer RQ3.

4.8 Limitations

This research was limited to the data structures and relationships that can be identified through the documentation of instructions for the collection of admitted patient morbidity data. This limitation was addressed to some degree by testing the data against data in a real morbidity collection, but these data are affected by the data collection rules. The model developed may not, therefore represent the requirements of all users.

This research did not investigate the reasons for variations between the data required by the State collections, however the production of an accurate model required an understanding of the variations between the States. The inclusion of the four(4) State systems was made to obtain a range of approaches to morbidity data collection, and thereby a more robust model of the data.

The scope of the research was limited to the clinical data elements of admitted episode morbidity data collection and did not include requirements for patient demographic or administrative episode data. The comparison of the data to EHR models gave an indication of the potential for optimizing data collection but was affected by the early stages of development and limited use in Australia of this form of model.

It is recognised that requirements, data included and relationships required for admitted patient morbidity reporting has been influenced by the real or perceived impact upon software development and maintenance costs. This may have influenced the current structure and relationships of the data collection.

4.9 Researcher's Role in the Process

The researcher is a member of the National Health Data Standards Committee and the chair of Standards Australia Health Concept Representation working group and vice-convenor of the International Standards Organisation – Health Informatics – Semantic Content committee. This involvement improves the researcher's ability to gain access to documents and to obtain the cooperation of relevant participants. The researcher is a formal consumer representative to these committees and is therefore not compromised by, but is privy to jurisdictional, clinical or National policy on the decisions made upon the content or processes of admitted patient morbidity data collection.

4.9.1 Access to Resources

The data used in this research were from the public domain. There are no issues regarding obtaining access to, or the use of these data. Patient diagnostic information used to test the model were from publicly available de-identified data provided by the Department of Human Services – Victoria.

4.10 Methodology Summary

The methodology chosen is ontologically based, and created from a mixture of approaches used in the development of health and generic ontologies, designed to support the development of a sound, tested entity-relationship model that would be useful to explain the domain of admitted episode morbidity data to the stakeholders of these data.

Chapter 5 Testing the Ontology Model

Having determined the methodology and scope of the research this chapter identifies the results of the data investigation and modelling development. This process has defined entities and relationships in the clinical data collected in admitted patient morbidity data in Australia.

The entities and relationships to be included into the morbidity model to be developed from the results of this research were identified using the following data collections and standards:

- Conceptual Framework for Patient Findings and Problems in Terminology {ISO, 2007b}
- National Metadata together with the National requirements for morbidity reporting via METeOR{METeOR, 2007}
- Four State morbidity data collection user manuals or guidelines
- ICD-10-AM coding tabular list instructions and coding standards
- Development of entity-relationship model
- Testing of diagram using cases from real data
- Evaluation of the model in comparison to EHR ontology models.

The processes adopted and the results achieved are now presented in this order.

5.1 The domain and scope of the ontology

The domain of the research is the clinical data in the admitted episode morbidity data collections of Australia. These data are defined as those which use ICD-10-AM as the information domain. Throughout this process the ontological convention has been adopted to describe all entities in the singular.

5.2 Existing framework – Starting Point Development

The Conceptual Framework for Patient Findings and Problems in Terminology{ISO, 2007b} identifies entities that represent clinical information. Some of these concepts support direct patient care and are not included in morbidity reporting. These are identified in Tables 4-1 and 4-2 and were assessed in the following stages of the research to confirm that those components declared out of scope were not required for reporting of clinical morbidity data.

The decision made on whether to keep or reject the element was based upon a review of the classification system ICD-10-AM and the researcher's knowledge of the morbidity process. The later stages of the research challenged these assumptions and informed the development of a sound model.

Each component of The Framework was reviewed and identified in scope if it meets the following criteria.
The concept:

- Is clinical - represented in morbidity data using ICD-10-AM
- Is represented as a concept in the ICD-10-AM coding system. The ICD-10-AM coding system is a classification designed to aggregate data{National Centre for Classification in Health, 2004} and therefore does not include the level of detail or complexity of data needed for clinical care. Where a concept in ICD-10-AM includes a group of concepts the additional types of relationships represented in the ICD-10-AM classification were not included as additional concepts required to be separately identified in the model. Where the concept can be represented separately in ICD-10-AM it is considered to be in scope. Table 5-1 indicates the entities in the 'Framework', which were in scope and the ICD-10-AM example that supports that inclusion.

Table 5-1 Entities for morbidity from the Framework

Entity	Sub-Entity	In Scope	ICD-10-AM example of entity{National Centre for Classification in Health, 2004}
Clinical Finding		Yes	Disease – Asthma (J45.9)
	Physiological function	Yes	Dysphagia (R13)
	Interpretation of findings	No – required for clinical practice, but not included in reporting, other than through declared diagnoses (which are modelled in the clinical finding entity)	
	Clinical Test	Yes – where the tests represent clinical risk or significant expense	CAT Scan (60100-00)
Anatomical Structure	Embryonic structure	No - represented in ICD-10-AM procedure and disease code – not a separate entity for morbidity reporting.	Fracture of Nasal Bones (S02.2) represents both the injury (fracture) and the anatomical structures (nasal bones)
	Morphological structure	Yes	Basal Cell Carcinoma (M8090/1)
Organism		Yes	Salmonella (A02.2)
Substance		Yes	Anaesthetic Agent included as a cause of injury (T41.4) Also included 'bundled' with procedures.
Physical Agent		Yes	Fall down (W19)
Physical Object		Yes	Device Implanted (90551-0) Or represented as a 'bundled' concept with physical agent (hit by car where car is the physical object).
Pathological process		No – included in the diagnosis concept	Primary respiratory tuberculosis without mention of bacteriological or histological confirmation
Social Aspect		Yes	Lives Alone (Z60.2)
Pathological Aspect		No	
Events		Yes	Special screening examination for neoplasms
Geographic and environmental locations		Yes	Injury occurred at school (Y92.21)
Person		Yes – as it is included in family history	Family history of Bowel Cancer (Z80.0)
Severity		No – included in the disease code as a separate code	Acute sinusitis (J01.9)
Stages		No – included in the disease code as a separate code	Initial stage of traucoma A71.0
Onset		No – included in the disease code as a separate code	
Episodicity		No – included in the disease code as a separate code	
Temporal Periods		No – included in the disease code as a separate code	

A similar analysis of the relationships provided a starting point for further analysis

Table 5-2 Relationships with potential for morbidity data

Relationship Type	In Scope	Example
HasAbnormalAnatomicalLocation	No – included in the disease or procedure code	Laevocardia (Q24.1)
HasAnatomicalSite	No – included in the disease or procedure code	Fracture of Nasal Bones (S02.2) represents both the injury (fracture) and the anatomical structures (nasal bones)
HasStructuralEmbryologicalDefect	No – included in the disease or procedure code	Maternal care for CNS malformation of foetus (O35.0)
HasOnset	No – included in the disease or procedure code	Sudden cardiac death (I46.1)
HasEpisodicity	No – included in the disease or procedure code	Subsequent myocardial infarction of anterior wall (I22.0)
HasCourse	No – included in the disease or procedure code	Electroconvulsive therapy course < or = 12 treatments
HasOccurrence	No – included in the disease or procedure code	Other seasonal allergic rhinitis (J30.2)
HasMorphology	Yes	Basal Cell Carcinoma (M8090/1)
HasCausativeAgent	Yes	Exposure to excessive natural cold (X31)
HasSeverity	No – included in the disease or procedure code	Bipolar affective disorder, current episode, mild or moderate depression
HasStage	No – included in the disease or procedure code	Initial stage of traucoma A71.0
HasPathologicalProcess	No – included in disease process	Autoimmune Glandular Failure (E31.0)
HasExternalCause	Yes	Bus occupant injured in collision with railway train or railway vehicle (V75)
HasPlaceOfOccurrence	Yes	Driveway to home (Y92.00)
HasSubjectOfInformation	Yes	This concept is often included with others, as in the external cause example above, but it can also be separate to indicate diseases of others. Family history of malignant neoplasm of breast Outcome of delivery – Single live born (Z37.0)
HasQuantity	No	
HasPsychosocialAspect	No	
HasAssociatedFinding	Yes	Cerebral infarction has associated finding of hemiplegia (G81.9)
HasAssociatedFunction	No	Incorporated in the disease description.
HasInterpretation	No	
HasAssociatedTest	Yes	Gastroscopy (30473-00)

This activity identified underlying entities and relationships included in morbidity reporting through the ability to code represent them as individual concepts using ICD-10-AM.

Table 5-3 represents the entities in scope and included as individual concepts in ICD-10-AM.

Table 5-3 Entities established through review of ICD-10-AM and The Framework

<table>
<tr><th>Entity</th><th>Represented in ICD-10-AM as a:</th></tr>
<tr><td>Clinical finding</td><td>Disease code</td></tr>
<tr><td>Physiological function</td><td>Disease code</td></tr>
<tr><td>Clinical test</td><td>Procedure</td></tr>
<tr><td>Morphological Structure</td><td>Morphology code</td></tr>
<tr><td>Organism</td><td>Disease code</td></tr>
<tr><td>Substance</td><td>Procedure (use of substance) and
As cause of injury</td></tr>
<tr><td>Physical Agent</td><td rowspan="2">As external cause of injury (physical agent + physical object)</td></tr>
<tr><td>Physical Object</td></tr>
<tr><td>Social Aspect</td><td>Disease code</td></tr>
<tr><td>Events</td><td>Disease code</td></tr>
<tr><td>Geographic and Environmental Locations</td><td>Place of Occurrence of Injury</td></tr>
<tr><td>Person</td><td>Disease Code</td></tr>
</table>

Table 5-4 represents the relationships in scope and included as individual concepts in ICD-10-AM.

Table 5-4 Relationships established through review of ICD-10-AM and the Framework

Relationship	ICD-10-AM comment
HasMorphology	A clinical finding to be associated with morphology. In cases of cancer this is a documented requirement.
HasCausativeAgent	Some diseases require the underlying disease cause to be specified, others can have this included as an extra code.
HasExternalCause	Morbidity classification in Australia includes injuries as diseases – though these conditions can be identified as the code used to represent them starts with an S or a T. These disease codes require the specification of an external cause of injury
HasPlaceOfOccurrence	Injury disease codes require the indication of the place of occurrence of the injury, except where this is obvious (hospital procedure complications)
HasSubjectOfInformation	Some concepts included in morbidity classification are not for the patient, but to others associated with the patient, such as family history, or outcome of delivery
HasAssociatedFinding	A disease may have an associated finding that is recorded to more precisely describe the impact of that disease.
HasAssociatedTest	A person may be admitted for a specific test, without a disease being found, or a test may be undertaken related to a disease.

5.3 Entities and Relationships

The entities and relationships identified through the start up process using The Framework were analysed using the National Metadata and requirements for morbidity reporting to inform the entity and relationship tables.

5.3.1 National Metadata

METeOR identifies the following data components as essential data elements in the minimum data set for admitted patient care represented in METeOR.

Principal diagnosis is defined as…

> *"The diagnosis established after study to be chiefly responsible for occasioning an episode of admitted patient care, an episode of residential care or an attendance at the health care establishment, as represented by a code"* {METeOR, 2005}.

A principal diagnosis is a special case of Clinical Finding extracted as a special requirement to meet the needs of National reporting and casemix, for these reasons this concept has been established as a specific entity.

Additional Diagnosis is defined as..

> *"Generally, external cause, place of occurrence and activity codes will be included in the string of additional diagnosis codes. In some data collections these codes may also be copied into specific fields"* {METeOR, 2005}.

This definition indicates that, from the ontological perspective this is not one element, when referred to the Starting Point tables the content of additional diagnosis field may include:

Clinical finding

Morphological structures

Organism

External Cause of Injury (In ICD-10-AM Physical Agent is combined with Physical Object in a single code to represent the cause of injury)

Geographic and environmental location

Person

In addition to this the system requires that when the person was admitted for an injury specified by a code that starts with an S or a T the additional diagnosis field must include a code that indicates what the person was doing (activity) at the time of injury.

The cause of injury concept includes an indication of a substance to which a person has had an adverse experience.

Injury Event – Place of Occurrence is defined as..

> *"The place where the external cause of injury, poisoning or adverse effect occurred, as represented by a code"* {METeOR, 2007}.
>
> This entity can be represented by the Geographic and environmental locations concept in the Framework.

Injury Event – External Cause of Injury is defined as..

> *"The environmental event, circumstance or condition as the cause of injury, poisoning and other adverse effect, as represented by a code."* {METeOR, 2005}.

This entity is already represented in the Framework in the ideas of organisms, substances, physical agents, and physical objects with a relationship of 'has causative agent' to the injury.

Procedure is defined as

> *"A clinical intervention represented by a code that:*
> - *is surgical in nature, and/or*
> - *carries a procedural risk, and/or*
> - *carries an anaesthetic risk, and/or*
> - *requires specialised training, and/or*
>
> *requires special facilities or equipment only available in an acute care setting"*{METeOR, 2007}.

The Framework does not intend to cover the scope of procedures. Therefore the procedure entity was added to the list of required entities. Table 5-5 shows the entities identified at completion of this stage of the review.

