DESIGN CONSIDERATIONS OF A SEMANTIC METADATA REPOSITORY IN HOME-BASED HEALTHCARE

CECIL CLIFFORD VAN DER WATT

Cape Peninsula
University of Technology

Design Considerations of a Semantic Metadata Repository in Home-Based Healthcare

by

Cecil Clifford van der Watt

Thesis submitted in fulfilment of the requirements for the degree

Master of Technology: Discipline Information Technology

In the Faculty of Informatics and Design

At the Cape Peninsula University of Technology

Supervisor: Dr Retha de la Harpe

Cape Town

31 October 2011

DECLARATION

I, Cecil Clifford van der Watt, declare that the contents of this thesis represent my own unaided work, and that the thesis has not previously been submitted for academic examination towards any qualification. Furthermore, it represents my own opinions and not necessarily those of the Cape Peninsula University of Technology.

Signed **Date**

ABSTRACT

The research was conducted as part of a socio-tech initiative undertaken at the Cape Peninsula University of Technology. The socio-tech initiative overall focus was on addressing issues faced by rural and under-resourced communities in South Africa, specifically looking at Home-Based Healthcare (HBHC) primarily in the Western Cape.

As research into the HBHC context in rural and under-resourced communities continued numerous issues around data and data-elements came to light. These data issues were especially prevalent in relation to the various paper forms being used by the HBHC initiatives that attempt to deliver care in these communities. The communities have the tendency to suffer from poor access to formal healthcare services and healthcare facilities.

The data issues were primarily in terms of how data was defines and used within the HBHC initiatives. Within the HBHC initiatives that cater for rural and under-resourced communities there was a clear prevalence of paper-based systems, and a very low penetration of IT-based solution. Because similar and related data-elements are used throughout the paper forms and within different context these data-elements are inconsistently used and presented. The paper forms further obfuscate these inconsistencies as the paper forms regularly change due to internal and external factors. When these paper forms are changed date elements are added or removed without the changes to the underlying ontologies being considered.

These issues around inconsistent usage and representation of data-elements prevent HBHC from benefitting from the numerous advantages offered by Health Informatics and IT-based systems primarily in terms of interoperability and integration. Health Informatics has numerous potential benefits beyond interoperability which could help in addressing the many issues faced by HBHC in rural and under-resourced communities of South Africa. These same issues also prevent HBHC from efficiently and effectively providing the much needed level and quality of care.

The research thus looks at two problems prevalent in HBHC, specifically the limited penetration of IT and HI-based solutions in HBHC and secondly the issues surrounding data-elements. The research thus focused on the design and development of a tool intended to be used to capture the inconsistent definitions and representations of data-elements in HBHC. The intended tool is a semantic metadata repository which is able to create pseudo-ontologies, using metadata which represents the care-data-elements and their semantics. These pseudo-ontologies are

able to represent the care elements used in HBHC as well as their relations and context of these relations.

By looking at the design and development process the research also provides considerations that need to be taken into account when developing IT and HI-based solutions intended for the rural and under-resourced HBHC context. Hopefully increasing the likelihood of success in similar undertakings and indirectly increasing the penetration of IT and IS-based solutions in HBHC.

Because of the research is focused around design and development an IT-based artefact a Design Science Research (DSR) approach was used to conduct the research. DSR bridges the gap between research and practice and allows for software design and development, usually associated with practice, to be used for research purposes. An amalgamated DSR methodology is used to conduct the research, combining many of the more prevalent DSR methodologies found in the academic literature to ensure a high-level of research relevance and rigour.

To support the DSR process and to provide valuable insights and understanding for the design and development aspects of the research a supporting ethnographic study was undertaken to help better contextualise the problem area and the problem context. The ethnographic study was conducted at the Stellenbosch Hospice which provides HBHC-services to the Kayamandi Community just outside Stellenbosch in the Western Cape and at the Olive Leaf Foundation with provides HBHC-services to the Motherwell Community in the Eastern Cape.

The Design and Development produced a number of artefacts, ranging from initial conceptualisations, Entity-Relational-Diagrams and a prototype-artefact created using the Microsoft Windows Forms and the Microsoft WPF software development technologies. The research presents the design from conceptual-stage to the final prototyped-artefacts.

The artefacts produced during the research are evaluated using relevant criteria, found within the academic literature. The evaluation help to show the utility of the artefacts, thus providing the necessary research rigour and relevance but the evaluation also aided in producing the research findings.

The research produced a number of findings in the areas of: design science research, artefact evaluation and in terms of the artefacts themselves (the metadata and the repository).

The researched showed that in the academic body of knowledge design science research is usually not associated with a strong academic literature foundation. However the researched found this this to be problematic. Rather by basing the design, development and overall research process within a strong foundation of academic knowledge and by understanding the context of the problem and the context of the solution it helped to ensuring the success (in terms of relevance and utility) of the research and the artefacts produced by the research. Further few academic sources spoke of any particular design and development methodology (waterfall, agile, spiral etc.). Most academic literature sources spoke only of design science at a higher conceptual level, leaning either towards a strong research or strong practical interpretation. This research used an agile iterative approach to produce the artefacts as it better fitted the overall research goals and process. By iterating insight could be gained and reflected on and the impact could be gaged and if needed changes made.

In terms the evaluation of the artefacts produced by the research it became clear that no universal criteria could be created that would be sufficient to evaluate all the possible artefacts that could be produced by design science research. Rather the criteria had to be relevant to the particular artefacts, moulded around its unique characteristics and its intended uses. These criteria too had a strong foundation within the literature and the research / solution context.

In terms of the repository the research saw the importance of visualising the repository content (aiding in understanding, and limiting the complexity of the data that the user had to work with. The repository also needed to be designed around the intended users, the volume of information it would contain and the frequency of chances that would occur to the content.

In terms of the metadata the research found that the metadata model was fairly successful at capturing the home-based care data-elements at a meta-level (their structure, relations, context and semantics). Further the research found that the metadata allowed for the data-elements to easily be visualised. However the research saw that the metadata model had a trade-off, the more complex the less widely it could be used but the more general forms were less expressive and useful.

ACKNOWLEDGEMENTS

I wish to thank:

My mother, for her moral support.

Professor Retha de la Harpe for her diligent supervision and guidance.

Members and Colleagues of the Kujali Innovation Incubator, whose efforts contributed to the completion research.

Table of Contents

Index of Tables

Index of Figures

GLOSSARY

Terms/Acronyms/Abbreviations	Definition/Explanation
HBHC	Home-Based Healthcare
HI	Health Informatics
IT	Information Technology
HIS	Health Information System
EHCR	Electronic Healthcare Record
EHR	Electronic Health Record
CPR	Computerised Patient Record
HL7	Health Level 7
DSR	Design Science Research
NGO	None-Government Organisation
DS	Design Science
IS	Information System
BSR	Behavioural Science Research

Background

Chapter 1 Introduction

1.1 Background

The research was conducted as part of a South African Finland Partnership (SAFIPA) in Information Communication Technology (ICT) funded project conducted at the Cape Peninsula University of Technology (CPUT) in Cape Town, South Africa. The project overall goal is to address tensions and stressors, such as poverty and illiteracy, within rural and under-resourced communities. CPUT specifically focused on healthcare services and healthcare provisioning, which was known to be problematic within these communities before the project started.

A number of different projects were undertaken in partnership with several healthcare initiatives that cared to communities in need. Initially the project partnered the Stellenbosch Hospice, in the Western Cape, that provided care for the Kayamandi Township just outside of Stellenbosch and later with the healthcare initiative in Motherwell, in the Eastern Cape. Most of these projects were undertaken in an attempt to see how ICT can be used to solve a number of issues faced by the healthcare initiatives and healthcare workers, focusing on the usage of mobile technology, community education and data quality and usage.

From the interaction with these different healthcare initiatives it became clear that the usage of data and data-elements, and what constituted quality data, was different from one healthcare initiative to the next with little standardisation current available. The metadata repository project was specifically undertaken to create a means by which these differences within the interpretations of data could be captured and stored in a common place in a format that allowed it to easily be accessed and interpreted. The eventual goal of the repository is to compare these differences and possible to create a common model or common definition for the data-elements. However at this stage the repository is only in its second iteration of development and acts primarily as a common storage for the different elements using a semantic metadata model.

The research however does not look directly at the created repository (the product). The research assumes that the repository itself is specific to the intended context and cannot easily be used in different context without sufficient alteration and re-scoping thus the research focuses more on the design process and design consideration undertaken in the development of the repository.

The research primarily considers the technical design consideration, and does not heavily focus on the soft design consideration (such as usability). The reason for this particular focus is because it is assumed in this research that although both usability and functionality are important parts in creating quality software solutions it is possible to have something that is very usable but not functional or alternatively functional but not usable. This thus leads to the notion that there is a clear dichotomy exists between the two. Although there is no assumption made that either one of these is less or more important than the other and both play an important role in the overall success of a given software system.

1.2 Problem Statement

In South Africa there is a prevalence of paper-based systems being used by the various initiatives that attempt to deliver home-based healthcare services to rural and under-resourced communities (van Zyl, 2011). These paper based systems that are used to capture, store and manage data, but these systems also makes it difficult to see some of the underlying data issues, namely: that the way in which data is defined, interpreted and used varies from one home-based healthcare initiative to another. This obfuscation of data-elements by the paper systems and their context of use make it difficult to understand what is actually happening at the ground level and furthermore thus makes development of informatics solutions to meet various needs of the HBHC initiatives problematic. Overall there is a low penetration of IT in rural HBHC and a high failure rate amongst the few attempts made to bring IT into these areas.

1.3 Research Question

What is required to design and develop an appropriate repository for the HBHC context to adequately store and represent care data-elements at an ontological level?

1.4 Research Sub-Questions

Table 1.1 lists the research sub-questions along with the research methods used in order to answer these sub questions and the objectives. The objectives provide the reasoning for why a particularly question was asked and what ideally could be learned the particular sub-questions. By looking at the objectives it is hoped that the logic as to why these question are being asked can be shown as well as how these sub-question relate back to the primary research question.

Table 1.1: *Research Sub-Questions.*

Sub-questions	**Research method(s)**	**Objectives**
What are the contextual implications for designing a repository of HBHC data-elements?	Literature review. Ethnographic Study of the HBHC context.	To gain additional insight into the rural HBHC context of South Africa. To gain the necessary design requirements to successfully create a system that can be used in the South African home-based care context. To gain an understanding of how the HBHC characteristics impact design and development.
How are the care data-elements and the relationships between them appropriately represented	Design and Development based research. Literature review. Ethnographic Study of the HBHC context.	To determine an appropriate means representing care data-elements. To gain an understanding of the types of care data-elements and how they relate to each other.
How can the appropriate design considerations for a semantic metadata repository be identified within a research process.	Literature review. Design and Development based research.	To understand the best way to use the process of design and development as a means of research. To understand how, if at all, design knowledge and innovative thinking can be captured.

1.5 Objectives

An objective of the research is to look at the design considerations that go into the development of a facility-level IT-based solution intended to be used within the rural HBHC context of South Africa.