Table 5-5 Entities at National metadata level

Entity	**Represented in ICD-10-AM as a:**
Principal diagnosis	Disease code (not an injury or morphology code) – additional entity
Clinical finding	Disease code
Injury	Disease code
Physiological function	Disease code
Clinical test	Procedure
Morphological Structure	Morphology code
Organism	Disease code
Substance	Procedure (use of substance) and As cause of injury
External Cause of Injury	Combination of Physical Object and Physical Agent
Social Aspect	Disease code
Activity	Disease code – added to meet requirements of morbidity reporting
Events	Disease code
Place of Occurrence of Injury	Specification of geographic and environmental locations for the purpose of national morbidity reporting
Person	Disease Code
Procedure	Procedure Code – added to include the scope of morbidity reporting

The relationships to entities relevant to morbidity reporting are not indicated in the National minimum data set which does not have a structure that permits this type of metadata. The addition of procedures and activities requires the additions to represent the relationships between these entities. Table 4-5 below lists the possible relationship types that will be considered when reviewing State and coding instructions.

Table 5-6 Relationships at National metadata level

Relationship	Modification
HasMorphology	No change
HasCausativeAgent	No change
HasExternalCause	No change
HasPlaceOfOccurrence	No change
HasSubjectOfInformation	No change
HasAssociatedFinding	No change
HasAssociatedTest	No change
HasActivityAtOccurrence	Added to represent relationships between injuries and the activity being undertaken by an individual when the injury occurred.
HasTreatment	Indicates that a disease or injury has a treatment procedure performed.

5.3.2 State data collection review

ICD based data elements in the 4 State Collections are summarised in Table 4-6. Where there were variations between the National approach and the State data elements these were investigated to identify any refinement of the entities and relationship requirements.

Table 5-7 State collection data elements

Data Element	States
Principal Diagnosis	All States (using National METeOR definition)
Co-Diagnosis	Western Australia
Additional Diagnosis	All States with some minor variations
Diagnosis Type	Victoria and Queensland
External Cause of Injury	All States collect this but not all have it as a separate data element
Place of Occurrence of External Cause of Injury	All States collect this but not all have it as a separate data element
Activity when Injured	All States collect this but not all have it as a separate data element
Morphology	All States collect this but not all have it as a separate data element
Procedures	All States (using National METeOR definition)

Co-Diagnosis

Specified as an individual field in Western Australia, though this same concept is included in all data collections within the additional diagnosis field. This field is described as

> *"Some diseases can affect two different body systems. Where there is a relationship between the disease and its manifestations (commonly called dagger and asterisk codes), the co-diagnosis field is used to record the disease matching the Principal Diagnosis. The field may also be called "CodeAlso" or "Co Diagnosis". These matched codes can also be reported as additional diagnoses. These codes must be reported with their manifestation. It is easy to recognise these codes in the ICD-10-AM Tabular List as they are enclosed in brackets after the code description. If the Principal Diagnosis is a Dagger code then the next sequenced code must be an Asterisk (Code Also) code. However, if the Dagger code is not the Principal Diagnosis then the next sequenced code is still an Asterisk code, but is entered as an Additional Diagnosis in the record sent to HMDS"* {Department of Health, 2006}.

This requirement does not represent a new entity, rather the requirement for a relationship that indicates the manifestation of a condition. For the purpose of morbidity reporting this is considered to be the relationship called "HasCausativeAgent" in The Framework.

Additional diagnosis

Victoria, Queensland and NSW use the METeOR definition for this data element, which includes manifestation codes in the additional diagnosis data element. Western Australia defines additional diagnosis as:

> *"An unrestricted number of additional diagnoses including morphology codes, external cause, place of occurrence and activity codes (depending on the local information system)*" {Department of Health, 2006}.

This definition does not require modification to the entities or relationships as it is consistent with the data included by the other States and represents the exclusion of manifestations from this data set in Western Australia.

Diagnosis Type

This concept is not present in the data collection of NSW or Western Australia. In Victoria a prefix has been used to each individual code entered in the principal diagnosis and additional diagnosis fields to indicate whether the diagnosis is:

P – principal diagnosis (and any additional diagnosis that were 'code also' relationships to the principal diagnosis or where there are multiple injuries as a result of one accident, one is THE principal diagnosis but the others are also principal in nature. The causes of injury, place of occurrence and activity associated with the principal injury are also prefixed with a "P" {Health Data Standards and Systems Unit, 2006}.

A – Associated diagnosis – ICD-10-AM codes that are not a principal diagnosis but are relevant to the episode of care are prefixed with an "A" {Health Data Standards and Systems Unit, 2006}.

C – Complications are recorded with a prefix of "C". This includes the code that describes the disease or injury and the cause of injury, place of occurrence and activity. In the 1990's sequelae were recorded with an S prefix. This convention is no longer used {Health Data Standards and Systems Unit, 2006}.

Queensland has a more extensive approach than Victoria. The categories they used to indicate these relationships are:

"P Principal Diagnosis PD
A Other/Associated Diagnoses
C Complications
PE External Cause associated with the Principal Diagnosis
AE External Cause related to Associated Diagnosis
CE External Cause associated with the complication
PM Morphology associated with the Principal Diagnosis
AM Morphology related to Associated Diagnosis
CM Morphology associated with the Complication"
{Data Services Unit, 2005 }.

These requirements are indicators of relationships in the data collection. This approach is a clumsy approach, for example there is no facility to indicate which of the associated diagnoses the AE – external cause code relates, other than through sequencing of the codes in the additional diagnosis field.

The research has already identified the need for the HasMorphology relationship. The requirement to indicate when a condition is a complication of a disease, and injury or potentially a procedure (which is not a clear option on the list of diagnosis types) has been identified. Though this can be expressed by the ability to represent a causative relationship between diseases, causes and procedures.

This relationship can be represented by relating a disease with the underlying cause of the disease or a procedure with a disease that represents the complication caused by that procedure.

Example
When a post operative infection of a procedure occurs the relationship is Infection(diagnosis) HasExternalCause Procedure. This describes a post-operative infection that can be specific about the infection type and the relationship applies directly to the specific procedure involved).

Additional Example
A person is admitted to hospital for treatment of Kidney disease which is a result of hypertension. This could be represented as:

> Hypertension HasComplication Kidney Disease IsAPrincipal Disease
>
> or as
>
> KidneyDisease HasCausativeAgent Hypertension.

In this research no new relationship has been established as complications are seen to be either diseases with an underlying cause or an external cause. Further investigation of this decision would be warranted as it is not proven to be the most appropriate response to the requirement.

External Cause of Injury

An external cause cannot be recorded as a Principal Diagnosis. An individual can have more than one injury in an episode of care. Some States include external causes of injury codes only in the string of additional diagnoses, and computers can identify them by the code used and the assumption is made that there is a relationship between this cause and the immediately preceding disease code/s The relationship of HasExternalCause is already represented in the relationships available to the model.

Western Australia, New South Wales and Queensland have specific fields to accept the external cause of injury. This field is used to hold the cause of injury that relates to the principal diagnosis, or most important additional disease which is an injury. The Western Australian requirement which is consistent with METeOR, is described in the user manual as:

> *"The external cause is the environmental event, circumstance and other condition that caused an injury, poisoning or adverse effect. This information is important because it is reported to the National Injury Surveillance Unit (NISU) for the examination of causes of injury and poisoning, and the setting of targets for reduction of these events. Guidelines for coding external causes:*
>
> - *Select the most specific code possible.*
> - *Unlimited external cause codes may be recorded.*
> - *The external cause code should be sequenced after the diagnostic code to which it relates.*
> - *All external cause codes require a place of occurrence and activity code as set out in the WA Coding Standards"* {Department of Health, 2006}.

Place of Occurrence of External Cause of Injury

When an injury is recorded, or a disease associated with an external cause of injury, the place of occurrence of the injury is a required additional diagnosis code in all States. In some cases there is a specific field for this entry, in other cases it is recorded after the external cause of injury code. This is a concept required as an entity and as a relationship that indicates the specific injury to which this place of occurrence applies, to handle the situation where a person has more than one injury recorded in the admitted episode.

Activity when injured

When an injury is recorded, or a disease associated with an external cause of injury, the activity the person was undertaking at the time of injury is usually required (except where the injury is post-operative, where the activity is implied by the injury). In some cases there is a specific field for this entry, in other cases it is recorded after the external cause of injury code. This entity and relationship are in the entity relationship collection identified by the National morbidity data and therefore there is no need for addition or change.

Morphology

The HasMorphology relationship has been defined previously and is consistent with the requirements indicated in the State documentation.

As a result of this review the relationships there were no additional entities or relationships established.

5.3.3 ICD-10-AM Relationships

A search of ICD-10-AM e-book(NCCH, 2006) (e-book) was undertaken to identify coding instructions that represent the requirement for additional entities or relationships. This review was not seeking coding details, rather the relationships implied when the use of multiple codes is required.

A detailed search of the e-book identified the following instruction types and entities.

Use additional code for

Use additional code if applicable to identify

Use additional code with subcategories to identify

Dagger/Asterisk – manifestation combinations

Code also

Code when related to reason for admission or treatment

Use additional external cause code to identify

Code also when performed

Every such instruction in the code book was reviewed and identified with an entity and the relationships represented by the code obtained by following the instruction. Table 5-8 shows the resultant relationships and an example of each.

Table 5-8 ICD-10-AM instruction relationships

Entity	Relationship	Example and comment
Principal Diagnosis (PD)	HasCausativeAgent, (Dagger/Asterisk)	PD: Meningitis (G01) HasCausativeAgent Salmonella (A022) This is a dagger/asterisk combination required in ICD-10-AM
	HasExternalCause, HasPlaceofOccurrence HasActivityatOccurrence	PD: Fracture of upper end of tibia (S82.1) hasExternalCause Pedestrian injured in collision with car(V02), HasPlaceof Occurrence on public road (Y92.49) HasActivityatOccurrence on the way to work as a shop assistant (U73.04) Or the case above with a procedure that has a complication. PD: Fracture of upper end of tibia (S82.1) hasExternalCause Pedestrian injured in collision with car(V02), HasPlaceof Occurrence on public road (Y92.49) HasActivityatOccurrence on the way to work as a shop assistant (U73.04) hasTreatment Open reduction of fracture of shaft of tibia with internal fixation (47566-01) Additional Diagnosis: infection of procedure site (T81.41) HasExternalCause 47566-01 (the procedure)
	HasAssociatedFinding	PD: Cellulitis of face (L03.2) HasAssociatedFinding Eyelid Involvement (H00.0) PD: Second Degree Burn of Chest (T21.22) HasAssociatedFinding Burn to 15% of Body Surface
	HasMorphology	PD: Malignant neoplasm of nose (C44.3) HasMorphology Squamous Cell Carcinoma (M8070/3)
	HasAssociated	PD: Fracture of upper end of tibia HasAssociated Open Wound(S81.81) over the fracture site.
	HasHistoryOf	PD: Type 2 Diabetes HasHistoryOf Long term use of insulin (Z92.22)
	HasDuration	PD: Spontaneous Abortion HasDuration 16 weeks of pregnancy (O09.2)
	HasCauseofDeath	Event (admitted episode): HasCauseofDeath pneumonia (J18.9)
	HasHazard	Event (admitted episode): HasRiskFactor Smoker (Z72.0)
	HasTreatment	PD: Acute Appendicitis (K35.9) HasTreatment Appendicectomy (30375-30)
	HasAssociatedTest	PD: Dyspepsia (R13) HasAssociatedTest Panendoscopy to duodenum (30473-00)
	HasSubjectofInformation	PD: Normal Vaginal Delivery (O80) HasSubjectofInformation single live birth (Z37.0)

A review of the tabular list of codes also identified additional entities that are not diseases, though they are recorded in the principal, co-morbidity or additional diagnosis fields. These are:

Reason for admission not associated with a disease/injury

- encounters with health service for examination and investigation
- potential health hazards associated with communicable diseases
- circumstances related to reproduction

- encounters for specific procedures and health care (eg removal of stitches where the disease is no longer present)
- socioeconomic and psychosocial health hazards
- other reasons for encounter with the health service
- potential health hazards (such as smoking)
- environment issues

These have been aggregated into two groups of similar concepts that represent different administrative and epidemiological types of information. Though the aggregation used could be argued, the intent is to represent information of clinical relevance (the hazards) and information of administrative or service provision relevance. These groups of data have different users and serve different purposes in the data collection.

Hazards (including psycho/social hazards and replaces the previous entity – social aspect as only those social aspects that represent a hazard are required in morbidity data)

- potential health hazards associated with communicable diseases
- socioeconomic and psychosocial health hazards
- potential health hazards
- environment issues

Services (called events in the earlier examples but more correctly an indicator of a service)

- encounters with health service for examination and investigation
- circumstances related to reproduction
- encounters for specific procedures and health care (eg removal of stitches where the disease is no longer present)
- other reasons for encounter with the health service

A similar review was taken of the Australian Classification of Interventions (ACHI) (volumes 3 and 4 of ICD-10-AM) and this and the associated standards for procedures identified the following additional requirements

Table 5-9 ICD-10-AM procedure concepts

Entity	Instruction	Relationship
Procedure	Use additional anaesthesia code when appropriate	HasAnaesthesia
	Code also when required	HasAssociatedProcedure
	Code also when required, reconstruction procedures	HasAssociatedProcedure
	Code also when required graft procurement	HasAssociatedProcedure
	Test performed for disease	Disease HasAsssociatedTest
Anaesthesia	A specialisation of procedure	

Table 5-10 ICD-10-AM entities and relationships from procedure instructions

Entity	Relationship	Example and comment
Procedure	HasAnaesthetic	PD: Acute Appendicitis (K35.9) HasTreatment Appendicectomy (30375-30) HasAnaesthetic General Anesthetic in emergency(92514-30)
Procedure	HasAssociatedProcedure	Procedure: Neuroendoscopy HasAssociatedProcedure insertion of shunt (40003-00) HasAssociatedAnaesthetic General anaesthetic not in an emergency (92514-29)

Though ICD-10-AM provides instruction on the manifestation, aetiology relationship as represented in the classification using dagger and asterisk pairs, there are also instructions to code underlying causes that are not part of such pairs, largely, it appears, because there are a large number of alternative options. I have chosen to use the one relationship HasCausativeAgent to represent this idea.

The concept of physiological function was removed as the entity clinical finding will represent this in a manner that meets the relationship and entity requirements of ICD-10-AM there is no semantic or structural implication to morbidity data in the removal of this entity. Clinical Tests are represented in ICD-10-AM as a procedure. Many are not coded in morbidity reporting. For this reason they have been aggregated into the procedure entity.

Table 5-11 includes the results of the identification of entities in morbidity data in Australia.

Table 5-11 Entities identified in review

Entity	Represented in ICD-10-AM as a:
Principal diagnosis	Disease code (not an injury or morphology code) – additional entity
Clinical finding	Disease code (additional diagnoses)
Injury	Disease code
Morphological Structure	Morphology code
Organism	Disease code (as a causative agent of disease)
Substance	Procedure (use of substance) and As cause of injury or disease
External Cause of Injury	Combination of Physical Object and Physical Agent
Hazard	Disease code
Activity	Disease code – added to meet requirements of morbidity reporting
Service	Disease code
Place of Occurrence of Injury	Specification of geographic and environmental locations for the purpose of National morbidity reporting
Person	Disease Code
Procedure	Procedure Code – added to include the scope of morbidity reporting
Anaesthesia	Procedure Code

The relationships were reviewed and updated to reflect the requirements and content discovered in the ICD-10-AM review.

Table 5-12 Relationships identified after review

Relationship	Modification
HasCausativeAgent	No change
HasExternalCause	No change
HasPlaceOfOccurrence	No change
HasActivityAtOccurrence	Added to represent relationships between injuries and the activity being undertaken by an individual when the injury occurred.
HasAssociatedFinding	No change
HasAssociated	Added to cover the relationship between multiple injuries occurring with one cause, or diseases associated by a relationship that is not one of the ones specified.
HasMorphology	No change
HasHistoryOf	Added to cover this concept of historical references to conditions or procedures
HasDuration	Added to include the ability to indicate duration of pregnancy, but could be used to cover other duration based information.
HasCauseofDeath	Indicates the relationship between a Principal disease or a clinical finding to the individual's death.
HasAssociatedTest	No change
HasTreatment	Indicates that a disease or injury has a treatment procedure performed.
HasSubjectOfInformation	No change
HasAnaesthesia	Added to indicate the anaesthetic related to a given procedure or group of associated procedures
HasAssociatedProcedure	Added to indicate a group of procedures that are related or occurred at the same time as the procedure to which they are associated.

These entities and relationships were analysed to produce an entity-relationship model

5.4 The Model

The entities and relationships identified were represented graphically in an information model. The methodology used to represent the concepts is a combination of the representation of entities used in The Categorial Framework for Patient Findings and Problems and the relationship linkages identified through the investigation of morbidity data requirements and meaning.

Rectangles are used to represent entities. The depiction of an entity is within another entity indicates that the internal entity is a subset or specialisation of the external entity within which it is contained. Where the rectangle overlaps another entity the group of entities are a set used to describe a given situation (for example the cause of injury, place of occurrence and activity are a group of entities required to describe an injury.

The clinical finding rectangle overlaps itself. This convention is used to indicate that all relationships that go into the clinical finding can be relationships between that clinical finding and another clinical finding.

Connective lines are used to indicate relationships, the name of the relationship is on the line, and the arrow indicates the semantic relationship represented. This example represents the entity below the diagram was caused by the external factor in the entity to which it points.

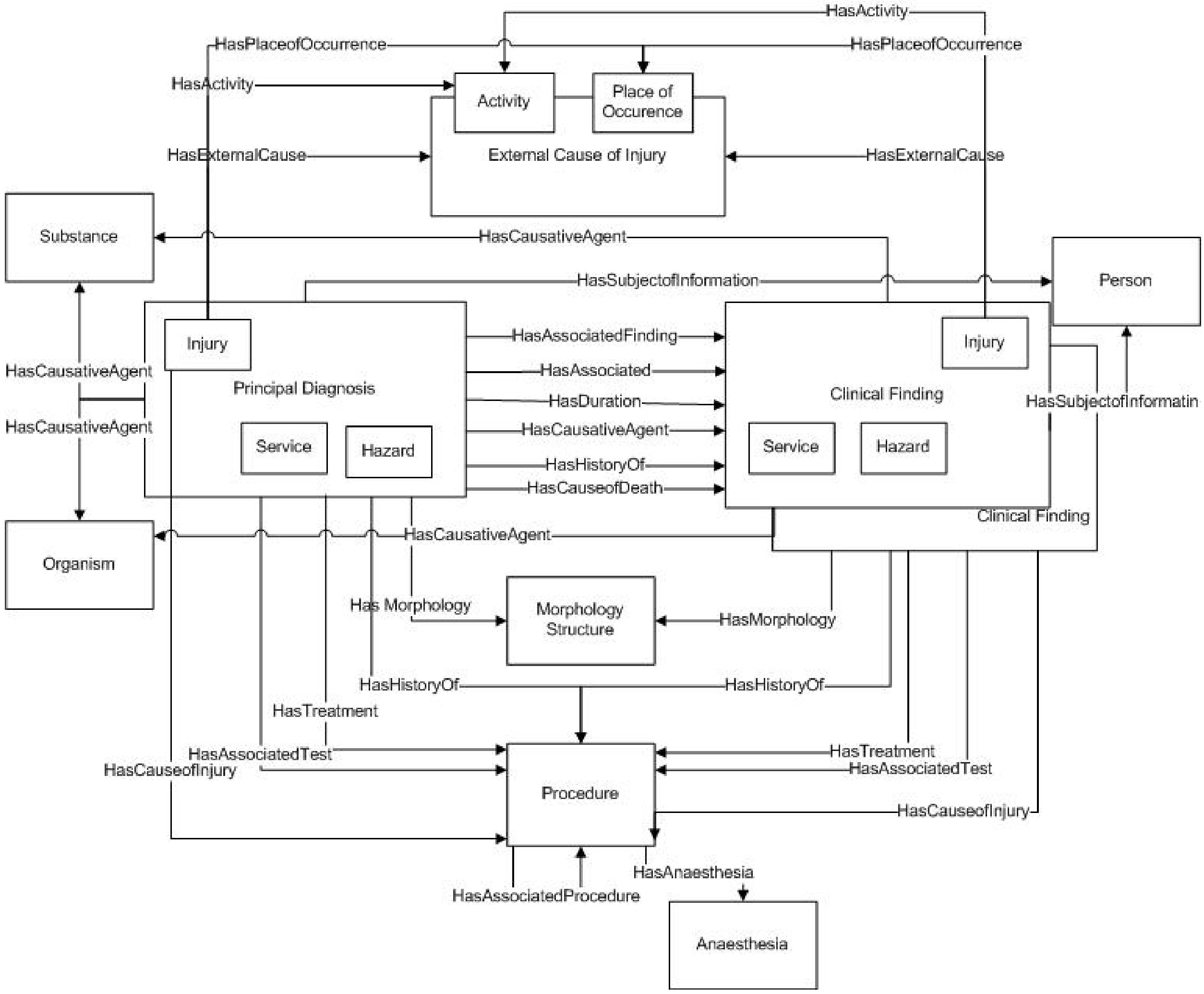

Figure 5-1 Model of morbidity data

To confirm and improve this model a test of the model using real data was undertaken.

5.5 Testing the Model

The entities and relationships identified in the model were tested by extracting the five sample cases from 12 months of Victorian Morbidity data 1,311,678 cases. These cases were chosen to represent particularly complex patterns of data. A query was built to seek cases that meet the query criteria and the case with the lowest record number in the file that meets that criteria was selected for investigation in each case.

5.5.1 Case 1 Many diagnoses and no procedures

Table 5-13 Case 1

Case Number	Rationale	Query Criteria
1	A case with a large number of diseases and no procedures	Retrieve all records with an entry in the 20th disease field and no entry in the 1st procedure field.

There were 27 cases that met these criteria.
The first case.
1: Principal diagnosis: Sepsis due to other gram negative organisms
Additional Diagnoses:
2: Type 2 diabetes with diabetic ischaemic cardiomyopathy
3: Stage 4 decubitus ulcer
4: Staphylococcus aureus as the cause of diseases classified to other chapters
5: Arteriosclerosis of arteries of extremities – with gangrene
6: Arteriosclerosis of arteries of extremities – with ulceration
7: Cellulitis of Trunk
8: Cellulitis of Lower Limb
9: Type 2 diabetes with other specified renal complications
10: Acute Renal Failure
11: Type 2 diabetes with advanced renal disease
12: Unspecified chronic renal failure
13: Hyperkalaemia
14: Type 2 diabetes with features of insulin resistance
15: Hypertension
16: Unspecified urinary incontinence
17: Faecal incontinence
18: Anaemia
19: Pneumonia
20: Palliative Care
21: Use of long term insulin

This case is represented in Table 5-14 using the entities and relationships from the morbidity model.

Table 5-14 Relationships and entities in Case 1

Entity	Content	Relationship	Comment
Principal Diagnosis	Sepsis due to other gram negative organisms		It is not possible to identify relationships
Clinical finding	Type 2 diabetes with diabetic ischaemic cardiomyopathy	HasAssociation with Type 2 diabetes with advanced renal disease HasAssociation with Type 2 diabetes with features of insulin resistance	It is not possible to identify relationship other than the other conditions of type 2 diabetes.
Clinical finding	Stage 4 decubitus ulcer	HasCausativeAgent	This relationship is assumed as the infectious agent follows immediately after this condition, though it is not possible to be sure that this is an accurate representation
Organism	Staphylococcus Aureus as the cause of diseases classified to other chapters		
Clinical finding	Arteriosclerosis of arteries of extremities – with gangrene	HasAssociated	This is possibly related to the diabetes, but it is not possible to say this with certainty. However it is a reasonable assumption that this condition is related to the same condition with ulceration.
Clinical finding	Arteriosclerosis of arteries of extremities – with ulcerations		
Clinical finding	Cellulitis of Trunk		This may or may not be related to the cellulitis of lower limb which follows – no association made due to lack of relationship between the sites
Clinical finding	Cellulitis of Lower Limb		It is not possible to identify relationships
Clinical finding	Type 2 Diabetes with specified renal complications		The diabetes code indicates a relationship exists between these two codes
Clinical finding	Acute Renal Failure	HasCausativeAgent	

This case represented a need to include a relationship of history of substances and the need to be able to consider diseases in a group, the multiple conditions associated with Type 2 diabetes. The model was modified to reflect the HasHistoryOf requirement between diseases and substances.

5.5.2 Case 2 Few diagnoses and many procedures

Table 5-15 Case 2

<table>
<tr><th>2</th><th>A case with few diagnoses and a large number of procedures</th><th>Retrieve all records with no entry in the second diagnosis field and an entry in field 10 of the procedure fields.</th></tr>
<tr><td colspan="3">There were 38 cases meeting these criteria. The first case had the following codes.

Principal Diagnosis: Other Schizophrenia
Procedure 1: Electroconvulsive therapy > 12 treatments
Procedure 2: General anaesthesia
Procedure 3: General anaesthesia
Procedure 4: General anaesthesia
Procedure 5: General anaesthesia
Procedure 6: General anaesthesia
Procedure 7: General anaesthesia
Procedure 8: General anaesthesia
Procedure 9: General anaesthesia
Procedure 10: General anaesthesia
Procedure 11: General anaesthesia
Procedure 12: General anaesthesia
Procedure 13: General anaesthesia
Procedure 14: General anaesthesia
Procedure 15: General anaesthesia
Procedure 16: General anaesthesia
Procedure 17: General anaesthesia
Procedure 18: General anaesthesia
Procedure 19: General anaesthesia
Procedure 20: General anaesthesia
Procedure 21: General anaesthesia
Procedure 22: General anaesthesia
Procedure 23: General anaesthesia</td></tr>
</table>

Analysis of this case indicates a new requirement, the capacity to indicate multiple occasions of a procedure or set of procedures. This will be represented by an entity of Quantity and a semantic relationship of HasQuantity The entities and relationships of this case are described in Table 5-16.