Another objective of the research is the creation of a prototype-version of a semantic structural-metadata repository to be used within the research context.

The research also attempts to detail some of the steps and activities that would go into the development of an IT based solution intended to be used within a context similar to the research.

The research will consider how a DSR methodology can be utilised in research by providing a practical example of design science research in practice.

A final objective is to gain a deeper understanding of HBHC within rural communities of South Africa.

1.6 Scope of Study

The research is focus on the development of a software based solution conducted using an iterative agile development methodology.

The research also focuses on the development of a system to be used at an institutional level, with the necessary infrastructure in place (computers, electricity etc.).

The overall research itself only focuses on the first iteration of the development process in which a prototype system was developed, as a proof of concept.

The research focuses at the analysis and fundamental design and development activities.

The human computer interaction and aesthetic elements of the solution are touched on by but are not a primary focus of the research.

1.7 Research Methodology

The research can be broken up into two different yet closely related parts, the overarching research and the underlying practical software development.

The research uses qualitative-pragmatic approach following a design science research paradigm. The research ontology is based-on a nominalism stance from a subjective viewpoint. The overarching research relieves on the development component to produce data and research results.

The practical component of the research is based-on the design and development of several artefacts, amongst other a software based system. The practical component of the research follows a pragmatic approach.

The overall DSR research methodology involves the following seven steps:

1. Problem identification, understanding and motivation.
2. Identifying the objectives/focus of the research and solution.
3. Concept design.
4. DSR artefact design and development.
5. Artefact evaluation.
6. Research contribution.
7. Communication.

1.8 Importance

During the course of the project, and as shown in the literature review in Chapter 2, there is clearly a lack of Information technology based solution implementation within the home-based healthcare environment, especially within middle-class and developing countries. The research is provide needed insight into which could possibly be used for future research or aid and guide individuals and groups that seek to implement information technology based solution within these communities. The research provides a practical example as well as clearly list and discuss the design and design consideration that were required within the development of a semantic metadata repository. These design considerations are presented in such a way that the context is clear and can be generalised and applied to similar projects and undertaken.

1.9 Delineation of Research

Research is focus primarily on HBHC within the Western Cape with supplementary findings form the Eastern Cape. The research does not involve an exhaustively look at the numerous care activities or at the patient involvement at the care process. The focus is on the institution and on the institution-caregiver interactions, with limited focus on institution-governmental body and caregiver-patient interactions.

It is far too ambitious and time consuming, to attempt to cover all aspects involved in designing and developing a semantic metadata repository.

The research is design and development-based, thus the research looks primarily at the technical aspects involved in the development of the system. Attention is only being paid to the human related aspects where needed as part of the design and development process of the repository.

The brief ethnographic study is used to address some of these issues, to help better understand the environmental factors, the problem and solution. The ethnographic study is not intended to be complete and comprehensive or a standalone component of the research. The ethnographic data collected is simply meant to help define the environmental factors and to clarify the complex issues surrounding the problem domain.

The design consideration addressed by the research is those design consideration closely related to the technical aspects of the system and not the usability aspects.

1.10 Intended Audience

The research considers the home-based healthcare context within South Africa but the target audience is not explicitly healthcare workers or healthcare employees, the thesis contains some technical information and a pre-cursory understanding of software development and database design is required. Although most of the information is presented in such a way that a person without such knowledge would still be able to understand most of the thesis and the work within.

1.11 Expected Outputs and Outcomes

The primary output of the research is a Master Thesis detailing the process and consideration that go into creating a Semantic Metadata Repository for the Home-Based Healthcare Context.

A secondary output is the actual prototyped Semantic Metadata Repository, intended to be used to understand the usage and definitions of data-elements used in Home-Based Healthcare.

The primary outcome of the research is a number of considerations that need to be taken into account when developing a Semantic Metadata Repository for the under-resourced Home-Based Healthcare Context. These considerations could also potentially be applicable to development of other types of software solutions within the same context.

Secondary outcomes include contributing to the academic body of knowledge relating to Design Science Research and a deeper understanding of the HBHC context within rural and underdeveloped communities of South Africa.

1.12 Layout of the Thesis

This thesis consists out of a total of ten chapters including this introductory chapter.

Overall the thesis' layout is based around a fairly rudimental structure namely: the first chapter provides an introduction to the thesis, the second chapter details the literature review, the third chapter details the research methodology utilised to conduct the research, from the fourth chapter onwards the research process is detailed final chapters details the research findings and contributions.

The chapters representing the research process were based-on the implemented research methodology, with each chapter representing at least one step in the methodology.

Figure 1.1 demonstrates visually demonstrates how the chapters of the thesis corresponds to the implemented research methodology and its phases seven phases.

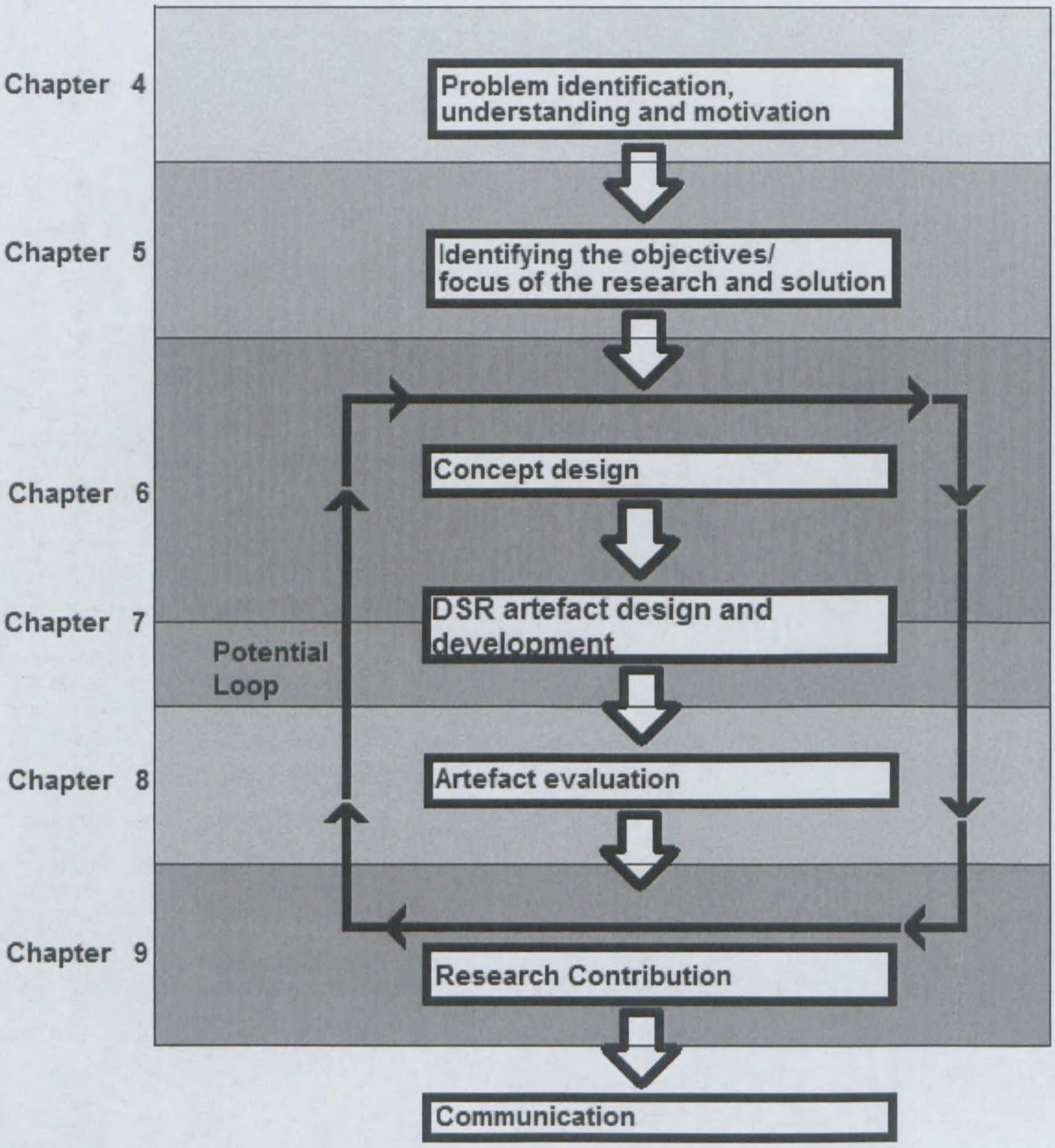


Figure 1.1: *Chapter Layout in relations to the Research Methodology.*

Chapter 4 details the research problem and motivates the need for the research. Chapter 5 Details the research objectives. Chapter 6 encompasses two phases in the research methodology The Concept Design phase and the design and development phase. Chapter 7 also encompasses the design and development phases, but is more focused on demonstrating the instantiated prototype artefact. Chapter 8 demonstrates the evaluation of the research artefacts created as part of the research. Chapter 09 identifies the research contributions.

The final chapter, chapter 10, is the conclusion which reflects on the research, answers the research question and identifies potential future research.

Chapter 2 Literature Review

In order to provide solid theoretical base for the research a literature review was undertaken with a strong component of technological due-diligence. The technological due-diligence involved looking at what technology and findings were available in the literature at the time and specifically what could be relevant to the research.

The literature review made contributions to several areas of the research. Firstly it helped to define the various concepts and terminologies used in the research area. Secondly the knowledge gained from the literature review was used to help define and refine the design science methodology used in the research. Thirdly during the research the knowledge gained from the literature review helped to contextualise the research as well as the problem domain. Finally the literature helped in designing and creating the various research artefacts and further helped to create the metrics used in the evaluation of the design science artefacts.

Most of the literature was sourced online using Google Scholar which in turn is linked to a number of online academic repositories, databases and catalogues such as ScienceDirect, ACM, IEEE and SpringerLink. Sourcing the literature online was a preferred method since it allowed for more expedient access to a large number of relevant articles and papers. Since the research touched on a large and diverse number of topics it was conceived that going through a number of printed journals would be more time consuming, and would not necessarily provide any additional insights.

The initial search for literature looked primarily at the topic of home-based healthcare (HBHC) and health informatics (HI). The literature sourced on the topic of HI was fairly general but provided a better understanding of some of the issues faced in HI domain. It was notable that interoperability was raised several times within the HI literature, and thus interoperability was looked at in more detail. Interoperability/integration literature touched on several other topics such as metadata and semantics. The literature on metadata and semantics eventually led to the topic of ontologies. During the later phase literature on the topic of repositories was briefly sourced in order to help clarify the intended solution artefact.

The structure of this chapter follows the above indicated logic with each Section providing an introduction that provides more information on the content and flow of that given Section.

2.1 Home-based Healthcare

This Section considers the available academic literature relating to HBHC. This Section does not touch heavily on technology or informatics used and/or required in HBHC but is rather meant to provide some insight into the HBHC context. Specifically this section focuses on under-resourced and developed areas of South Africa. It is important to first look at HBHC because the research was conducted within the HBHC context of the Western Cape and Eastern Cape provinces of South Africa and the literature provided an important starting point for further research and understanding. The sourced literature considers HBHC from a macro level briefly touching on the difference between HBHC in the developing and developed world and then concludes by looking specifically at the HBHC in the rural South African context.

The sourced literature on HBHC further helped to clarify and analyse the ethnographic findings produced as part of the research and helped to place HBHC within the greater healthcare environment. The literature also touched on the role and advantages of HBHC and some of the challenges faced by HBHC in order to show the importance and need for research addressing some of the issues faced by HBHC in countries such as South Africa.

Within the overarching healthcare context HBHC, also known as Home Care especially in the context of developed countries (Hongoro & McPake, 2004), is seen as being a subset of Primary Care (Wouters et al., 2009). HBHC is aimed at providing individuals and families within a given community with quality healthcare services, as the name implies, at the actual homes of the patients where it is needed most. HBHC also serves as a link between the patients and their community with the available facility and/or national healthcare systems (Strasser, 2003; Klaas, 2007).

HBHC and the overarching primary care rely heavily on nurses to meet its intended goals; the nurses serve as both the primary source of patient care and as the first line of contact between the healthcare facilities and with the patients and their community (Klaas, 2007). In the literature Padarath et al. (2003) saw nursing as being the key factor in the provisioning of HBHC services to those in need who could otherwise not make it to an healthcare facility or for whom long term stay in such facilities aren't practical or desirable.

By its nature HBHC decentralises the provisioning of care in that care is no longer provided at facilities such as clinics or hospitals but that the homes of patients. Because the care services are decentralised it can be adapted to specifically cater to the individual's and community's needs. Formal institutions however often have to cater to the healthcare needs of multiple communities and struggle to provide care for specific community needs. Thus the decentralisation means that HBHC can cater for the differing individual and community needs and provide the needed care services more effectively (Wouter et al., 2009).

The following Section considers the literature relating to the roles of HBHC in developed and developing countries and then finally looks specifically at the role that HBHC plays within rural areas of developing countries and in South Africa. Ideally this should provide an adequate amount of clarification as to the role and importance of HBHC and provide sufficient clarification of the context in which the research was conducted.

2.1.1 Home-Based Healthcare in Developed and Developing Countries

As has been shown in the section above, HBHC from at least a superficial point of view is the same in developing and developed countries. Simply put HBHC can be seen as: the provisioning of care at the home of a patient focusing on chronic and palliative care. However there are notable differences between HBHC in developed and developing countries when examined closely.

In order to better contextualise the research and to better understand the importance of HBHC specifically in developing countries such as South Africa, it is necessary to better understand the different roles that HBHC plays in developing and developed countries. This Section attempts to identify and discuss these differences between HBHC in developing and developed countries using the sourced literature.

Since the 1990s HBHC has started to play an increasingly important role in developing countries (Lewin et al., 2009) which face a number of healthcare crises such Acquired Immune Deficient Syndrome (AIDS) and tuberculosis (TB) (Evans et al., 2003; Fraser et al., 2005). These developing countries have seen an increase in the need to treat these and other chronic diseases but a failure of the formal healthcare systems to address the community's need for care (Lewin et al., 2009).

Developing countries have a number of unique characteristics which prevent the formal healthcare systems from meeting a community's healthcare needs. Some of these characteristics include: a notable lack of human resource including doctors, nurses and other healthcare professionals which helps to lead to a widening gap between healthcare supply and demand (Evans et al., 2003; Ludwick & Doucette, 2008); a lack of infrastructure such as water, electricity and transport, poor purchasing capacity of both patients and caregiving organisations; a lack of access to repair and maintenance services (Martinez et al., 2005) and poor organisation of current healthcare facilities and resources (Evans et al., 2003). All these mentioned factors collectively lead to issues surrounding the availability, provisioning and quality of healthcare services.

The number of available trained nurses to patients in the developing world is far more disproportionate than it is in the developed world to such an extent that the provisioning of healthcare services can be extremely difficult to provide (Hongoro & McPake, 2004, Litsios, 2004).

HBHC is novel in how it attempts to bridge the gap between supply and demand for healthcare. HBHC actively seeking to involve members of the community in helping to address the healthcare needs of the community in order to work around the insufficient number of available nurses (Litsios, 2004). HBHC in developing countries moves the provisioning of basic nursing care activities from strained formally-trained caregivers such as nurses to informally-trained caregivers such as volunteer community members (Schneider, 2008; Wouters et al., 2009).These informal caregivers are also sometimes called Lay health workers (Litsios, 2004).

Informal caregivers have been used throughout the world especially in developing countries for decades especially at the local level, dating back to the late 1970s. Although more prevalent in the developing countries, informal caregivers are also used in developed countries (Schneider et al., 2008)

The HBHC in the developing world attempts to address the inverse care law. The Inverse Care law according to Hart (1971) states that: "The availability of good medical care tends to vary inversely with the need for the population served... operates more completely where medical care is most exposed to market forces, and less so when such expose is reduced." The inverse care law simply states that those who need care services the most, especially those in poor communities where exposure to market forces are limited, normally do not have access to the needed level and quality of healthcare services.

The role that HBHC plays in the developing world is to attempts to bridge the gap between supply and demand for quality healthcare services, caused directly and indirectly by the various characteristics unique to developing countries, reducing the overall mortality rate. HBHC brings quality healthcare services into the communities which need them the most, communities which normally do not have access to more conventional forms of healthcare services and facilities, such as hospitals, clinics etc. (Strasser, 2003).

From the above it is clear that the literature sees HBHC in the developing world as addressing the lack of access to quality healthcare services within the communities of need. The literature also recognises the usage and importance of informal caregivers who only have basic training treating a range of illnesses that are usually less severe, and rarely terminal in nature.

In developed countries, as briefly touched before, there are a number of unique characteristics which impact the role HBHC plays compared to developing countries including: a greater access to healthcare services and a greater availability of trained healthcare personnel (Hongoro & McPake, 2004; Litsios, 2004). Developing countries such as Australia HBHC are associated with palliative care, caring for the terminally ill (Hudson, 2003).

Palliative care can be defined as: "comprehensive, interdisciplinary care of patients and families facing a terminal illness, focusing primarily on comfort and support" (Billings, 1998). Hudson (2003) also saw palliative as focussing on the terminally if but another definition of palliative care given in the literature is: "Palliative care relieves suffering and improves quality of life for both patients and families throughout an illness experience, not just at the end of life" (Stjernswärd, 2007). These two definitions seem somewhat at odds, specifically in regard to the terminal status of the patient being referred to. The literature does note however that there is some debate as to what exactly is included under the term palliative care, and as such several definitions exist (Materstvedt, et al., 2003). However broadly speaking both definitions involves the patients and the families and attempts to relieve suffering and stress directly or indirectly caused by illness.

Palliative care services has the tendency to be better developed in developed countries, specifically in urban areas, where palliative specialisation is more feasible and likely to occur because of the denser clustering of people (Evans et al., 2003).

Based-on the findings from the literature it is clear that HBHC in the developing and developed world differ in two regards, firstly HBHC in the developing world cares for a far larger range of illnesses ranging from simple nursing to chronic disorders while in the developed world HBHC is more focused, because of market and social forces, on palliative care and caring for terminal ill patients. Secondly HBHC in the developing world functions in an environment noted for a lack of resources, human and otherwise, and a poor ability to meet even basic healthcare needs. This thus leads to the utilisation of more informal caregivers while in the developed world HBHC enjoys a better access to resources and better healthcare systems to meet basic medical needs.

The above Section looked at HBHC in the developed and developing world as addressed in the available academic literature to give a broader view of HBHC and point out the possible different views on the subject based-on where it is being applied. The next Section looks specifically at HBHC within the rural context, in order to better place the research.

2.1.2 Home-Based Healthcare in the South African Rural Context

As seen above HBHC plays a slightly different role in the developing world than in the developed world. This is most likely because each of these types of countries have their own unique characteristics influencing the ability to provide healthcare services in general which then directly influences the role that HBHC plays. For example in developed countries where general healthcare services are more readily and effectively provided HBHC primarily takes the form of elderly and palliative care. In developing countries were healthcare is difficult to provide, HBHC attempts to provide basic healthcare services and breach the supply and demand gap. However rural areas in both developing and developed countries also contain their own unique characteristics influencing the provisioning of care services.

This Section follows the previous Section but narrows the scope looking specifically at HBHC in the South African rural context, in which the research was conducted. The sourced literature spoke directly to the research context and provided a good starting point for the research as well as valuable insights and clarifications to the findings produced by the research.

The provisioning of healthcare within rural areas of developing and middle-class countries is inhibited by a number of factors. Inhibited factors includes many of the previously discussed factors common to the developing world, but also by various factors unique to rural areas including: poor levels of organisation of healthcare services and a high concentration of poverty (Evans et al., 2003), as well as poor access to and availability of basic services, insufficient infrastructure, low levels of literacy (Wouters et al., 2009; Schoeman et al. (2010), general unhealthy conditions, poor quality of food (Schoeman et al., 2010).

Rural, and generally smaller, hospitals suffer from a relatively high demand for care that usually exceeds the available human resources. These smaller hospitals also suffer from a lack of infrastructure and support in terms of money, supplies and technology which in turns leads to an unattractive job prospect for skilled workers. Those that have the needed skills do not want to work in an institution which is already lacking the means to provide successful healthcare services (Jacucci et al., 2006).

This lack of skilled workers and infrastructure causes a number of challenges for managers within the HBHC domain. Managers need to ensure that specific activities and tasks are performed but issues such as a lacking a sufficient number of care workers or the necessary skills to perform specific tasks prevent these tasks from being successfully performed (Jacucci et al., 2006). HBHC however decentralises care and involves informal caregivers, or lay health workers to compensate for the lack in trained medical staff.

Even if the available healthcare facilities, such as small hospitals and clinics, were not facing the problems listed above, individuals in poor communities do not always have access to or the means to get to these facilities. The facilities are usually located too far away from the rural communities and the community members do not always have the monetary means of paying for transport to the facilities (Schoeman et al., 2010). HBHC however delivers care services to the community at the homes of the patients directly addressing this issue (Strasser, 2003; Klaas, 2007)

Not only are local factors in these rural areas inhibiting the ability to provide care but macro factors such as government bureaucracies as well. This is because the care programs assigned to HBHC do not always reflect the needs of the HBHC domain and have a tendency to be heavily influenced by the needs of "health bureaucracies". Also the government interpretations of the needs in HBHC are not always correct and the needs of individual communities are not always reflected in government care plans (Strasser, 2003). Because of this disparity in what is expected or deemed sufficient by governing bodies and what exists in reality cause issues to arise when implementing ICT such as a misunderstanding of context and diversity (Jacucci et al., 2006).

It is clear that a number of causal factors exist; both on a local level and a national level, within and around rural communities prevent them from having access to the latest and greatest health technologies and services. HBHC however seeks to improve the healthcare situation in these communities by providing care specific to the communities needs and to the individuals within the communities in their own homes.