Table 5-16 Relationships and entities in Case 2

Entity	Content	Relationship	Comment
Principal Diagnosis	Other Schizophrenia	HasTreatment	
Procedure	Electroconvulsive therapy >12	HasAnaesthesia	This relationship to anaesthesia creates a link between this concept and the anaesthetics given for each therapy session.
Anaesthesia	General Anaesthesia	HasQuantity	The relationship allowing entry of quantity is proposed as a more user friendly and reliable mechanism for recording repletion than collecting the same code many times. This also directly relates the quantity to the procedure.
Quantity	22		

5.5.3 Case 3 Reason for admission is injury with no procedure

Table 5-17 Case 3

<table>
<tr><th>3</th><th>Injury Case – a case where the reason for admission is an injury with no procedure</th><th>Retrieve all records where the principal diagnosis field contains a code that begins with an S or a T and where the 1st procedure field is blank.</th></tr>
<tr><td colspan="3">Number of cases: 30,160 , the first retrieved entry had the following entries:

Principal Diagnosis: Fracture of calcaneus
Associated Diagnoses:
2: Other Fracture of shaft of tibia
3: Fracture of Triquetral bone of wrist
4: Other fall from one level to another
5: Occurring at other and unspecified place at home
6: While undertaking domestic duties</td></tr>
</table>

Table 5-18 Relationships and entities in Case 3

Entity	Content	Relationship	Comment
Principal Diagnosis	Fracture of calcaneus	HasAssociated HasExternalCause	These conditions are associated together through one external cause and one place of occurrence.
Clinical finding	Other fracture of shaft of tibia	HasAssociated HasExternalCause	
Clinical finding	Fracture of Triquetral bone of wrist	HasAssociated HasExternalCause	
External Cause of Injury	Other fall from one level to another	HasPlaceofOccurrence HasActivityatOccurrence	
PlaceofOccurrence	Other places at home		
Activity	Undertaking Domestic Duties		

5.5.4 Case 4 Injury not reasons for admission with many procedures

Table 5-17 Case 4

4	Injury case – a case where the reason for admission is an injury and there are at least 10 procedures	Retrieve all records where the principal diagnosis field contains a code that begins with an S or a T and where there is an entry in the 10th procedure field

There were 2,893 cases returned that meet this criteria. The first returned case had the following composition.

Principal Diagnosis: Injury of cervical spinal cord, unspecified
Associated Diagnoses:
2: Functional spinal cord injury, C1
3: Other dislocation of spinal vertebrae
4: Fracture of first cervical vertebra
5: Loss of Consciousness of unspecified duration
6: Traumatic cerebral oedema
7: Injury of oculomotor nerve
8: Traumatic pneumothorax
9: Haematoma of spleen
10: Pulmonary collapse
11: Pedestrian in collision with car
12:: Occurring on a Roadway
13: Unspecified activity
14: Harmful use of opioid
15: Paralysis of vocal cords or larynx, bilateral, partial
16: Epilepsy
17: Chronic viral hepatitis
18: Cytomegaloviral disease
19: Pneumonia due to staphylococcus
20: Methicillin resistant agent
21: Chronic renal impairment
22: Depressive episode, unspecified
23: Pneumonia due to pseudomonas
24: Nosocomial condition
25: Pleural effusion
26: Urinary tract infection
27: Sepsis, unspecified
28: Stage 3 decubitus ulcer
29: Pseudomonas as the cause of diseases classified to other chapters
30: Other medical Procedures as cause of injury
31: Other complications due to nervous system device, implant and graft
32: Surgical operation with implant of artificial internal device
33: Other complications of genitourinary prosthetic devices, implants and grafts
34: Mechanical complication of urinary (indwelling) catheter
35: Urinary catheterisation
36: Tracheostomy malfunction
37: Surgical operation with formation of external stoma
38: Elevated blood glucose level
39: Fever
40: Complication of other systemic antibiotics

Procedure 1: Closed reduction of fracture / dislocation of spine with spinal cord involvement, with immobilisation
Procedure 2: Posterior spinal fusion, >3 levels
Procedure 3: Segmental internal fixation of spine, 3 or 4 levels
Procedure 4: Procurement of bone for graft via separate incision
Procedure 5: General anaesthetic
Procedure 6: Percutaneous tracheostomy
Procedure 7: Management of tracheostomy
Procedure 8: Management of continuous ventilation support
Procedure 9: Lumbar puncture
Procedure 10: Initial insertion of percutaneous endoscopic gastroscopy [PEG]tube
Procedure 11: Sedation ASA 49
Procedure 12: Repeat incision of percutaneous endoscopic gastroscopy [PEG] tube
Procedure 13: Sedation ASA 49
Procedure 14: Percutaneous cystotomy
Procedure 15: Sedation ASA 49
Procedure 16: Bronchoscopy through artificial stoma
Procedure 17: General anaesthesia, ASA 49
Procedure 18: Fibreoptic colonoscopy to caecum
Procedure 19: Sedation, ASA 49
Procedure 20: Wedge resection of ingrown toenail
Procedure 21: Surgical removal of a tooth or tooth fragment, requiring removal of bone.
Procedure 22: Treatment of acute periodontal infection
Procedure 23: General anaesthesia, ASA 49
Procedure 24: Transfusion of packed cells
Procedure 25: Computerised tomography of brain
Procedure 26: Computerised tomography of spine, cervical region
Procedure 27: Computerised tomography of soft tissue of neck with intravenous contrast medium
Procedure 28: Computerised tomography of abdomen and pelvis with intravenous contrast medium
Procedure 29: Retrograde micturating cystourethrography
Procedure 30: Computerised tomography of chest, abdomen and pelvis with intravenous contrast medium
Procedure 31: Allied health intervention, occupational therapy
Procedure 32: Allied health intervention, social work
Procedure 33: Allied health intervention, dietetics
Procedure 34: Allied health intervention, prosthetics and orthotics
Procedure 35: Allied health intervention, physiotherapy
Procedure 36: Allied health intervention, speech pathology
Procedure 37: Allied health intervention, pastoral care
Procedure 38: Allied health intervention, music theory

Table 5-20 Relationships and entities in Case 4

Entity	Content	Relationship	Comment
Principal Diagnosis	Injury of cervical spinal cord	HasExternalCause HasAssociated HasTreatment	all of these are injuries related to the one set of external cause information.
Clinical finding	Functional spinal cord injury C1	HasExternalCause Has Associated HasTreatment	
Clinical finding	Other dislocation of spinal vertebrae	HasExternalCause Has Associated HasTreatment	
Clinical finding	Fracture of first cervical vertebra	HasExternalCause Has Associated HasTreatment	
Clinical finding	Loss of consciousness	HasExternalCause Has Associated	
Clinical finding	Traumatic cerebral oedema	HasExternalCause Has Associated	
Clinical finding	Injury of oculomotor nerve	HasExternalCause Has Associated	
Clinical finding	Traumatic pneumothorax	HasExternalCause Has Associated	
Clinical finding	Haematoma of spleen	HasExternalCause Has Associated	
Clinical finding	Pulmonary collapse	HasExternalCause Has Associated	
Cause of Injury	Pedestrian in collision with car	HasPlaceofOccurrence HasActivityofOccurrence	
Place of Occurrence	Roadway		
Activity	Unspecified		
Clinical finding	Paralysis of vocal cords	HasExternalCause (Procedure)	Tracheostomy relationship
Clinical finding	Epilepsy		No clear relationship identified
Clinical finding	Chronic viral hepatitis		No clear relationship identified
Clinical finding	Cytomegaloviral disease		No clear relationship identified
Clinical finding	Pneumonia due to staphylococcus	HasAssociated	Though it isn't clear there may be a relationship to the methicillin resistant agent…;..
Clinical finding	Methicillin resistant agent		

Clinical finding	Chronic renal impairment		No clear relationship identified
Clinical finding	depressive episode		No clear relationship identified
Clinical finding	Pneumonia due to pseudomonas		No clear relationship identified
Clinical finding	Nosocomial condition		No clear relationship identified
Clinical finding	Pleural effusion		No clear relationship identified, possibly associated with the Pneumonia
Clinical finding	Urinary tract infection		No clear relationship identified
Clinical finding	Sepsis, unspecified		No clear relationship identified
Clinical finding	State 3 decubitus ulcer	HasCausativeAgent	
Organism	Pseudomonas as cause of disease		
Cause of Injury	Medical procedure as cause of injury	HasPlace of Occurrence HasActivityatOccurrence	It is unclear which of the conditions above represent the injury.
Clinical finding	Other complication due to nervous system device implant and graft	HasExternalCause	This is a confusing group of complications where it is difficult to ensure appropriate relationships.
Cause of Injury	Surgical operation with implant of artificial internal device	HasPlace of Occurrence	
Clinical finding	Other complication of genitourinary prosthetics devices	HasExternalCause	
Cause of Injury	Mechanical complication of urinary (indwelling) catheter		
Cause of Injury	Urinary Catheterisation		
Clinical finding	Tracheostomy malfunction	HasExternalCause	
Cause of Injury	Surgical operation with formation of external stoma		It is not clear which surgical procedure this relates to.

Clinical finding	Elevated blood glucose level		No clear relationship identified
Clinical finding	Fever		No clear relationship identified
Clinical finding	complication of other systemic antibiotics		No clear relationship identified
Procedure	Closed reduction of fracture / dislocation of spine with spinal cord	HasAssociatedProcedure HasAnaesthesia	
Procedure	Posterior spinal fusion, > 3 levels	HasAssociated	
Procedure	Segmental internal fixation of spine	HasAssociated	
Procedure	Procurement of bone for graft via separate incision	HasAssociated	
Anaesthetic	General Anaesthetic		
Procedure	Percutaneous tracheostomy	HasAssociated	
Procedure	Management of tracheostomy	HasAssociated	
Procedure	Management of continuous ventilation support	HasAssociated	
Procedure	Lumbar puncture		
Procedure	Initial insertion of percutaneous endoscopic gastric [PEG]	HasAnaesthesia	
Anaesthesia	Sedation		
Procedure	Repeat incision of percutaneous endoscopic gastroscopy [PEG}	HasAnaesthsia	
Anaesthesia	Sedation		
Procedure	Percutaneous cystotomy	HasAnaesthesia	
Anaesthesia	Sedation		
Procedure	Bronchoscopy through artificial stoma	HasAnaesthesia	
Anaesthesia	General Anaesthesia		

Procedure	Fibreoptic colonoscopy to caecum	HasAnaesthesia	
Anaesthesia	Sedation		
Procedure	Wedge resection of ingrown toenail	HasAnesthesia HasAssociatedProcedure	Because of the limitation on the number of disease codes that can be accepted and the comparative unimportance of the diagnoses associated with these treatments, there are no associated diagnoses in the collection
Procedure	Surgical removal of tooth or tooth fragment	HasAssociatedProcedure	
Procedure	Treatment of acute periodontal infection	HasAssociatedProcedure	
Anaesthesia	General Anaesthesia		
Procedure	Transfusion of packed cells		
Procedure	Computerised tomography of brain		There are likely to be relationships between these tests and injuries / diseases. These relationships are unclear.
Procedure	Computerised tomography of spine		
Procedure	Computerised tomography of soft tissue of neck		
Procedure	Computerised tomography of abdomen		
Procedure	Retrograde micturating cystourethrography		
Procedure	Computerised tomography of chest, abdomen		
Procedure	Allied health intervention – occupational therapy		
Procedure	Allied health intervention – social work		
Procedure	Allied health intervention – dietetics		

Procedure	Allied health intervention – prosthetics and orthotics		
Procedure	Allied health intervention – physiotherapy		
Procedure	Allied health intervention – speech pathology		
Procedure	Allied health intervention – pastoral care		
Procedure	Allied health intervention – music therapy		

It is difficult to determine some of the relationships in this string of concepts. However it is clear that the use of relationships to group and associate conditions in the codes make the case easier to understand.

5.5.5 Case 5 Complication

Table 5-21 Case 5

5	**A complication – a case where there the reason for admission is not an injury and there is a code that indicates a complication of a procedure in the additional diagnoses. This case is used to analyse the content of a case that was admitted for a non injury and where a complication occurred during the patient's stay.**	**Retrieve all records where the principal diagnosis field does not start with an S or a T and where there is a diagnosis code indicating an injury (starting with an S or a T) in any of the other diagnoses.**
There were 38,164 cases meeting the criteria. The first returned entry had the following codes: Principal diagnosis: Person awaiting admission to another health care facility Additional diagnosis: 2: Congestive heart failure 3: Faecal incontinence 4: Unspecified urinary incontinence 5: Urinary tract infection 6: Contusion to other and unspecified parts of lower leg 7: Exposure to unknown factor 8: Place of occurrence of injury – health service 9: Unknown activity 10: Purulent endophthalmitis Procedure: Allied health intervention, dietetics		

Table 5-22 Relationships and entities in Case 5

Entity	Content	Relationship	Comment
Principal Diagnosis (service)	Person awaiting admission to another health care facility		
Clinical finding	congestive heart failure		
Clinical finding	Faecal incontinence		
Clinical finding	Unspecified urinary incontinence		
Clinical finding	Contusion to other parts of lower leg	HasExternalCause	
Cause of Injury	Unknown factor	HasPlaceofOccurrence HasActivityatOccurrence	
Place of Occurrence	Health service		
Activity	Unknown		
Purulent endophthalmitis			
Procedure	Allied health intervention – dietetics		

Though there are a number of procedures in this case, the model is appropriate to represent it, as all entities and relationships required are included in the model. Again it is difficult to ensure what relationships are required. It should be noted that some of the clinical findings are not associated with any other disease or procedure.

The literature review also identified the issue of transience. This issue related to the previous recording of post operative conditions, such as postoperative vomiting, which are no longer included in the collection unless they persist at discharge or for more than 7 days. A transient entity has been added to the model to allow the identification of transient conditions. By including this concept it would be possible to allow these data to be included in the collection, but excluded from casemix calculations, thereby meeting the requirements of epidemiology and finance/planning.

5.6 *Harmonisation of modifications identified*

At completion of this review the concepts required to represent morbidity reporting are included in the model shown in Figure 5-2.

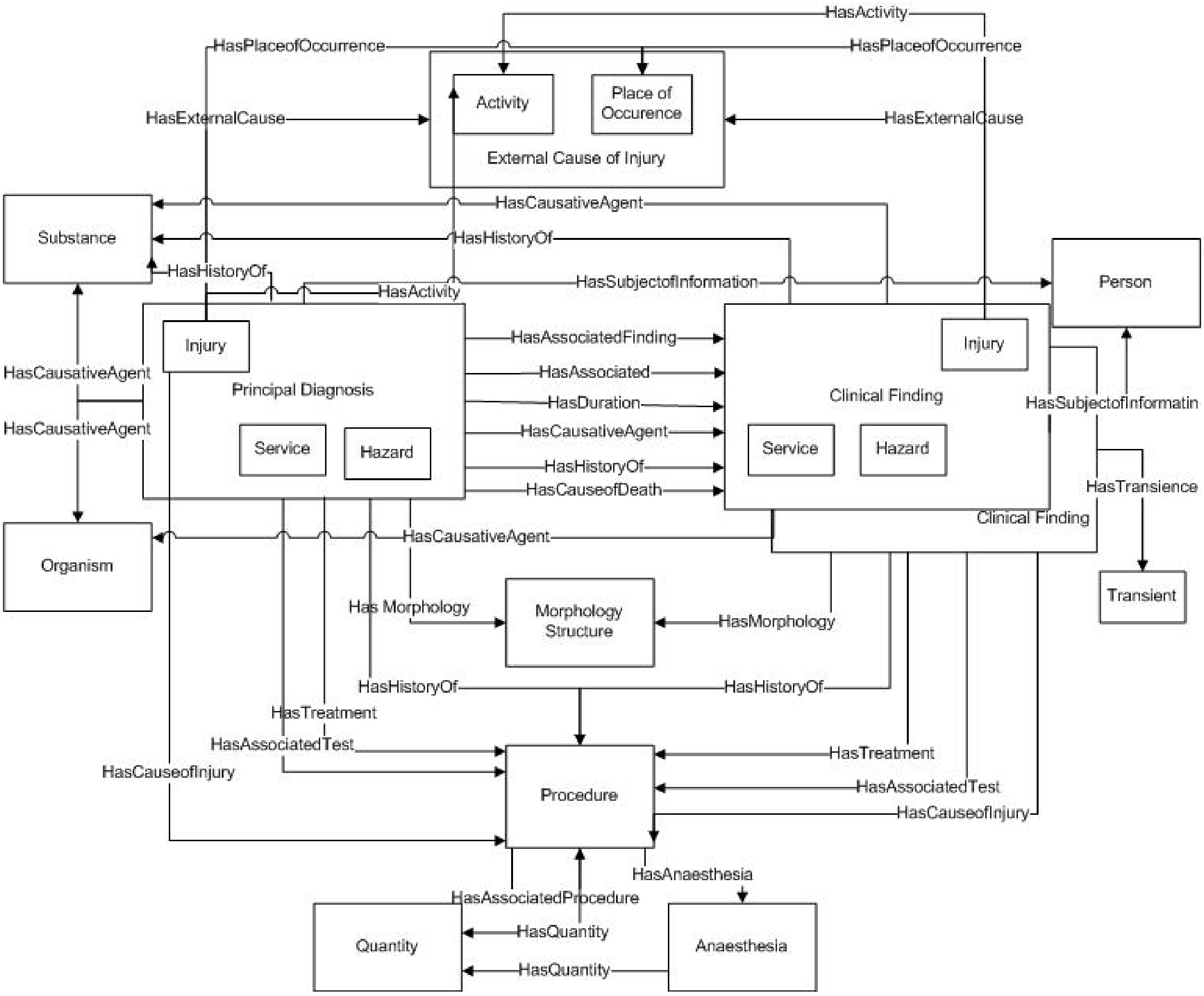

Figure 5-1 Final morbidity data model

The components of morbidity data have been derived from metadata and collection rules to produce an ontologically based model of clinical admitted episode morbidity data in Australia. This refined model included the following entities:

Principal diagnosis with sub entities of

Injury, Service and Hazard

Clinical finding with sub entities of

Injury, Service and Hazard

Morphological Structure

Organism

Substance

External Cause of Injury

With additional components of Place of Occurrence and Activity

Person

Procedure

Anaesthesia

Transience

Quantity

The set of relationships required to represent the data effectively are:

HasCausativeAgent

HasExternalCause

HasPlaceofOccurrence

HasActivityatOccurrence

HasAssociatedFinding

HasAssociated

HasMorphology

HasHistoryOf

HasDuration

HasCauseofDeath

HasAssociatedTest

HasTreatment

HasSubjectOfInformation

HasAnaesthesia

HasAssociatedProcedure

HasTransience

HasQuantity

5.7 Comparison to EHR Models

The morbidity data collection is an abstraction of the clinical environment represented by ontologies such as SNOMED-CT and OpenEHR Information Model. After identifying the entities of the morbidity data collection these were compared to the high level concepts in SNOMED-CT and the elements of the OpenEHR Information Model to identify similarities and variations thereby indicating where potential relationships may support morbidity data collection in the future

5.7.1 SNOMED-CT

The top levels of the SNOMED-CT ontology hierarchy were compared to the entities in the morbidity model. The results of this analysis are provided in Table 5-23

Table 5-23 High level concepts in SNOMED-CT

SNOMED-CT Concept	Morbidity Entity	Comment
Substance	Additional Codes used to indicate substance as the cause of injury but not as a separate concept	This information is of relevance to the identification of causes of injury that relate to specific substances and is included in the morbidity model
Linkage concept	Not required	These are covered by semantic linkage concepts in the model
Specimen	Not required	Specimens are not included in morbidity data collection
Body Structure	Represented as location of disease.	The structure of ICD-10-AM incorporates body structure into the coded concept. It is therefore not represented as an entity in the model
Staging and scales	Used to represent some forms of cancer morphology	Like body structures, staging and scales are incorporated into the ICD-10-AM coded concept.
Physical object	Related to cause of injury, such as struck by 'ball' where the ball is the physical object.	Included in the morbidity concept of cause of injury
Event	Events such as admission	The morbidity model includes service events, but does not include any other event details amongst the clinical information of morbidity data
Environment	Can be used to indicate place of occurrence of injury	Incorporated into the morbidity model as Place of Occurrence

Qualified value	Not a separate concept in morbidity data collection.	There are no concepts equivalent in meaning to this concept in morbidity reporting, though the concept is included in some individual codes.
Organism	Coded HasCausativeAgent to define infections and other parasitic, fungal agents etc.	This concept is included in the model.
Special concept	Awaiting admission to nursing home	This covers many concepts, particularly administrative concepts that are included in morbidity reporting. These are represented in the Service entity of the model
Pharmaceutical / biological product	Coded HasCausativeAgent to define infections and other parasitic, fungal agents etc.	Substances are included in the model, and these include pharmaceutical and biological agents
Clinical finding	There majority of morbidity codes fall into this category, diseases and symptoms of disease	Included as clinical finding in the model
Record artefact		There were no cases in the morbidity reporting that fell into this group of concepts, though one could consider that the morbidity record is an artefact
Social context	There are many morbidity concepts that include these conditions, such as lives alone.	Social context is included in the Hazards element of morbidity reporting. These issues are only included in morbidity when they represent a potential health risk. This is a good example of the difference between the requirements in the domain of morbidity reporting and that of clinical care.
Procedure	After clinical findings, this group is the largest group of morbidity concepts	This concept is included in morbidity reporting.
Physical force	Eg hit by	Included, with physical object, this concept is part of the external cause of injury I the morbidity reporting data

Though many of the codes represented in the Hazard entity in morbidity reporting are collected in SNOMED-CT they are represented within the clinical findings heading of SNOMED-CT. The only entity that would be difficult to identify from SNOMED-CT concepts is

Transience. The concept of postoperative vomiting (one of the conditions where inclusion in the collection is dependent upon persistence) has the option to specify episodicity of 'ongoing' but this concept does not have the same meaning as persistent vomiting postoperatively. The persistence of the condition is a definition that relates to definitions for data collection purposes and SNOMED-CT is intended to represent concepts for direct patient care, further investigation could be undertaken to investigate the variations in structure necessary to support the requirements of data aggregation and summarisation for morbidity reporting.

There is a strong similarity between the structure of the concepts in SNOMED-CT and those in the morbidity data collection, particularly when considered from the perspective of morbidity being a selected picture of the entire clinical event and in some cases an aggregation of concepts. The detail provided by SNOMED-CT as a representation of clinical content would support the representation of the entities in morbidity reporting.

The relationships in SNOMED-CT are designed to define the concepts accurately and unambiguously representing each individual clinical concept. The relationships in the morbidity model indicate relationships not to medical knowledge, but to the individual person's 'story'. In this sense the morbidity concepts and those of a terminological ontology such as SNOMED-CT are quite different. SNOMED-CT provides a sound basis for the generation of morbidity data, but the context of the information is required to gain full utility of the information for the purpose of morbidity extraction. Further investigation of this potential could be undertaken

5.7.2 OpenEHR

The OpenEHR information model provides a high level representation of the clinical information documented in an EHR. The OpenEHR information model can be represented in more detail using Archetypes which represent all the data of relevance in the clinical record in details using SNOMED-CT or other clinical terminologies. The archetypes provide context to the terms in the record and to the events of patient care. The different scopes and objectives of the different ontological perspectives provide a richness for consideration and further analysis. The concepts of the OpenEHR information model are compared to the Morbidity Model in Table 5-24.

Table 5-24 Comparison of OpenEHR Information Model to Morbidity Entities

OpenEHR	Morbidity Equivalent
Composition	Non Clinical Information is included in this concept; it would also indicate some relationships such as history. The composition has the capacity to define the concepts that occurred within an episode of care. This is an important structural element when identifying morbidity data to be included in a collection
Section	Sectional components of the OpenEHR all other identification of specific purpose data, rather than data that fits generically into the clinical knowledge of the individual. This is useful in morbidity reporting as it supports the clinical decision of principal diagnosis and cause of death.
Structure	Structure could represent groups of entities such as multiple injuries which occur as a result of a single cause of injury, place of occurrence and activity, or multiple procedures undertaking with a single anaesthetic.
Clusters	Morbidity data does not use the concept of a cluster, except for the instances of disease and manifestation pairs, represented by the HasAssociatedFinding, HasDuration
Action	This component of the information model includes details of procedures, referrals etc. Selected procedures are required in morbidity reporting and this structure is consistent with the capacity to represent these concepts.
Evaluation	This would indicate specific conditions and the relationships between them supporting improved relationship reporting in morbidity data
Instruction	There are few instructions in morbidity reporting. However there is a requirement that to include in morbidity reporting concepts such as – awaiting a bed. If an instruction is recorded in the record and the patient must wait in hospital to receive the required attention this can be coded and reported. These components are included in the service entity of the morbidity model.
Observations	Some observations are required to be included in morbidity reporting to more clearly describe a given condition through the associated findings relationship.

Additional entities such as hazards, organisms, substances, causes of injury are all able to be recorded in the more detailed archetypes. The issue of transience is currently not indicated in the OpenEHR model, however this model is in development and the opportunity to extend it to meet this need exists. It could also be possible to use associated compositions and the dates recorded as part of the composition to derive the currency of clinical data and thereby automate the allocation of transient status where relevant.

As morbidity data collection is sought from EHRs data repositories the terminology represented by SNOMED-CT requires the contextual structure provided through OpenEHR archetypes to describe and reference the clinical data in a manner that supports relationships required for the generation of morbidity reporting. This research indicates that this approach would be fruitful but this would require further investigation to confirm the detailed requirements and strengths and weaknesses of this approach.

Chapter 6 Conclusion

The hypothesis of this research was that an entity relationship ontology model could be used to define the structure components of the clinical components of admitted patient morbidity data. This research developed such a model and when tested against both governance concepts and actual data from the collection this model was proven sound. The research identified the structural components of the data collection. The model's entities and the semantic relationship which express the knowledge in an individual person's morbidity data have been demonstrated to be able to represent live morbidity data.

The process of ontology development was used and information has been learnt about the process as it relates to health information and about the existing ontological models of clinical data used in the electronic health record environment. Detailed outcomes are now summarised in accordance with the research aims, objectives and specific questions posed demonstrated that all have been achieved as a result of undertaking this study.