Interestingly Togno and Topps (2000) noted that rural communities easily adopted new technology especially technologies which reduced the communities' isolation and poor access to services. But this is contradicted by the literature review conducted by Tomasi (2003), who, although not being able to explain why, noted a very low adhesion to the various protocols required for computerisation in primary health care. It is possible that the statement by Togno and Topps (2000) is not fully relevant to the South African rural context, which we've already seen from the literature by Wouters el al (2009) and Schneider et al. (2008) is characterised by a lack of IT infrastructure and wariness towards government and non-government initiatives who promise to implement innovative IT solutions.

Even though South Africa is seen as a leader in ICT development these previously-mentioned inhibiting factors for both the rural and developing countries also apply to South Africa (Wouters et al., 2009), leaving citizens who live within these poor rural communities with little option but to rely on HBHC services for the provisioning required healthcare services (Evans et al., 2003).

As mentioned nursing plays an important role in providing HBHC services to those in need, but it was noted that Africa had one of the more severe shortages of healthcare workers. Specifically workers in Southern Africa were described as facing a number of issues such as: low remuneration, work related risks due to possible exposure to illnesses such as TB and HIV/AIDS, unrealistic workloads placed upon caregivers, sub-standard working environments and insufficient infrastructure such as facilities and equipment (Padarath et al., 2003).

Although in the South African context there is a shortage of trained nurses to meet the primary care needs of the community (Klaas, 2007). Semi-trained caregivers, or lay health workers, are thus an important part of HBHC, since they help to address the gap between need and supply.

Since the mid-1990s new policies and programs implemented in South Africa caused an increase in the number of informal carers. By 2004 it was noted that the number of lay workers or informal caregivers was almost equal to the number of trained nurses and by 2006 the number of informal caregivers had reached 62,445. These informal caregivers were collectively called Community Healthcare Workers by the government (Schneider et al., 2008).

Based-on the literature it is seems clear that rural communities face many issues in terms of healthcare. Most of these rural communities lack the monetary means, skills and infrastructure to ensure a generally healthy environment. Formal healthcare facilities that are intended to provide the needed care to these communities themselves face a number of issues that prevents them from doing so. But HBHC attempts to provide an additional means by which healthcare services can be provided in order to ensure that the communities have access to better quality care services.

Issues such as a lack of trained nurses, lack of funding and infrastructure, which prevent the formal healthcare facilities from providing adequate levels of care, also impact HBHC initiatives and their ability to provide care. HBHC still seemingly manages to help bridge the gap despite these challenges, but there is still much room for future research and improvement.

2.2 Health Informatics

This section considers the literature relating to HI as a whole. In Section 2.1 HBHC was discussed both broadly as it related to the developed and developing world, and then specifically as it related to the rural communities in South Africa. It was shown that HBHC plays and important role but faces a number of issues. This Section considers Health Informatics as it holds a number of advantages which could possible help overcome the issues faced by HBHC in South Africa.

Unfortunately little relevant literature was sourced about HI in HBHC and specifically in the South African context. However in order to better understand the micro level of HI in HBHC it would be important to have a look at the macro-level briefly of HI as well. Insight into the currently available academic knowledge on the topic of HI is furthermore useful, as it helps to show the importance of the research and what technologies are currently available and could be used directly or indirectly to meet the objectives of the research.

A number of articles were sourced from the academic literature on the topic of HI. Initially the articles were generic providing an overview of the current body of knowledge available on the topic in order to better ground the overarching context of HI in HBHC. From the initial sourced literature there was a strong theme that revolved around the advantages and disadvantages of HI, which is discussed in more detail later on. However these sources which spoke to the advantages and disadvantage barely touched on the technology components involved and seemed more to be a 'sales-pitch' then an in-depth discussion of action implementation details. Out of these advantages and disadvantages arose the topic of what constituted the successful application and usage of HI in order to ensure the advantages and mitigate the disadvantages.

We start off in this section by giving an overview of HI. The second part of this Section considers the advantages of HI provided in the literature and then attempts to identify any commonalities between the identified advantages and group them together. The third part of this Section considers the issues of implementing HI systems in practice in order to provide a more balanced view of HI. The fourth part of this Section considers how HI systems can most effectively be utilised which leads to the topic of integration and interoperability. This Section is concluded with a look at standards within HI.

Over the years IT has played an increasingly more significant role in the healthcare environment (Beyer et al., 2004). Eysenbach (2000) defined HI as: "the branch of medical informatics that analyses consumers' needs for information; studies and implements methods of making information accessible to consumers; and models and integrates consumers' preferences into medical information systems."

Investment in both Information Systems (IS) and IT within the healthcare domain has grown at a rapid pace over the years (Cho et al., 2008) with the simple goal of contributing and improving the quality, availability and efficiency of patient care (Stead, 2000; Haux 2005).

These factors range from simply increasing quality of care, the various cost pressures in healthcare to increasing information intensive nature of modern healthcare. These factors are discussed in more detail in the following Section.

2.2.1 Factors leading to the increase in Health Informatics

This Sections looks briefly at the factors which led to the increase in the investment and application of IT and IS in healthcare, broadly called HI. This Section is intended to highlight the reasons why HI has become an important part of healthcare and further emphasise the importance of HI and the need and advantages that further research into the topic of HI can add as a whole. Although as it is shown later in this section it is important to remember that these factors are not mutually exclusive and have notable overlap.

The Table 2.1 briefly lists some of the factors that lead to the growth of HI and in the remainder of this Section some of the key points are discussed in more detail.

Table 2.1: *Factors that led to the increase in the investment in Health Informatics.*

Factor:	**Source:**
Improving the quality and efficiency of patient care.	Stead, 2000 Haux, 2005
Increasing information-intensive nature of the healthcare industry	Cho et al., 2008
bridge the widening gap between healthcare supply and demand	Ludwick & Doucette, 2008
Cost pressures	Kuhn & Giuse, 2000 Rowe, 2008

Cost pressure is a driving force for further investment in HI because it promises to decrease overall costs of providing healthcare services (Kuhn & Giuse, 2000). Decreasing the cost of healthcare is especially attractive since the high cost of medical equipment and supplies is a fairly large inhibiting force as to who can and cannot afford the best quality of healthcare (Rowe, 2008). Cost pressures especially hit the poor the hardest since they do not have the needed financial ability to afford the biggest and best new treatments, again touching on the Inverse Care Law (Hart, 1971) mentioned in Section 2.1 under HBHC. Lowering costs thus increases the availability to quality healthcare (or at least removes one of many inhibiting factors) and lowering the costs can thus help to make healthcare more universally available, bridging the gap between supply and demand (Ludwick & Doucette, 2008).

The hope is that the adoption of HI in developing countries, especially the rural areas can help breach, or at least narrow, this gap between healthcare supply and demand (Ludwick & Doucette, 2008; Wouter et al., 2009). Even simply the implementation of routine HI solutions, in rural, middle-class and developed contexts, is seen as being a critical component in improving the ability to provision health service (Jacucci et al., 2006). So far in the developing world no scalable healthcare system has successfully been implemented (Wouter et al., 2009; Chetley, 2006). This might indicate that although the solution is seemingly obvious, the implementation is more difficult.

An interesting and complex additional factor is the information-intensive nature of the healthcare industry (Cho et al., 2008). The information-intensive nature has two contributing causal factors, the first being the implementation of IT itself and the second being the appearance of distributed healthcare networks.

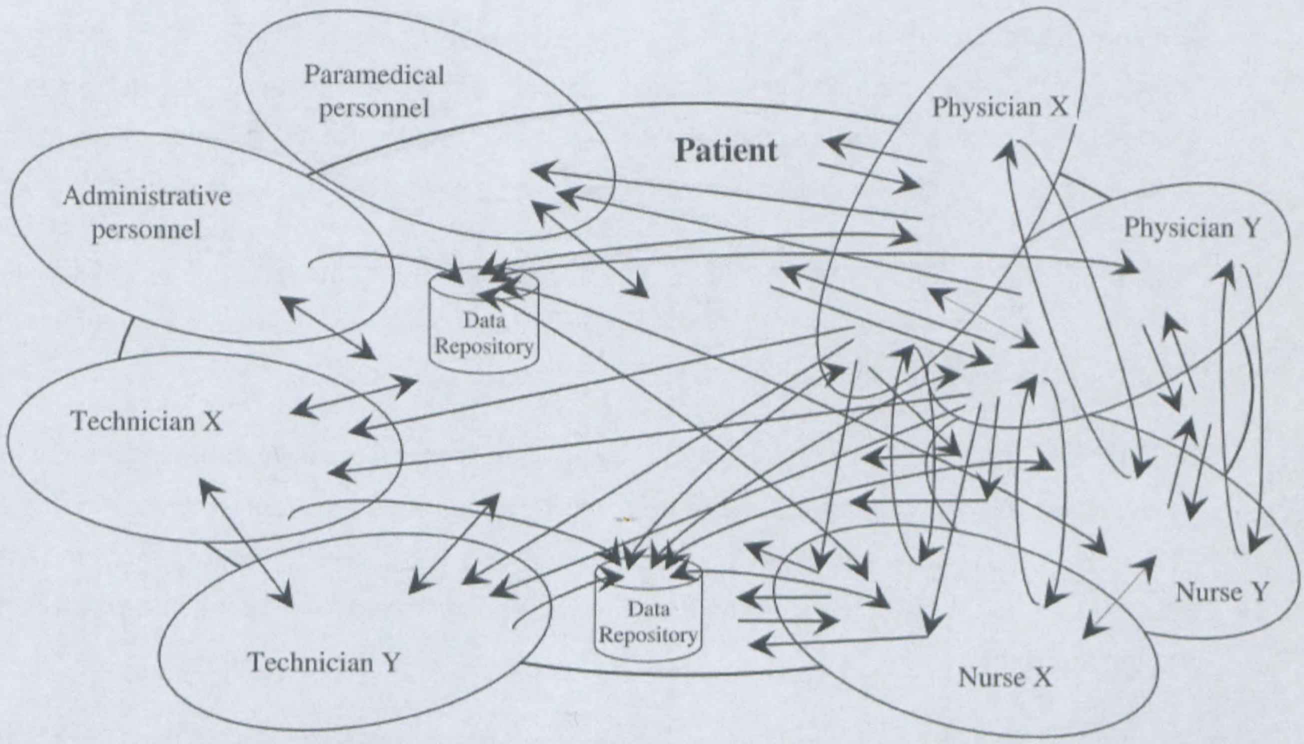


Figure 2.1: *Patient health care management and health record filling are examples of extremely complex co-operative works (Masseroli et al., 2006).*

Some healthcare professionals specialise in specific fields of health and do not have much knowledge or expertise in other fields, or a healthcare professional might choose not to specialise and become a General Practitioner (GP). None the less all these specialists have their own data processing and storage needs. Sometimes a patient might need to be treated by a number of specialists; an example provided by Masseroli et al. (2006) was of a GP who suspected that a patient might have breast cancer. This patient needs to see a specialist in order to confirm the diagnosis and if confirmed the patient needs to be treated by a number of other healthcare works such as nurses or undergo surgery.

In order to enable this distributed care it is necessary that the different healthcare professionals are able to successfully interact with each other and efficiency share information about the patient with each other (Masseroli et al., 2006). Although as the Figure 1.1 shows this can possibly be a very complex number of interactions and data requirements. If one assumed that each specialist has his own set of systems that meet his individual requirements then it is clear that sharing information between these various systems can be a difficult prospect.

Effective and efficient interoperability between the various systems used by healthcare professionals is a possible solution that can enable distributed care. Interoperability is simply the ability of two or more systems to exchange information and in order to be interoperable they need to have some form of integration (Mead, 2006). The goal of integration and interoperability is for information origination on one system to be easily transferred to a recipient system and for the recipient system to be able to effectively utilise that information, essentially to be able to understand the information that it received (Beyer et al., 2004).

The second factor that caused healthcare to become more information centred is HI itself. Not only has distributed care contributed to making healthcare more distributed and complex but the implementation of HI helps to facilitate this too (Stead, 2009).

By implementing increasingly sophisticated healthcare systems and equipment the amount, variety and complexity of data that can be explicitly captured is constantly increasing (Haux 2005). New advances in diagnostic equipment and data management, storage and dissemination technology have created not only a huge volume of data but also new ways of sharing this data.

The data being produced in healthcare ranges from patient specific data, such as patient details or patient diagnoses, to resource management data, such as medical stock supplies, (Kaur & Wasan, 2006) all of which need to be stored and processed. The large volume along with the large differentiation in the types of data makes the management of the data no easy task (Davies et al., 2005), which for the most part is addressed by implementing more advanced technology solutions to meet these processing needs.

Most of the contributing factors to the implementation are closely related. The desire to improve healthcare, enable distributed care, bridge the supply/demand gap and cut back on the rising cost of care has essentially led to the investment in HI. The active usage of HI in turn has allowed a larger volume of more complex and differentiated data to be produced, which for obvious reason is more complex to manage (Davies et al., 2005) thus leading to the need for more IT in order to process and disseminate this data.

The interrelated factors above contributed to the investment and utilisation of HI. Most of these factors also lightly touched on the advantages that HI holds if successfully implemented and effectively utilised.

This section looked at some of the reasons found in the sourced literature that led to the investment in HI being increased in recent years. However it is difficult to fully discuss these factors without touching on the advantages that HI promises to those willing to invest and use since investing in HI is done in order to achieve some form of advantage. Otherwise if HI offered no advantage then there would be no point in investing.

The following Section looks more closely at these advantages that HI purposes to bring.

2.2.1 Advantages of IT in healthcare

This Section considers the advantages that IT brings to the healthcare domain. These advantages are not specific to HI in HBHC but apply to the general healthcare domain. Ideally the advantages would apply uniformly to the broader healthcare domain and sub-domains such as HBHC, but this is more than likely not the case because of the various complexities found within HBHC and the unique characteristics of not only HBHC but also the communities in which they operate. But it is presumed that these advantages can, if contextualised correctly, be related to and become relevant in the HBHC environment.

This Section came about because of the strong theme of the good that IT can do within healthcare, thus initially source literature broadly looked at the topic. Although more detailed sources were found, few of these sources related directly to the context of the research, specifically towards the advantages that IT holds within HBHC in the South African context.

These advantages found in the literature further help to emphasise the importance of the role that HI can play. This ideally coupled with the section on HBHC in South Africa above that shows some of the issues faced by HBHC. Furthermore this section helps to emphasise the need for HI solution to be developed for the South African HBHC context.

Table 2.2: *Advantages of IT in healthcare as found the in literature.*

Source	Advantage	Context
Dixon and Dixon (1994)	Increased quality of care	Patient
	decreased transcription errors	Facility
	decreased reliance on clerical staff	Facility
	possible attracting new physicians to hospitals	Facility/ Healthcare workers

Table 2.2: *Advantages of IT Healthcare as Found in the Literature (continued).*

Source	Advantage	Context
Kuhn and Giuse (2000)	improve coordination of care	Healthcare workers
	link institutions to trusted suppliers improve supply chain	Facility
	delivering care remotely where appropriate	Patient
	Offer net perspectives for both patients and caregivers	Patient/ Healthcare worker
Ammenwerth et al. (2003)	Means of reducing clinical errors such as diagnostic and medication errors	Healthcare workers / Patient
	A means of supporting healthcare professionals by providing them with relevant timely and up-to-date information	Healthcare Workers
	Increasing the efficiency of healthcare, allow for a higher turnover of patient without negatively influencing quality of care	Facility / Patient
	A general improvement of the quality of care	Patient
Beyer et al. (2004)	Increased availability of patient data	Healthcare workers
	A means of preventing adverse events	Patient
	Prevention of unnecessary duplication of test and procedures	Facility/ Patient/ Healthcare workers
	A means to reduce costs	Facility
	A means to reduce waiting times	Facility / Patient
Jacucci et al. (2006)	Ability to provide timely and accurate information, to effectively monitor and address health issues, such as epidemics in rural areas.	Facility
Garde et al. (2007)	An enabler of efficient communication	Patient/ Healthcare workers
	A means of reducing costs	Facility
	An improvement in the quality of care	Patient/ Healthcare workers
	patient empowerment	Patient
Rowe, (2008)	Facilitate data exchange leading to better access to data	Facility/ Healthcare workers
	A tool to improve communication between colleagues	Healthcare Workers
	A means by which healthcare professionals can continue to improve their knowledge	Healthcare workers
	A means to facilitate deep learning not only for healthcare professionals but also for patient and the community at large	Patient/ Healthcare workers
	Allows primary healthcare centres to gain access to clinical specialists	Facility

It was not the intent of the literature review to source articles on the advantages of HI, but it was a strong theme amongst the source articles none the less. Only seven articles spoke directly to the advantages of HI but based-on these articles it seemed obvious enough that these advantages could broadly be placed into three categories: the advantages for the facilities, the advantages for the healthcare workers and finally the advantages for the patients. Although these advantages could then again be sub-divided, by doing so would be time-consuming since a lot of overlap exists and many of the advantages would fall into several sub-categories.

The advantages to the facility are fairly straight forward such as: the reduction of costs (Beyer et al., 2004; Garde et al., 2007), an increase in the number of patients that can be treated (Ammenwerth et al., 2003; Beyer et al., 2004), or attracting new staff and skills (Dixion and Dixion, 1994). However many of the facility advantages are just as relevant to healthcare workers or patients, since patients also benefit from a higher turn-over rate and both patients and healthcare workers benefit from reducing clinical errors (Ammenwerth et al., 2003).

Many of the advantages of the patient and the healthcare worker is overlapping and multifaceted.

The Information Technology Association of America (ITAA) defines IT as: "the study, design, development, implementation, support or management of computer-based information systems, particularly software applications and computer hardware" (Taylor, et al. 2009):139. IT also encompasses a component of data management; most of the systems that are encompassed under IT have some information components to them and seek to achieve specific goals by the generation, processing and dissemination of data.

It was considered by several of the sourced authors that an advantage of HI is that it enables the improvement of communication in some form or another. Communication improvements included the communication between patients and their caregivers or between colleagues enabling healthcare workers to get in contract with clinical specialists and make use of their skills and insights which might not have been available locally (Kuhn & Guise, 2000; Garde et al., 2007; Rowe, 2008). Garde et al. (2007) also identified that HI allowed data to be effectively and efficiently disseminated. Kuhn and Guise (2000) identified another advantage that formed a subset of communication; specifically that HI enables better communication between the healthcare institution and trusted resource suppliers.

Access to information is another one of the multifaceted advantages of HI namely: the dissemination of relevant and timely information to the right parties (Ammenwerth et al., 2003; Beyer et al., 2004; Jacucci et al., 2006). Furthermore Ammenwerth et al. (2003) and Beyer et al. (2004) saw this as potentially reducing the possibility of clinical errors from occurring; because in order to make the correct medical decisions there is a need for both the decision maker to have the relevant task and domain knowledge (Bird, 2003).

But access to information not only supports people in making the right medical decisions but can also help to inform and educate, enabling further empowerment of both patients and caregivers. Technology enables the patients and their families to learn more about the given illness or even how to prevent or treat certain illnesses. The caregivers on the other hand can also increase their medical skills and knowledge to be able to better perform their duties (Rowe, 2008).

Not only does the reduction of clinical errors but also the reduction of transcription error exist as an advantage of HI. Reduction of transcription errors was noted by Dixion and Dixion (1994) which leads to another of the advantages noted by Dixion and Dixion (2004), being the reduction in the need for clerical staff, which would usually have been needed to catch and correct these errors.

Several more vague advantages also came out of the literature, increase in the quality of care was noted by several authors (Dixion & Dixion, 1994; Ammenwerth et al., 2003; Garde et al., 2007), although little actual metrics for what constituted quality care was provided. It is presumed however that quality care encompasses many of the above-mentioned advantages and that by achieving these the quality of care will also be increased.

By increasing the skills and knowledge of the caregivers (Rowe, 2008), ensuring that they have the relevant information at the right time improving overall coordination of the care activities (Kuhn and Giuse, 2000) and thus making the care process more efficient and reducing turnover and waiting times (Beyer et al., 2004), the overall quality of care is also improved.

Many of these generic HI advantages are also relevant to the rural HBHC context, which has been shown to suffer because of a number of disadvantages when it comes to the effective and efficient delivery of healthcare services. Disadvantages such as a lack of human resources can be addressed by increasing the efficiency of the care activities and providing education for the healthcare workers and communities they catered for. Reducing the clinical and administrative errors and the need for clinical staff and overall reducing the costs of care makes it accessible to those who would previously not have had access to these healthcare services.

This Section looked at some of the advantages highlighted in the literature; the next Section considers the issues of HI to provide a more balanced view of HI.

2.2.2 Issues involved in IT in healthcare

This Section considers the issues involved in the implementation of a HI solution. Similar to the advantages listed above most of these issues are not specific to any individual domain within healthcare, although a couple are specific to the developing world. None the less the issues involved with HI are a strong theme that emerged within the sourced literature on HI. It is believed that looking at the issues and disadvantages provides a balanced view of HI and by contextualising it correctly is of use to the research.

Although HI is clearly important and offers a number advantages as discussed above, there are also issues impeding the successfully exploitation of IT in the healthcare industry. The following tables list the issues which were found within the literature looking at healthcare in general.

From the issues listed in Table 2.3 it follows that the issues involved in HI can broadly be placed in 4 categories, namely; human, technical, organisational (or facility) and cost/effectiveness.