6.1 Entities and relationships in the current admitted episode morbidity data set

The research identified the entities and relationships included in the current data collection.

6.1.1 Data elements in the National collection

The data elements required in the collection were identified and found to represent a complex domain of concepts in the diagnosis and procedural data fields. These data structures offered little structure and did not support the identification of meaning in the data. The representation of the data in single fields with multiple coded entries does not support computer query of the data and therefore does not take advantage of the capacity of computerisation to represent knowledge and support both knowledge re-use and accessibility.

6.1.2 Data structure represented by the coding instructions and state collections

This study has demonstrated that it is possible to represent the components of morbidity data through the use of an ontological approach applied to the structure of that data. The components of data identified through a review of the governance documentation from the States and the coding instructions used when morbidity data is collected were identified the structure of the data included in morbidity data. These were expressed as entities (see section 4.6).

Each of these entities represents a different concept within the morbidity collection, concepts that describe the clinical aspects of the individual's care and provide a structure suited to the representation of that data using ICD-10-AM and the additional algorithms applied (AR-DRGs). The requirement for other entities was derived from analysis of the content of the data. Quantity is an entity clearly required to improve the capacity to represent multiple procedures in a clear and concise manner, while Transience allows the capacity to differentiate between information that is to be used in the calculation of AR-DRGs and that which is not.

6.1.3 The relationships between entities

Relationships between the entities were identified and expressed ontologically. These relationships are listed in section 4.6. These relationships were tested in a small selection of real world cases and shown to be appropriate to represent the meaning within that collection, as far as it is possible to do so given that the original collection does not have any instruction and much meaning has to be assumed. The process of undertaking an ontological review of the data in the actual morbidity data collection could be extended to test the model further.

6.1.4 Variations between State collections

The instructions provided in State collections indicated a desire to identify some of these entities individually as they had extracted them from the National grouped data structure.

Variations in the State data collection indicated a need consistent with the entities that had been identified. The variations were found not to reflect a difference in the actual content of the information collected from a clinical perspective, but they do indicate the need for clear relationships and entities to represent the different concepts in the collection in order to extract and analyse that data effectively. The results of this research would inform further development of morbidity collection to meet this need.

The following issues were identified during the study:

- The clear tensions regarding what should and should not be included in the collection should be addressed in the design of the system. This tension comes from the range of users of the data, particularly the use of data for analysis of hospital casemix using Diagnosis Related Groups. The Diagnosis Related Group computer algorithm is applied to morbidity data, particular to principal diagnosis and additional diagnoses to determine whether a case is complicated by significant co-morbidity. This process necessitated the exclusion from the morbidity data collection system of some 'transient conditions". The capacity to indicate whether a condition is to be included in the AR-DRG calculation has not been included in the model as this is not a strictly clinical component of the data, however further investigation could be undertaken to add this to the model. The model proposes that the capacity to indicate whether the status of a clinical finding is enduring (long term) would allow the differentiation between these two situations and has therefore addressed this specific issue but further investigation of the variations in the requirements of the users of this data could be undertaken to improve the model provided here.

- One of the opportunities offered by technological advances is that a sound ontologically structured data collection should be able to identify the inclusion and exclusion characteristics for morbidity data collection and allow automated extraction of those components from EHR data. This study has introduced this issue by considering the structure of the morbidity data, but further study is required to identify the capacity to automate the data collection.

6.2 Entity-Relationship model of clinical morbidity data

An ontologically based model of clinical morbidity data was developed and tested. That model is presented in Section 4.6. The ontological approach to the development of this model was useful as it permitted comparison and evaluation of the relationship between morbidity data and the Electronic Health Record from which that data is extracted.

The confusion and proliferation of models is often seen as a problem. In this research the availability of different views of the world provided by ontological models of patient findings and problems, OpenEHR information models was useful and assisted the result of the research.

This approach was particularly useful as the traditional approach of establishing a firm boundary around the domain can risk a scope so firm that it restricts the ability to recognise the need to function not only within the scope of the specific system needs, but within the wider world. This research did not investigate the value of other models, nor was a detailed research undertaken into the stakeholders' requirements for morbidity data but it is clear that the different perspectives need to be able to function in a more integrated manner if we are to gain maximum advantage of the EHR systems being introduced.

The methodology strengths were the use of existing ontologies as a starting point and the availability of an actual documentation about the data collection combined with the use of real data. This holistic approach to the development of ontology using the abstraction of exiting models, the governance information and the real world was a great strength of the methodology.

The weakest element of this research was the complexity of the topic. Further research is required to confirm the model, particularly through more detailed analysis of the data in the real world data collection. The resultant model could also be improved by a review of the actual requirements of the users, as opposed to the representation of their requirements assumed through the existing data collection and the documentation about that collection.

6.3 Comparison of the model to existing ontologies

The model developed was compared to the concepts in existing health ontology models. This comparison identified the utility of such a review, but also highlighted the difference in perspective of the world as represented by the domain of the ontology. The patient based ontologies of OpenEHR, The Framework of Clinical Findings and Problems and Morbidity reporting have similarities in their structure that support the identification of information about the individual and the event or contact with clinical care that is being described, while the whole of medicine views of SNOMED-CT and ICD-10-AM are useful to describe the individual medical ideas in the data collection but less useful in describing the relationships between those ideas.

6.4 Utility of the findings and results

6.4.1 EHR Opportunities

Australian healthcare is investing heavily in infrastructure to support the development of electronic health records {NEHTA, 2006}. This study is significant because it will identify the potential structure of information for hospital admitted episode morbidity data that will support improvements in Australia's morbidity data collection system at a time when the issues of clinical terminology, electronic health records and health data collection are being reviewed.{Health Data Standards Committee Secretariat, 2007}

The strong relationships between the ontological structure of morbidity data and the structures representing the patient's clinical environment were found which suggests that EHR technology may be used to improve the collection of morbidity data. However, this is only true if the data structure required for morbidity reporting is identified early. For example if there is no capacity in the electronic system to indicate that a procedure was conducted for a given disease (clinical knowledge often obvious to the clinician, but to be computable this needs to be in the system). Traditional morbidity collection systems have not changed in design since their paper based collection origins. To achieve the National objective of retrieving morbidity data as a by-product of EHR systems, the entities and relationships of the morbidity model need to be either incorporated into the EHRs concepts or be able to be derived from them. This is considerable further research required to investigate this possibility if morbidity data is to be automatically produced as a by-product of EHR systems. This automation is clearly not as simple as converting individual SNOMED-CT concepts into ICD-10-AM concepts.

6.4.2 ICD-10-AM

ICD-10 is being reviewed internationally to prepare the next version of the international classification – ICD-11. The availability of a model of morbidity and the potential to build systems that can collect data using this model could inform the development of ICD-11.

6.5 In Conclusion

This research has developed an ontology based representation of clinical morbidity data. This model can serve as a basis for further development of morbidity reporting systems at State and National level in the environment where EHR systems are being introduced throughout Australia and for the development of the medical domain classification used to represent the concepts within the morbidity collection ICD-11.

Methodological use of multiple ontological domains to inform the development of an ontology in a related domain has proven useful, as has the incorporation of information through the whole data collection continuum. This approach provided a more robust solution and highlighted the need to understand the barriers and overlapping concepts in the models of EHR systems to improve each of those models and better understand the purpose served by each model.

Further research should be undertaken to extend the model developed here, particularly to test it against a broader set of real world data, and to return to the origin of the data extracted into the data collection to identify the 'real' relationships prior to extraction in order to better understand the world being described. An understanding of the requirements of morbidity data sought directly from the stakeholders would be useful as the requirements are currently assumed from the documented governance procedures as a representation of requirements; this research recognises that this is only one view of those requirements.

6.5.1 Ethics approval

This research did not require ethical approval to proceed as it involved the investigation of publicly available data

6.5.2 Intellectual Property

This research complies with the requirements of CQU's policy on Intellectual Property.

Appendix A– Organisations

Organisation	Background
Australian Institute of Health and Welfare	The Australian Institute of Health and Welfare (AIHW) is Australia's agency for health and welfare statistics and information.

Clinical Advisory Group	The rules and standards for coding of Australian morbidity data are approved by the Clinical Advisory Group. Each State has its own advisory group, and each of these provides direction for coding in that jurisdiction. Each of these committees is represented at the National level for management of National rules and standards. The terms of reference of the National Coding Standards Advisory Committee (CSAC) are: " 1. *Advise the NCCH on the implementation and publication of new and amended ICD-10-AM codes and Australian Coding Standards.* 2. *Advise the NCCH on activities and products relating to coding and coding quality measures.* 3. *Report to and from organisations/jurisdictions represented on this committee.* 4. *Ensure that standards of definition and convention are maintained when ratifying changes to ICD-10-AM and the Australian Coding Standards.* 5. *Review public submissions for changes to ICD-10-AM.* 6. *Receive feedback from users of coded data on the impact of standards and codes on current data collections.* 7. *Ratify coding advice from the NCCH prior to publication in Coding Matters.* 8. *Recommend to Health future changes to the Australian Refined – Diagnosis Related Groups classification system as they relate to coding.* 9. *Recommend to National Health Information Management Group the national adoption of ICD-10-AM modifications on a biennial basis.* 10. *Provide input to relevant authorities on morbidity and mortality coding related issues such as data edits, coding quality measurement, design of data collection systems.* 11. *Provide coding advice to the National Health Data Committee on definitions relating to relevant classification items in the National Health Data Dictionary.* 12. *Provide advice to NCCH and the Australian Bureau of Statistics on the relationship between Australian Coding Standards for morbidity coding and rules for cause of death coding.* 13. *Provide advice on other relevant health classification systems"* (NCCH, 2006).

Health Data Standards Committee (HDSC)	*"The Health Data Standards Committee (HDSC) is a standing committee of the National Health Information Group - a body established under the Australian Health Ministers Advisory Council to oversee development of health data standards. All data element definitions to be included in the National Health Data Dictionary (NHDD) require approval by the Health Data Standards Committee and endorsement by the National Health Information Group.* *The Health Data Standards Committee has three major roles.* *The first is to assess data definitions proposed for inclusion in the NHDD (available on the AIHW page) and to make recommendations to the National Health Information Group on revisions and additions to each successive version of the Dictionary. Expert groups develop the data definitions and work to the HDSC when developing proposals for their inclusion in the NHDD.* *The second is to assess the business case for the modification or creation of endorsed data definitions for use in National Minimum Data Sets (NMDS). Recommendations are submitted to the Statistical Information Management Committee (formerly the National Health Information Management Group) which has the final approval.* *The third function is to assess the creation and modification of classifications and terminologies as proposed by the Classifications and Terminologies Working Group.* *The Committee also makes recommendations to the Statistical Information Management Committee (SIMC) to endorse new (NMDS) or modifications to an existing NMDS (Important note: see business rules for special requirements)*" (AIHW, 2006d).

National Centre for Classification in Health	The NCCH (Sydney) was established in "*March 1993, when T The then Commonwealth Department of Health, Housing, Local Government and Community Services called for tenders from interested groups to establish an Australian coding authority, funded by the* ***Casemix Development Program****. This followed the recommendations of the Patient Abstracting and Coding Project Report by K Eagar and K Inns (1992), which included the first version of the Australian National Coding Standards for Admitted patient Data Collections. The National Coding Centre (NCC) was established as an independent body by the School of Health Information Management, in December 1993. The NCC and NRCCH became 'NCCH' under a joint agreement on 1 January 1997. NCCH (Sydney) continues to be funded by the present. Prior to the establishment of the former NCC, new codes for use with ICD-9-CM were developed in the USA and coding standards were determined at a state level within Australia*"{National Centre for Classification in Health, 2005}. **"*Our vision*** *The NCCH enhances health outcomes through our contributions to clinical terminologies and statistical classification development that underpin health information knowledge systems. We value our team, whose commitment to enhancing our business processes and to developing health information solutions for a diverse and growing user base, ensure outcomes that meet users' needs*" (NCCH, 2006) ***We will realise our vision by:*** • *continuing to develop, refine and support ICD-10-AM and other classifications on behalf of Australia's Department of Health and Ageing, for the people of Australia* • *taking a leadership role in clinical terminologies and health classification systems development* • *providing a centre of theoretical clinical terminologies and health classification systems excellence* • *investigating, analysing and developing IT strategies to optimise current and planned activities* • *developing the knowledge and skills of our team* • *providing education and training to ensure best benefits from our products for their users* • *contributing to health data quality initiatives* • *maintaining our commitment to research and development in our own and allied fields* • *providing consultancy advice in all aspects relating to health information management systems* • *initiating and participating in international collaborative efforts* • *publishing our work in formats and media that are state-of-the-art and designed to meet users' needs* • *continuous monitoring of health sector developments* • *remaining focussed on our users' needs and demands* • *developing terminology and classification tools for specific sections of the health community*

Nehta	*Nehta's origins lie in a meeting of Health Ministers on 29 July 2004, at which the Ministers endorsed the immediate formation of a National E-Health Transition Authority team, responsible for establishing a new national health information management and information and communication technology (IM&ICT) entity and, simultaneously, progressing work on the most urgent national IM&ICT priorities. The establishment of this team was a demonstrable reaffirmation by the Ministers of the importance of IM&ICT to the health sector.* *On 5 July 2005, the National E-Health Transition Authority team became NEHTA Limited, a not-for-profit company limited by guarantee, with continuing responsibility for developing national health IM&ICT standards and specifications. NEHTA Limited is jointly funded by Australian state, territorial and national governments and the Board of NEHTA Limited are comprised of chief executives from health departments within these jurisdictions*." (National E-Health Transition Authority, 2006)

Reference

ABBOTT, B. (2001) Requirements set the mark. *InfoWorld.* Info World Media Group

ACUTE HEALTH DIVISION (1993) Input Transactions - Diagnosis Record. *PRS/2 Manual.* Victoria, Department of Human Services.

AIHW (2004a) Australian Family of Health and Related Classifications Reference classification. Canberra, AIHW

AIHW (2004b) MetaData Online Registry (METeOR). www.meteor.aihw.gov.au

AIHW (2005a) Australian hospital statistics 2003 - 2004. Canberra, Australian Institute of Health and Welfare , www.meteor.aihw.gov.au

AIHW (2005b) Metadata Online Registry (MeTEOR). Canberra, Australian Institute of Health and Welfare, www.meteor.aihw.gov.au

AIHW (2006a) Australia's Health. Canberra, Australian Institute of Health and Welfare

AIHW (2006b) About the Australian Institute of Health and Welfare. Canberra, Australian Government

AIHW (2006c) National hospital morbidity database (NHDB). Canberra, Australian Institute of Health and Welfare, http://www.aihw.gov.au/hospitals/nhm_database.cfm

AIHW (2006d) The National Hospital Morbidity Dataset. Canberra, Australian Institute of Health and Welfare

AIHW (2007a) A picture of osteoarthritis in Australia. Canberra, Australian Institute of Health and Welfare

AIHW (2007b) The National Hospital Morbidity Data Service. Canberra, AIHW

AIHW (2007c) Australian Health Statistics 2005-2006. Canberra, Australian Institute of Health and Welfare

AIHW & COMMISSION FOR SAFETY AND QUALITY IN HEALTH CARE (2007) Sentinal events in Australian public hospitals 2004-2005.

ALEXANDER, I. & STEVENS, R. (2002) *Writing Better Requirements*, Addison Wesley.

ARTIM (2003) Association (Semantic) Relationship. IN ARTIM, J. M. (Ed., PrimaryView.org

AUSTRALIAN CENTRE FOR ASTHMA MONITORING (2008) Statistical methods for monitoring asthma. Canberra, AIHW

AUSTRALIAN HEALTH CARE AGREEMENT REFERENCE GROUPS (2002) A report to the Australian Health Ministers' Conference.

AUSTRALIAN HEALTH INFORMATION COUNCIL (2008) AHIC E-Health Future Directions Briefing Paper. Canberra, Department of Health and Ageing March 2008

BEALE, T. (2001) Health Information System Manifesto. 2.5 ed.,

BEALE, T. (2006) Archetypes - An Interoperable Knowledge Methodology for Future-proof Information Systems. 2.2.1 ed.,

BEALE, T. (2007) Ontologies. OpenEHR

BEALE, T. & HEARD, S. (Eds.) (2007a) *Aims of the OpenEHR Architecture,* London, OpenEHR Foundation.

BEALE, T. & HEARD, S. (2007b) An Ontology-based Model of Clinical Information. IN KUHN, K., WARREN, J. & LEONG, T. Y. (Eds.) *MedInfo 2007.* Brisbane, International Medical Informatics Association

BOOCH G. ET AL (1999) *The unified Modelling Language User Guide,* Boston, Addison-Wesley.

BOWER, A. G. (2005) *The Diffusion and Value of Healthcare Information Technology,* Santa Monica, CA, Rand Corporation.

BRACHMAN, P. S. (2006) Public Health Surveillance. *Bacterial Infections of Humans: epidemiology and control.* 3rd ed., Springer.

BRAY, I. K. (2002) *An Introduction to Requirements Engineering,* Harrow, Pearson Education Limited.

BROWN, E. (2006) Fact Sheet Specifying Clinical Information. Sydney, NeHTA

BRYAN, M. (2004) NSW to Revamp Clinical System. *Australian Financial Review.* Sydney,

BUSH, G. W. (2004) Transforming Health Care: The President's Health Information Technology Plan. Washington, The White House

CAMPBELL, J. R., CARPENTER, P., SNEIDERMAN, C., COHN, S., CHUTE, C. G., WARREN, J. & CPRI WORK GROUP ON CODES AND STRUCTURE (1997) Phase II Evaluation of Clinical Coding Schemes: Completeness, Taxonomy, Mapping, Definitions and Clarity. *JAMIA,* 4**,** 238-250.

CAMPBELL, K. E., OLIVER, D. E., SPACKMAN, K. A. & SHORTLIFFE, E. H. (1998) Representing Thoughts, Words and Things in the UMLS. *JAMIA,* 5**,** 421-431.

CAMPBELL, R. J. (2004) Database Design: What HIM Professionals Need to Know. *Perspective in Health Information Management*.

CANADA HEALTH INFOWAY, (2008) Canada's EHR initiatives the benefit from SNOMED-CTCT terminology standard. Toronto, Canada Health Infoway

CEN ECFS (1997) Medical Informatics - Categorial structure of systems of concepts: Model for representation of semantics. Brussels, Belgium, European Committee for Standardisation
CEN TC251 WGII TERMINOLOGY AND KNOWLEDGE BASES (2005) Medical Informatics - Categorial structure of systems of concepts - Medical Devices. Brussels, European Committee for Standarization
CHEN, H., FININ, T. & JOSHI, A. (2004) An ontology for context-aware pervasive computing environments. *The knowledge Engineering Review,* 18**,** 197-207.
CHEN, P. (1976) The Entity Relationship Model - towards a unified view of data. *ACM Transactions on database systems,* 1**,** 1-36.
CHUTE, C. G. (2000) Clinical Classification and Terminology: some history and current observations. *JAIMA,* 7**,** 298-303.
CHUTE, C. G., COHN, S. P., CAMPBELL, K. E., OLIVER, D. E. & CAMPBELL, J. R. (1996) The content coverage of clinical classifications. . *JAMIA,* 3**,** 224-233.
COLLINS, H. (2002) *Enterprise Knowledge Portals*, Amacom.
COMMONWEALTH OF AUSTRALIA (2004) *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Australian Modification (ICD-10-AM),* Sydney, National Centre for Classification in Health, Faculty of Health Sciences, University of Sydney.
CONRICK, M. (Ed.) (2006) *Health Informatics: Transforming Healthcare with Technology,* Melbourne, Thomson Social Science Press.
DAILEY, L., WATKINS, R. & PLANT, A. J. (2007) Timeliness of Data Sources Used for Influenza Surveillance. *JAMIA,* 14**,** 626-631.
DATA SERVICES UNIT (2005) Queensland Hospital Admitted Patient Data Collection: Manual of Instructions and Procedures for the Completion of Patient Identification and Diagnosis Data. Brisbane, Queensland Health
DATA SERVICES UNIT (2007) *2005 - 2006 Queensland Hospital Admitted Patient Data collection,* Brisbane, Queensland Health.
DAVIDSON, S. M. & HEINEKE, J. (2007) Toward an Effective Strategy for the Diffusion and Use of Clinical Information Systems. *JAIMA,* 14**,** 361-371.
DENTAL HEALTH SERVICES VICTORIA (2001) Specifications - Electronic Patient Management System (EPMS). Melbourne,
DEPARTMENT OF HEALTH (2004) HMDS Reference Manual. Perth, Government of Western Australia
DEPARTMENT OF HEALTH (2006a) *Hospital Morbidity Data System Reference Manual,* Perth, Government of Western Australia.
DEPARTMENT OF HEALTH (2006b) Hospital Morbidity Data System Reference Manual. Perth, Government of Western Australia
DEPARTMENT OF HEALTH & AGEING Australian Health Care Agreement 1998 - 2003 Performance Report.
DEPARTMENT OF HUMAN SERVICES - VICTORIA (2003) Whole of Health Information and Communication Technology Strategic Plan 2003 - 2007. Melbourne, Victorian Government Department of Human Services
ELKIN, P. L., HEARD, S. & BEALE, T. (2004) The Relationship between Archetypes, Templates, Terminology and the HL7 RIM. Mayo Foundation for Medical Education and Research,
ENDRES, A. & ROMBACH, D. (2003) *A handbook of software and systems engineering: empirical observations, laws and theories,* London, Pearson Educational Limited.
ENDRES, A. & ROMBACH, D. (Eds.) (2006) *Handbook of research on informatics in healthcare and biomedicine,* Hershey, PA, Idea Group Reference.
FARRE, T. (2001) Get It Right the First Time - Improved requirements management is vital to the success of software development projects. *Information Week.* Information Week
FREAN, I. (2006) Modelling Health Care Information. IN CONRICK, M. (Ed.) *Health Informatics: Transforming Healthcare with Technology.* Melbourne, Thomson Social Science Press.
FREITAS, A. A. (2002) *Data Mining and Knowledge Discovery with Evolutionary Algorithms,* Berlin, Springer-Verlag.
GANGEMI, A., CATANACCI, C., CIARAMITA, M. & LEHMANN, J. (2006) Modelling Ontology Evaluation and Validation. *ESWC 2006.* Springer
GANGEMI, A., CATENACCI, C. & BATTAGLIA (2004) Information Ontology Design Pattern: An exercise in building a core biomedical ontology with descriptions and situations. *Ontologies in Medicine,* 102.
GASEVIC, D., DJURIC, D. & DEVEDZIC, V. (2005) MDA Standards for Ontology Development. *4th International Semantic Web Conference.* Galway, Ireland, Semantic Web Science Association
GEORGE, J. (2005) An ontology model for clinical documentation templates. *Massachusetts Institute of Technology*.

GLASS, R. L. (1994) *Software Runaways. Lessons learned from massive software project failures,* Upper Saddle River, NJ, Prentice Hall 1998.
GÓMEZ-PÉREZ, A. & BENJAMINS, V. R. (2002) Knowledge engineering and knowledge management : ontologies and the semantic web. *European Knowledge Acquisition Workshop (13th.* Sigüenza, Spain, Springer-Verlag
GOSTIN, L. O. & HODGE, J. G. (1998) The Names Debate: the case for national HIV reporting in the United States. *Alb. L. Rev,* 61**,** 679-743, 689-698.
GRAIN, H. (1979) *Patient Reporting System Manual,* Melbourne, Health Computing Services - Victoria.
GRAIN, H. (1980) *Patient Reporting System Manual,* Melbourne, Health Computing Services - Victoria.
GRAIN, H. (2005) Preparing for the new information environment. *HIMJ,* 34.
GRUBER (1993) Toward principles for the design of ontologies used for knowledge sharing. *Formal Ontology in Conceptual Analysis and Knowledge Representation.* Kluwer Academic Publishers.
GUARINO, N. & WELTY, C. (2001) Identity and Subsumption. Padova, Italy, LADSEB-CNR
HARRISON, J. & BERRY, J. G. (2007) Serious injury du to land transport accidents, Australia 2003 - 2004. Canberra, AIHW
HARTE-HANKS TRILLIUM SOFTWARE (2006) Data Quality Essentials: for any data-intensive project. DMReview.com
HAUSAM, R. R. (1996) Vocabulary, Coding, and Concept Representation. *Florida Family physician,* 46.
HEALTH COMPUTING SERVICES VICTORIA (1979) *Patient Reporting System Manual,* Melbourne, HCS.
HEALTH DATA STANDARDS AND SYSTEMS UNIT (2003) VAED User Manual. Melbourne, Department of Human Services
HEALTH DATA STANDARDS AND SYSTEMS UNIT (2004) VAED User Manual. Melbourne, Department of Human Services
HEALTH DATA STANDARDS AND SYSTEMS UNIT (2006a) VAED User Manual. Melbourne, Department of Human Services
HEALTH DATA STANDARDS AND SYSTEMS UNIT (2006b) VAED User Manual. Melbourne, Department of Human Services
HEALTH DATA STANDARDS AND SYSTEMS UNIT (2007a) VAED User Manual. Melbourne, Department of Human Services
HEALTH DATA STANDARDS AND SYSTEMS UNIT (2007b) Victorian Admitted Episode Data Set 2005 - 2006. Department of Human Services Victoria,
HEALTH DATA STANDARDS AND SYSTEMS UNIT, ACUTE HEALTH DIVISION (1999) *PRS/2 Manual,* Melbourne, Department of Human Services, Victoria.
HEALTH DATA STANDARDS COMMITTEE SECRETARIATE (2007) Progress with approval of AHMAC-funded work items for 2007–08. Canberra, Australian Institute of Health and Welfare
HEALTH STATISTICS BRANCH - HEALTH INFORMATION CENTRE (2002) Hospital Morbidity Data System - Reference Manual. Perth, Department of Health - Western Australia
HEALTHCONNECT (2003a) Health Connect - Clinical Information Project: Interim Phase 1 Report. Canberra, Government of South Australia and Department of Human Services
HEALTHCONNECT (2003b) HealthConnect Project Plan 2003 - 2005. Canberra,
HEALTHCONNECT (2005) HealthConnect Implementation Strategy Version 2.0 June 2005 (Revised). Canberra, Department of Health and Aging
HENLEY, G. (2007) Hospitalised football injuries 2003-05. Canberra, AIHW
HL7 (2004) HL7 EHR System Functional Model: A major development towards consensus on electronic health record system functionality.
HL7 (2007a) Clinical Document Architecture Release 2. HL7
HL7 (2007b) HL7 RIM. IN HL7 (Ed., HL7
HL7 AUSTRALIA (2007) What is CDA? , HL7 Australasia
HOVENGA, E. J. S., TICKLE, K. & WINKLER, C. (2002a) Limitations of current health information systems. *HIC02.* Melbourne, Health Informatics Society of Australia
HOVENGA, E. J. S., TICKLE, K. & WINKLER, C. (2002b) Limitations of Current Health Information
HUFFMAN, E. (1994) *Health Information Management,* Berwyn, Illinois, Physicians' Record Company.
HUMPHRIES, K. H., RANKIN, J. M., CARERE, R. G., BULLER, C. E., F., K. M. & SPINELLI, J. J. (2000) Co-morbidity data in outcomes research: are clinical data derived from administrative databases a reliable alternative to chart review? *Journal of Clinical Epidemiology,* 53**,** 343-349.
IHTSDO (2007a) Our Standards - SNOMED-CTCT. Copenhagen, IHTSDO
IHTSDO (2007b) *SNOMED-Clinical Terms User Guide,* Copenhagen, IHTSDO.