Table 2.3: *Issues involved with IT in healthcare.*

Source	Issue	Context
Kuhn & Giuse (2000)	Data security concerns	Human
	Concerns about the confidentiality of data	Human
Grimson (2001)	Multimedia presentation complexity of medical data	Technical
	Data capture as unstructured text, issues with important and usage of unstructured data sources	Technical
	Terminology compatibility issues	Technical
	Coding and classification issues	Technical
	Concerns about the security and confidentiality of patient data	Human
	Absence in many countries of unique national patient identifier	Organisational
	Inability to fully comprehend the benefits of a new HI solution	Human
	Impact of ICT solution on the organisational structure	Organisational
	Wariness due to number of highly publicised failures	Human
Ammenwerth et al. (2003)	Users and workflow needs a lot of time to get used to new tools	Organisational
	Takes time to completely exploit the new possibilities	Human

Table 2.3: *Issues involved with IT in healthcare (continued).*

Source	Issue	Context
Hersh (2004)	Cost	Cost/effectiveness
	Technical- complicated, issues conforming to workflow / workflow modified to fit system	Technical
	System interoperability	Technical
	Privacy and confidentiality – exists whether the medical record is paper or electronic	Cost/effectiveness
	Lack of well-trained clinician informatics workforce to lead process	Human
Garde et al. (2007)	Physician concerns about work flow	Organisational
	Broad environmental issues	Organisational
	Perceived lack of added value	Cost/effectiveness
Ludwick & Doucette, (2008)	Concerns about the safety of new technologies	Cost/effectiveness
	Negative impact on the provider/patient relations	Human
	Staff anxiety about the impact of new technology	Human
	Time needed to implement	Cost/effectiveness
	Negative impact on the quality of care	Cost/effectiveness
	If new technology will be financially efficient	Cost/effectiveness
	Liability concerns	Human
Cho et al. (2008)	Knowledge and management	Organisational
	People and organizations	Human
	Social communication patterns	Human
	Organization structure and culture	Organisational
	Resistance to change	Human
	Divergent interests across stakeholder groups	Organisational

The human issues are raised by most of the above authors especially Cho et al. (2008). The technical issues are slightly touched on, mostly by Grimson (2001) but based-on Table 3.2 above these are more perceived technical issues than real tangible issues. This perception makes it hard to be completely sure about the categorisation being used i.e. would the concerns of healthcare workers about some technical issue fall under human or under technical? Erring on the side of caution these perceptions were categorised under human issues, since it is quite possible for the perception to be inaccurate or based-on misinformation and that in reality there are no actual technical issues.

The human issues of HI are mostly centred on human perceptions. Some of these perceptions including: an anxiety around the impact that new technology could have, such as drastic changes to workflow and processes or to authority structures, and the lack of perceived benefit of IT (Ludwick & Doucette, 2008; Grimson, 2001). Other human issues refer to the perception that technology would negatively impact the relationships between the patient and the carer (Ludwick & Doucette, 2008), or that technology is intrinsically less secure and confidential (Kuhn & Giuse, 2000).

This anxiety leads to further human issues such as the question of cost effectiveness and questions around the safety of new technological solutions (Kuhn & Giuse, 2000; Grimson, 2001; Hersh, 2004; Ludwick & Doucette, 2008), although these issues, around cost and effectiveness, seemed notable enough to be placed in their own category. It is also apparent that there is in some cases a resistance to change or even a culture of resistance to change, amongst the individuals and users involved with the HI solutions (Cho et al., 2008). More than likely this resistance to change is because of these above-mentioned perceptions of HI; these perceptions are confirmed by the highly published failures of previous HI initiatives (Grimson, 2001).

A number of organisational issues were identified, although these issues are also somewhat perception based as well but they are seemingly more concrete. The organisational issues are seemingly two-folded, firstly based around fear that HI might negatively impact the workflow and work activities and actually get in the way of the care process (Grimson, 2001; Ammenwerth et al., 2003; Garde et al., 2007; Cho et al., 2008) and secondly that organisational characteristics might prevent the effective implementation of HI, such as differing stakeholder interests (Cho et al., 2008) or time concerns related to implementation of HI (Ludwick & Doucette, 2008), or the lack of unique patient identification in some countries (Grimson, 2001).

The final identified category of issues in the literature related to the various technical issues involved in the implementation of HI. A more straightforward and simple of the technical issues is the technologies' ability to conform to be current workflow, ensuring that HI acts to aid the care activities (Hersh, 2004). More complex issues involve questions around how to effectively store and utilise the data that is produced by the day to day medical activities. Grimson (2001) brought up several of these issues, particularly issues around how multimedia can effectively be used to structure the complex medical data. Of a particular note was the issue around storing data as unstructured text. It can be difficult to discern what is and isn't useful information amongst a wall of plain text, or to ensure that the data is being interpreted and used correctly.

Further technical issues were raised by Grimson (2001) namely issues of compatibility between the various terminologies used in healthcare and issues around coding and classifying medical data. These issues between relate directly to how the data is collected, stored and utilised, which might differ between HI systems.

The above issue of formatting data and choosing how to store it leads into another of the technical issues of HI identified in the literature, interoperability. Interoperability was mentioned in the literature as a notable issue that prevents HI systems from being utilised to their fullest (Hersh, 2004). On the topic of interoperability Kuhn & Giuse, (2000) noted that medical data which is intended for interoperability can originate from a number of HI system or even from various paper forms. Furthermore Mead (2006) noted the terminological and classification issues mentioned by Grimson (2001) referring to them as syntax inconsistencies.

Interoperability promises to effectively deliver information to those who need to improve decisions and relevant processes. Getting the correct information to the correct people is important in healthcare where people's lives and the future quality of life are at stake. Thus achieving successful interoperability can effectively achieve many of the advantages that were listed above. Interoperability has however been a major focus of research in healthcare, dating back all the way to the 1960s, but has yet to be seamlessly achieved (Stead, 2000).

One of the possible solutions to achieving successful interoperability was the creation and usage of an electronic healthcare record (EHR) (Hersh, 2004). The EHR promises to resolve issues of using different types of data structured and unstructured and coding and classification issues (Grimson, 2001). The EHR helps to standardise the structure in which electronic records are stored and used in hopes of solving these issues.

The following table lists issues with the implementation of the Electronic Healthcare Record.

Table 2.4: *Issues with the implementation of EHR.*

Source	Issues
Grimson (2001)	Medical record can be structured in a variety of ways Record fulfils many different functions which require somewhat disparate data. Actual data recorded in records range from unstructured text, structured numeric +/ char data, images, graphics, sound, videos. Patient longitudinal record typically will be distributed between many different care providers

The concept of the EHR was created to address a pressing issue within HI, namely: interoperability (Hersh, 2004). The effective and efficient interoperable systems are an important goal within HI, one which on a universal scale to this day is still illusive.

It is ironic that the EHR which was intended to be a solution to numerous issues within HI itself has numerous issues. The underlying idea behind the EHR is to create an electronic record which functions as a standard and uniform means of storing and more importantly exchanging medical information (Hersh, 2004; Grimson, 2001). Although the concept of the EHR might sound fairly simply there are key issues preventing its realisation as listed by Grimson (2001) in Table 2.3. One of the more notable issues is the large variety of records and differing goals (blood tests, x-rays etc.) between the records being used. It is not difficult to imagine that creating a single standard record which can replace the myriad of different records can be extremely problematic.

Whereas most of the above authors looked at the more generic issues involved in HI and gave a good overview of some of the issues, Jacucci et al. (2006) looked specifically at the issues involved in implementing HI in developing countries, although these issues looked more at the human and organisational issues and didn't touch heavily on technical aspects.

Table 2.5: *Issues involved in implementing HI in developing countries.*

Source	Issues
Jacucci et al. (2006)	Limited duration of donors' financial support Inadequate focus on local expertise Too-Narrow interventions (often sustainable health information system requires a parallel reform of the health sector) Technical bias of projects (inadequate focus on human resource development) Pilot project orientation

Developing countries have their own limitation and issues to implementing HI solutions such as limited local skills and knowledge and lacking infrastructure. The issues listed by Jacucci et al. (2006) focus on some of these issues, as in most other cases where funding is listed, this might relate to the time required to implement a solution possibly because of the complexity of the solution.

From the other listed points above it is clear that Implementation of a HI solution is more than simply implementing some new technology, there is a need to develop the local expertise, ensure that the adequate support infrastructure and reforms are in place. The following Section looks specifically at what is required for the effective usage of HI.

2.2.3 Effective Usage of Health Informatics

With little doubt HI plays an important role in healthcare since it offers a number of advantages and through its increased usage in the healthcare domain, healthcare services can more readily be made available and hopefully the quality of such services can be improved.

It is, however, clear from the above that simply implementing IT in any given domain doesn't intrinsically lead to effective utilisation and improvement. Many of the issues faced by HI solutions are not related to technology at all, such as perceptions and misperceptions of HI and the numerous characteristics of the environment (such as the unique characteristics of the developing world) that inhibit the successful implementation of HI.

Any attempt to implement healthcare solutions in developing countries needs to do it on a number of different levels, taking into account a number of different contextual and environmental issues which was discussed in more detail above under HBHC in Developed and Developing Countries.

In this section we look at how HI is being used and what would be required to ensure that HI is being effectively and efficiently being utilised to try and prevent most of the listed issues and listed advantages.

Doctors primarily use computers for three purposes: research, writing reports and for presentations preparation (Rowe, 2008). Rowe (2008) indicates that clinicians and researchers who use their computers for research purposes tend to be unable to do so effectively. This seems to indicate that IT is fairly pervasive within certain context of healthcare but that it is still not being used to its full potential, as listed above under the advantages of HI.

For example Li et al. (2000) saw the need for tools to be available for medical experts to create and manage their own knowledge basis and import knowledge from numerous sources. The ideal was to allow medical experts to play a role in the development of knowledge bases allowing for more relevant information to be retrieved.

The opinion of Li et al. (2000) was also voiced by Brandt et al. (2003) who called for the need for database application that managed clinical data to be easily modified and to be able to adapt to new scientific advances.

No strong single theme for how IT can be utilised successfully was found in the literature however. Li (2000) saw a need for IT to be usable by medical professionals with little computer science experience while on the other hand Rowe (2008) saw a need for medical professionals to become more computer literate in order to utilise advances in IT. Although Li (2000) and Rowe (2008) both purpose two different requirements for the effective usage of IT, it follows that the utilisation of IT by healthcare professionals is important and from that that IT wasn't being fully utilised.

Wouters et al. (2009) identified that a need for proper needs analysis to ensure communities accept IT based solutions has being critical and this means that there should be no gap between the design of the perceived usage of a solution and its actual use in practice. It seems Wouters et al. argued for IT to conform to the indented user which is similar to what Li (2000) has indicated.

There is a need in HI systems for appropriate data input mechanism and adequate structures to store medical data. For more advanced HI systems there are also the additional requirements for a critical amount of data to be present in order for the system to be usable. Clinical research systems for example tend to need large volumes of detailed information in order to produce relying findings (Kuhn & Giuse, 2000).

Dixion and Dixion (1994) identified a few key factors acceptance and use of IT innovations and these are: the perceived usefulness of an IT innovation, the perceived ease of use of the given IT innovation and the sophistication of the end user, areas of knowledge that the user possesses.

A common solution to the IT processing needs is the implementation of Best of Breed (BoB) systems. These systems are specialised to perform specific functions or meet specific requirements, and for example a given medical specialist might implement a system which meets his specific needs. But the issue with the implementation of BoB systems is that it creates silos of information because these BoB systems are normally proprietary and contain little ability to share information with each other which prevents many of the previously listed advantages from being achieved (Holland, et al., 2001; Moss, 2007).

Correctly implemented and utilised HI allows for healthcare professionals to easily and effectively share information which is needed for optimal patient care (Garde et al., 2007) and thus an important prerequisite for integrated care. As has been mentioned before interoperability enables many of the advantages that HI promises, namely improving the coordination of care processes (Kuhn & Giuse, 2000) and communication between healthcare colleagues (Garde et al., 2007; Rowe, 2008), the dissemination of timely information (Ammenwerth et al., 2003; Jacucci et al., 2006).

One of the proposed means for sharing patient information is the Computerised Patient Record (CPR), or the Electronic Healthcare Record (EHR) (Harsh 2004). Already in 1991 the idea of a standardised easily-shareable electronic record had been seen as essential to healthcare (van Ginneken, 2003). The concept of the HER since has steadily been growing into the dominate idea for sharing patient related information in healthcare (Davidson & Reardon, 2005).

The use of an EHR promises to solve many issues around sharing medical information. The issue of interoperability in HI dates back to the 1960s (Stead, 2000) but there is currently no single universal standard for an EHR or similar solutions (Jacucci et al., 2006).

In order to create a flexible, structured EHR the explicit definition of metadata is required which can support the data-elements relating to a given patient and the context in which these data-elements are created and managed (van Ginneken, 2003).

The Electronic Healthcare Record (EHCR / EHR) is a prime example of an attempt to deal with interoperability issues in HI. Hospitals use a number of documents for their day to day processes and these documents appear in a number of formats both electronic and paper based (Harper, et al., 1997). EHCR is described as being a dominate idea for standardising the electronic and paper based documents and creating a common computerised record for the sharing of patient-level data (Davidson & Reardon, 2005). What EHCR seeks to address is the issues surrounding interoperability and integration between the various electronic heterogeneous systems used in healthcare.

Although having forty years of research into the topic of interoperability, the overall penetration of a standardised record with more than basic information into the day to day operations of healthcare organization has been relatively small (Grimson, 2001), although adoption rates are on the rise (Ludwick & Doucette, 2008). A more detailed discussion of the issues surrounding the integration follows in the next Section.

2.2.4 Integration of Health Informatics Systems

Integration and interoperability are an important consideration in most fields that IT appear, although it is of a particular importance in HI where the interoperability challenge has not yet been overcome. Where successful however interoperability, as shown above, holds the key to enabling a number of the advantages promised by HI. Various attempts have been made to enable seamless interoperability amongst HI systems, amongst which is the concept of the EHR and similar standards.

However looking at integration or any specific integration standard is not the core focus on the research but the underlying issues of interoperability however are relevant to the research, and as such interoperability is discussed in order to highlight these issues.

This Section looks briefly at the literature on integration, primarily integration in HI, and attempts to discuss some of the underlying issues.

Interoperability is simply the ability of two or more systems to share information, these systems are then said to be integrated. In healthcare it is normally necessary for data to be pooled and integrated from a number of sources but the needed information usually originates from varied sources such as IT systems and devices to paper based systems (Kuhn & Giuse, 2000). These different systems usually perform specific tasks or cater for a specific medical specialisation. An example of these disparate or specialised systems is the case of shared care where a patient is being cared for by a number of specialists each with their own tasks and data requirements, which need to be shared and disseminated amongst each other to effectively take care of a patient (Masseroli et al., 2006). It was stated by Grimson (2001) that an important requirement for integrated care is a ubiquitous shareable record.

The problem within HI is that a number of proprietary specialized systems are used with little standardisation in how the data is stored or formatted (Beyer et al., 2004). This has resulted in a situation where patient information is being stored in a number of different locations resulting in several islands of data which can be hard to integrate (Kuhn & Giuse, 2000), and furthermore the data on these islands are stored in potentially different formats.

Interoperability and integration is not always guaranteed when using these heterogeneous systems because these systems might vary significantly in the technologies they use and the design logic used in their creation. These systems 'see' the world different, their ontologies are different and thus mapping similar concepts across these ontologies aren't guaranteed (Mead, 2006).

Integration promises to effectively share and combine the information from the disparate sources and deliver information to those who need to improve decisions and relevant processes (Stead, 2000). Getting the correct information to the correct people is important in healthcare where people's lives and the future quality of life are at stake.

This issue of heterogeneous systems in healthcare was not only voiced by Kuhn and Giuse (2000) but also by Ingenerf et al. (2001). To overcome the issue of heterogeneous systems they called for standardisation to be in place, more specifically for standardisation of patient data in order for systems to use compatible data structures to allow for interoperability. The ability of records to be shared between medical systems is based-on their adherence to standards (Grimson, 2001).

The various medical records used in healthcare, both within HI systems or on paper, can be structured in a number of ways with each performing a specific function. The data captured in these records also vary widely from simple plain text, highly formatted text and a number of multimedia formats such as images or videos. Thus the creation of a single global standard or single medical record can be extremely challenging (Grimson, 2001).

The different levels of interoperability issues were listed by Beyer et al. (2004) and included: syntactic compatibility, differences in the structure of data or messages being sent and the encoding of the data and or messages (Mead, 2006), ontological comparability, differences in the semantics that considers the meaning and relations of data (Mead, 2006) and finally the terminological comparability. The semantics of the data is usually hard-coded into the systems, for example the database schema. Semantic incompatible data is normally the product of users entering data which is incompatible with the systems data semantics.

Beyer et al. (2004) called for interoperability standards to be based-on common ontologies. An ontology is simply body of knowledge that describes objects within a specific domain such as healthcare or business and the relationships amongst these objects (Chandrasekaran, 1999).

As has been showed in the Sections 2.2.1 HI has the potential to positively impact healthcare in a number of areas, but several of these promised advantages require the various systems used in healthcare to be integrated. As has been previously mentioned and shown in this Section interoperability is not easily achieved. The differences in the scope, focus, purpose and ontologies of the various systems used in HI lead to a number of issues in terms of data sharing. Even if the systems can share data there is still the possibility that the data being transferred will not be interpreted the same by the sending and receiving system, leading to issues of semantic inconsistency.

Various articles called for the usage of standards to help structure the data, such as the creation of electronic records, with the ideal that the standards would not only enable the transfer of the data but also the correct interpretation of usage between sending and receiving systems.

But work in the field of interoperability in healthcare has been underway since the 1960s (Stead 2000), and no ubiquitous universal solution has yet been created. Various standards exists that attempt to enable integration, and the following Section briefly considers these standards.

2.2.5 Standards for Healthcare Records

A number of standards exist in healthcare, especially within the domain of HI. Standards such as Health Level 7 (HL7), CEN, GEHR (Good European Health Record) all attempt to structure medical data for the purpose of exchange or for storage (van Ginneken 2003).

The HL7 standard attempts to address issues involved in semantic interpretation of data during the process of data exchange (Kuhn & Giuse, 2000) by defining a comprehensive standard for messages to be sent between HI systems. (Ingenerf, et al., 2001)

HL7 is one of the many attempts to define a standard for EHRs within the healthcare industry (Chong et al., 2003) although the primary aim of the HL7 standard is for the usage in hospitals or similar facilities (Ingenerf, et al., 2001), raising some questions of its relevance and successful application in less structured medical contexts, such as rural healthcare.

The idea behind the HL7 V3 is that it defines the abstract concepts, the grammar, of information used within healthcare. HL7 V3 is a far cry from the original idea behind HL7 of defining a comprehensive model containing every single element of the HL7 standard and also several of the larger EHR models (Schadow et al., 2006). The intent of the standard is to define and structure between medical information in a way that is independent of any particular implementation thus ideally enabling consistent data usage and sharing (Smith & Ceusters, 2006).

Figure 2.2 shows the various objects within the HL7 RIM. The HL7 V3 defines five abstract concepts ((Mead, 2006; Schadow et al., 2006) :

- The Entity which is objects or nouns such as patients, materials etc.,
- The Role (or relators), time-based capabilities and capacities which relate Entities.
- The Participation (or prepositions) a Role in the context of an Act.
- The Act (or verbs), a clinical/administrative/financial occurrence or plan.
- The Act Relationship which is a semantic link between acts.

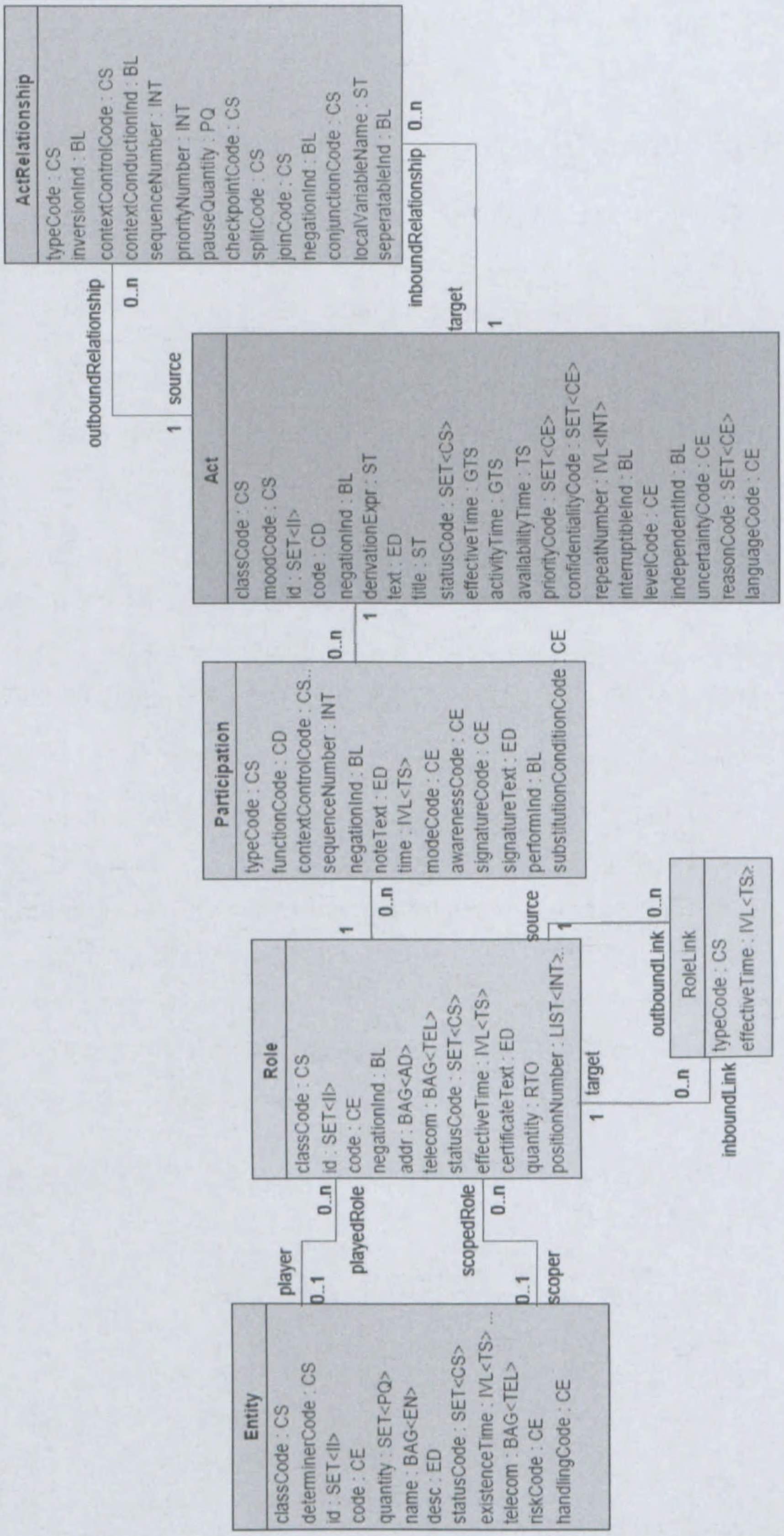


Figure 2.2: *Unified Modelling Language Class Diagram for HL7 Reference Information Model (Schadow et al., 2006).*

Possible inconsistencies of the HL7 V3 standard was discussed by Smith and Ceusters (2006), and Schadow et al. (2006) on the other hand defended the standard against some of the criticism raised in the literature. These topics, the merits and demerits of HL7, do fall somewhat outside of the scope of the research, as in part does these interoperability standards but their underlying goal is relevant.