IHTSDO (2007c) SNOMED-CT. 0707Intl ed. Copenhagen, IHTSDO
IHTSDO (2007d) SNOMED-CTCT. 20070731 ed., IHTSDO
INFORMATION COLLECTION AND MANAGEMENT (2006a) Inpatient Statistics Collection Governance. Sydney, Department of Health January 2008
INFORMATION COLLECTION AND MANAGEMENT (2006b) Section 5 - Coding Standards. *Inpatient Statistics Collection Governance.* Sydney, Department of Health.
INFORMATION COMMUNICATIONS TECHNOLOGY STANDARDS COMMITTEE (2004) Foundations for the Future: Priorities for Health informatics standards in Australia, 2005-2008. Canberra, National Health Information Group
INFORMATION MANAGEMENT AND REPORTING DIRECTORATE (undated) Western Australian Data Linkage. Perth, Department of Health WA
ISO (2002) Health Informatics - Controlled Health Terminology - Structure and high-level indicators. Geneva, International Standards Organisation
ISO (2003a) ISO/DIS 18104 Health Informatics - Integration of a reference terminology model for nursing.
ISO (2003b) Health Informatics - Reference Terminology Model for Nursing. Geneva, International Standards Organisation
ISO (2004) ISO/IEC WD 11179-3 Information Technology - Metadata Registries (MDR) - Part 3: Registry metadata and basic attributes. International Standards Organisation
ISO (2005) IHD 21731 Health Informatics: HL7 Version 3 - Reference information model, release 1. IN ORGANISATION, INTERNATIONAL STANDARDS. (Ed. V3 ed., ISO)
ISO (2007a) FDIS 13606-1 Health Informatics - Electronic Health Record Communication - Reference Model. Geneva, International Standards Organisation
ISO(2007b) 22789 Health Informatics - Conceptual framework for patient findings and problems in terminologies. Geneva, International Standards Organisation
ISO (2007c) Health Informatics - Criteria for Categorisation and Evaluation of Clinical Terminologies. DIS ed. Geneva, International Standards Organisation
JIMENEZ, L. G. (2006) Re-enhancing the entity-relationship model. *Data and Knowledge Engineering,* 56**,** 410-435.
LACEY, E. (2002) UK Collaborative Clinical Terminology Project: NHS Acceptance Criteria for SNOMED Clinical Terms. National Health Service Information Authority,
MCCORMICK, A., ROSENBAUM, M. & FLEMING, D. (1990) Socioeconomic characteristics of people who consult their general practitioners. *Population Trends***,** 8 - 10.
MERRIAM-WEBSTER (2007) Online dictionary. Merriam-Webster
METEOR (2005a) Additional Diagnosis. Canberra, AIHW
METEOR (2005b) Diagnosis. Canberra, Australian Institute of Health and Welfare
METEOR (2005c) Principal Diagnosis. Canberra, AIHW
METEOR (2005d) Procedure.
METEOR (2007) Admitted patient Care NMDS 2008 - 2009. Canberra, Australian Institute of Health and Welfare
METEOR (2008) Admitted patient care NMDS 2008-2009. Canberra, AIHW
MEYER, T., LEE, K. & BOOTH, R. (2005) Knowledge integration for description logics,. IN MVELOSO, M. & KAMBHAMPATI, S. (Eds.) *AAAI05, Twentieth National Conference on Artificial Intelligence* AAAI Press
MIGUEL E. RUIZ, O. B. E. L. P. S. (2005) Ontological research and its applications to the biomedical domain. *Proceedings of the American Society for Information Science and Technology,* 42**,** NA.
MOONEY, G. (2002) Information and casemix funding. *Health Information Management Journal,* 31.
MORITZ, V., SCORDEL, D., BRAITBERG, G. & HART, G. (2004) Clinical Information System Development and Implementation: The Austin Health Experience. *HIC.* Brisbane, HISA
NATIONAL CANCER INSTITUTE (2007) Dictionary of Cancer Terms. Washington, U.S. National Institutes of Health
NATIONAL CENTRE FOR CLASSIFICATION IN HEALTH (2006) Australian Coding Standards for ICD-10-AM. IN NATIONAL CENTRE FOR CLASSIFICATION IN HEALTH (Ed. The International Statistical Classification of Diseases and Related health Problems, Tenth Revision, Australian modification (ICD-10-AM) ed., National Centre for Classification in Health
NATIONAL CENTRE FOR CLASSIFICATION IN HEALTH NCCH (2005) Background Information. Sydney, National Centre for Classification in Health
NATIONAL CENTRE FOR CLASSIFICATION IN HEALTH (2000) Strategic Plan 2000 - 2005. Sydney, National Centre for Classification in Health

NATIONAL CENTRE FOR CLASSIFICATION IN HEALTH (2002) *The International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Australian Modification (ICD-10-AM),* Sydney, National Centre for Classification in Health.

NATIONAL CENTRE FOR CLASSIFICATION IN HEALTH (2004a) *The International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Australian Modification - 4th edition (ICD-10-AM),* Sydney, National Centre for Classification in Health.

NATIONAL CENTRE FOR CLASSIFICATION IN HEALTH (2004b) *The International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Australian Modification,* Sydney, Commonwealth of Australia.

NATIONAL CENTRE FOR CLASSIFICATION IN HEALTH (2004c) NCCH Annual Report 2002 - 2003. Sydney, National Centre for Classification in Health

NATIONAL CENTRE FOR CLASSIFICATION IN HEALTH (2004d) Objectives of the National Centre for Classification in Health. National Centre for Classification in Health

NATIONAL ELECTRONIC DECISION SUPPORT TASKFORCE (2002) Electronic Decision Support for Australia's Health Sector. Canberra, Department of Health and Aging

NATIONAL HEALTH INFORMATION MANAGEMENT ADVISORY COUNCIL (2001) *Health Online: A health information action plan for Australia,* Canberra, NHIMAC.

NATIONAL LIBRARY OF MEDICINE (2004) Medical Informatics. *Medical Subject Headings.* Washington, National Library of Medicine January 2005

NEHTA (2006a) Benefits Realisation Study for the Planning of e-Health Service Investment. Sydney, National E-Health Transition Authority

NEHTA (2006b) About NEHTA. Sydney, National E-Health Transition Authority

NEHTA (2007) Unified Modelling Language v2.0. *Nehta Standards Catalogue.*

NHIMAC (2001) Health Online: A health information action plan for Australia. 2nd ed. Canberra, Commonwealth of Australia

NHIMG (2003) National Health Information Model Version 2. Canberra, Australian Institute of Health and Welfare

NOY, N. & MCGUINNESS, D. (2001) Ontology Development 101: A guide to creating your first ontology. *Knowledge Systems Laboratory,* 2001.

NSW HEALTH (2001) ISC Instruction Manual Private Sector Inpatient Facilities. Sydney, NSW Health

NSW HEALTH (2003) NSW Health Data Collection Directory. Sydney, NSW Health

NSW HEALTH (2004) NSW Health Data collections: Inpatient Statistics Collection. Sydney, NSW Health

OLSON, J. (2002) *Data Quality: the accuracy dimension*, Morgan Kaufmann.

OPENEHR (2007) OpenEHR Archetypes. Melbourne, Australia, OpenEHR

OPENEHR FOUNDATION (2007) Archetype MindMap.

PISANELLI, D. M., AZACCAGNINI, D., CAPURSO, L. & KOCH, M. (2003) An ontological approach to evidence-based medicine and meta-analysis. *Medical Informatics Europe 2003.* Saint Malo, France,

PISANELLI, D. M., GANGEMI, A. & STEVE, G. (1999) Toward a standard for guideline representation: an ontological approach. *AMIA Fall Symposium.* Washington,

PISANELLI, D. M., GANGEMI, A. & STEVE, G. (2002) Ontologies and Information Systems: the marriage of the century? *Lyee Workshop.* Paris,

PROMETHEUS (2006) HealthWiz. Commonwealth of Australia

REPORT OF THE REGIONAL DIRECTOR TO THE REGIONAL COMMITTEE FOR THE WESTERN PACIFIC (1998) Fifty years of the World Health Organisation in the Western Pacific Region. Manila, World Health Organisation

ROBERTS, R. F. & HANSON, R. M. (2003) Medical records and population health. *Medical Journal of Australia,* 179**,** 277-278.

ROBERTS, R. F., INNES, K., LUXFORD, K. & WILLIAMSON, D. (1998) Evaluating the impact of ICD-10-AM. *Tenth Casemix Conference in Australia.*

ROBINSON, K. & SHEPHERD, J. (2004) Predicting the influence of the electronic health record on clinical coding practice in hospitals. *HIMAA Journal,* 32**,** 102-108.

ROMAN, I. (2006) Toward the universal electronic health record (EHR): an ontology for the EHR.

SAFRAN, C., BLOOMROSEN, M., HAMMOND, W. E., LABKOFF, S., MARKEL-FOX, S., TANG, P. C. & DETMER, D. E. (2007) Toward a National Framework for the Secondary Use of Health Data: an American Medical Informatics Association white paper. *JAMIA,* 14**,** 1-9.

SATZINGER, J. W., JACKSON, R. B. & BURD, S. D. (2005) *Object-Oriented Analysis and Design: with the unified process,* Boston, Massachusetts, Thomson.

SIMON, J. & SMITH, B. (2004) Using Philosophy to Improve the Coherence and Interoperability of Applications Ontologies: A field report on the collaboration of IFOMIS and L&C. *First Workshop on Philosophy and Informatics.* Cologne,

STANDARDS AUSTRALIA (2005) *AS5021: 2005 The Language of Health Concept Representation,* Sydney, Standards Australia IT14.

STEPHENSON, S., LANGLEY, J., HENLEY, G. & HARRISON, J. (2003) Diagnosed based injury severity scaling. *Injury Research Statistics Series.* Canberra, AIHW

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.

THE COMPUTER LANGUAGE COMPANY (2007) Entity. The Computer Language Company

THE NATIONAL OFFICE FOR THE INFORMATION ECONOMY (2003) Interoperability Technical framework for the Australian Government. Canberra, Australian Government

TONG, B. & STEVENSON, C. (2007) Co-morbidity of cardiovascular disease, diabetes and chronic kidney disease in Australia. Canberra, AIHW

TRAVAKS, G. (1975) Computers and Morbidity Data Capture in Victoria. Communication to Health Information Management Students ed. Melbourne,

TRELOAR, R. W. (2003) Clinical Resource Management: Driven by Data. *Journal of AHIMA,* 74**,** 27-29.

USTUN, T. B., JAKOB, R., CELIK, C., LEWALLE, P., KOSTANJSEK, N., RENAHAN, M., MADDEN, R., GREENBERG, M., CHUTE, C. G., VIRTANEN, M., HYMAN, S., HARRISON, J., AYME, S. & SUGANO, K. (2007) Production of ICD-11: the overall review process. Geneva, World Health Organisation

VICTORIAN CODING COMMITTEE (1983) PRS upgrade meetings. Melbourne, Department of Human Services - Victoria

WALKER, D. (2004) GP Vocabulary Project - stage 2: SNOMED-CTCT Report. Department of Health and Aging and General Practice Computing Group

WAND, Y. & WANG, R. (1996) Anchoring data quality: Dimensions in Ontological Foundations. *Communications of the ACM,* 39.

WANGLER, B., AHLFELDT, R.-M. & PERJONS, E. (2003) Process oriented information systems architectures in healthcare. *Health Informatics Journal,* 9**,** 253-265.

WEINER, E. (2006) Compact Oxford English Dictionary. Oxford University Press

WORLD HEALTH ORGANIZATION (2005) Library and Information Networks for Knowledge. World Health Organization