What the HL7 and similar standards attempts to do is define the structure and relation of medical information in order to standardise to storage, usage and transmission of data between systems attempting to interoperate. Ideally by conforming the way data is stored the issues listed above that prevent interoperability from being achieved can be overcome and the advantages of HI seamlessly achieved.

It was noted by Jacucci et al. (2006) that sustainable global standard in HI was highly dependent on first having to achieve sustainability at a local systems level. The various interoperability standards in HI promise to achieve a lot of advantages which could greatly benefit rural healthcare where IT penetration is notable low, as has previously been discussed in Section 2.2.1. However in the underprivileged rural context there might not be the needed local sustainability to implement and maintain advanced HI standards which then limit the potential usefulness and benefits of any HI solutions. Thus it seems before anything else there is a need to gain a comprehensive understanding of the local level first.

Section 2.2.1 above looked at the factors that caused the rapid increase in HI investment, in order to understand why HI is so important in the healthcare domain. From the literature two primary causes for investment were identified: firstly the emergence of integrated healthcare, and care networks and secondly the increase in the information intensive nature of healthcare. The literature also pointed out various advantages of HI, which also helped motivate investment, but were deemed prominent enough to be discussed on their own.

The following sections continues from the first section by looking at the advantages that the successful implementation and usage of HI hold, and then at the issues involved in the implementation of HI in order to balance the discussion (by looking at both the advantages and issues). By looking at the advantages and disadvantages of HI the topic emerged of what was required for the successful implementation and usage of HI, which touched on and brought up the topic of interoperability. Interoperability is a prominent topic in the HI literature and a necessary component in achieving the advantages listed by HI, but in order to achieve interoperability there is a need for standardisation to be in place. Before conformation to HI standards can take place there is a need to first understand what is occurring at the ground level.

The following sections consider the literature relating to topics relevant to understanding the interoperability challenges above as well as understanding the usage and modelling of data at a local level.

2.3 Metadata

This Section considers the literature on the topic of Metadata. Within this Section a fairly detailed look at the literature on the topic of metadata is undertaken, describing what metadata is, the different types of metadata and metadata standard. The overarching research considers the development and design of an IT-based artefact which can be used to help better understand the data-elements in rural HBHC, and metadata plays a role in the research.

Using the academic literature Section 2.1 contextualised the research area and showed the importance of HBHC and research within the rural HBHC context. Section 2.2 continued from Section 2.1 and used the literature to demonstrate the ability of HI to make meaningful contributions to healthcare in general but also to HBHC and address some of the issues within HBHC, specifically the need to understand the usage and structure of data and information at a local level.

This section forms part of the technological due diligence of the research, the research focuses on the design and development of a metadata repository to be used in the HBHC context.

The following sections consider the literature sourced on the topic of metadata specifically focussing on the different types of metadata found within the literature, the different usages and applications of metadata. The various standards which govern these different types are also discussed and in the final section the topic of what is considered the quality criteria of metadata and how metadata can be created.

2.3.1 Types of Metadata

During the course of the literature sourcing and review a number of definitions and examples of different sometimes vastly different, types of metadata was discovered. It was clear that before any consideration could be made on the usage of metadata due diligences had to be paid to the different types, since although clearly the academic literature is filled with these different types of metadata, rarely are the different types acknowledged and discussed.

Although it is not the goal or purpose of the research to shed light on the myriad of metadata definitions that are found in the academic literature, it is however still important to look at these different types because it offers a better insight into the topic of metadata ideally showing that metadata is a fairly wide topic within the literature. Because of the variety of different types of metadata there is also the fear that confusion might arise or misunderstanding of these different types. The research focuses on a specific type of metadata, that of structural metadata, which was only found in a handful of literature sources, and thus might be somewhat obscure. It is then likely that there might be some confusion because of the relative obscurity which only affirms the need to detail the different types of metadata.

This section lists and discusses the different types of metadata that were found in the literature. During the discussion of the different metadata types an attempt is made to see if any communality can't be identified between the different types.

Several attempts have been made in the academia body of literature to define what exactly metadata is (Park, 2009). One of the common definitions of metadata within the sourced literature is simply: data about or relating to data (Timpf et al., 1996; (Huynh, et al., 2000; Smail-Tabbone, 2005; Kim, 2005). This definition is however widely regarded as vague (Kim, 2005).

The common definition is so vague in fact that it can literally be applied to almost any aspect of computing, ranging from data warehousing (Huynh, et al., 2000), digital and non-digital cataloguing (Bowes, 2006), reporting and report generation (Sun 2005), statistics (Gillman & Appel, 1997), health informatics (Smaïl-Tabbone et al., 2005, Rehman, 2005, Darmoni & Thirion, 2010), database systems (Kerherve, 1997) and web pages (Najjar, 2006).

Despite the vague definition of metadata and the various different implementations, Kim (2005), tried to define what legitimate metadata is stating: ""Legitimate" metadata is metadata that the customer needs in order to understand the semantics and lineage of stored data, and in order to properly run the applications in support of the business needs." This provides 2 seemingly different dimensions in which to define metadata, metadata that aids the user and metadata that are required for applications to run.

Given the general vagueness of the standard metadata definition most of the literature that provides a more comprehensive definition only defines metadata in reference to some aspect of computing in which it is used, for example, Huynh, et al. (2000) defines metadata as it relates to data warehousing, Bowes (2006) defines metadata from a digital library perspective and Gillman and Appel (1997) defines metadata required within a statistical repository. Table 2.6 provides a more comprehensive list of metadata types and their relating contexts.

To further illustrate the point the following table shows the different identified types of metadata, the literature sources that they were found in and the context in which the metadata is used. The list does not contain all the sources, only those that explicitly named the type of metadata.

Table 2.6: *Types of metadata found within the literature.*

Source	**Context**	**Type**
Brandt et al. (2004)	Clinical Studies	Descriptive Process-Related/ Technical/ Active
Sun et al. (2005)	Clinical Reports	Report parameters Format and Display properties Column information Study/data limiting Data preparation for complex reports
Kim et al. (2004)	Speech recognition	Structural
Cleary et al. (1996)	Relational Database	Data Type Relational Rule-Based
Kim (2005)	Enterprise/Business	System Catalogues Relationship Content Data Lineage Technical Data Usage System Process
Bretherton (1996)	User's Perspective	Control Guide
Bowes (2006)	Digital Libraries	Descriptive Administrative Structural
Hodge (2001)	Digital Libraries	Descriptive Administrative Rights Management Structural
Bassiliades et al. (2003)	Education	Educational Metadata
Gillman & Apple (1997)	Statistics	Microdata Macrodata
Prahallad et al. (2006)	Digital Libraries	Descriptive Structural Administrative
Najjar et al. (2006)	Internet	Attention Metadata
Dushay (2002)	Digital Objects	Administrative Structural
Chou & Goethals (2009)	Digital Document Preservation	Technical
Al-Khalifa & Davis (2006)	Websites	Semantic metadata

In the next section the more comprehensive, context specific, definitions and types of metadata are discussed in detail. Although a number of different contexts are discussed, by the end of this section an attempt is be made to identify 'common' or 'core' metadata types.

2.3.2 Metadata Type discussions

It follows from the literature discussed in the previous section that there are standards available for metadata but no standard metadata. Various types of metadata exist within the academic literature as shown in Table 2.5, but these different types vary not only in name but also within its context of usage.

With all these different types and different usages of metadata it is not surprising that someone may become confused between the different options. This section is intended to discuss the different types of metadata as seen within the academic literature, to get a better understanding of these different types and to provide an explanation of what metadata in terms of this study means.

For webpages Najjar et al. (2006) identify Attention Metadata, which is data that keeps track of the content that a user or the entire user base and the number of accesses to a given website. Metadata also appears on the internet in the forms of meta-tags which can be added to webpages that describe the contents of the webpage for search purposes (Eyenbach, 2003).

From a data warehousing perspective Huynh, et al. (2000) defined 2 types of metadata namely: static and dynamic.

Static metadata as its name implies hardly ever changes during runtime and is primarily used to navigate and document the data within warehouse. An example of static metadata is the name of a store. The sales data for the store might change with subsequent transactions but the store name (or some other unique identifier) is more or less fixed and can be used to find and navigate its relevant data (Huynh, et al., 2000).

The second type of metadata defined by Huynh et al. (2000) is dynamic metadata. Dynamic metadata is the opposite of static metadata in that it changes frequently and usually changes at runtime. Examples of dynamic metadata in the data warehouse are things such as frequency of access or what is the most accessed piece of data, which changes over time.

Sun (2005) defined what constituted metadata for reporting purposes, although Sun doesn't specifically label this as reporting metadata. Amongst others, Sun defines the following reporting metadata: data about the report that includes the title of the types of web controls (if any) as they appear on the report; the format and display properties, such as legends, titles, captions etc.; limits placed on which data can be accessed and used to create the reports, for example only data from a specific time or on a specific topic, and data on tabular reports such as column information.

Digital libraries and metadata is a fairly well covered topic in the literature, especially since the creation of the Dublin Core metadata standard which over the years has found wide scale use. Most people when talking about metadata tend to refer to digital libraries or the Dublin Core as what they perceive to be metadata. Only three articles were sourced on the topic of metadata in digital libraries, since the focus of the research was not digital library metadata and the sourced articles were deemed detailed and informative enough, the reason for this could be the saturation of available digital libraries and metadata literature.

In the case of the digital libraries context, both Bowes (2006) and Prahallad et al. (2006) agree on three primary types of metadata namely: descriptive, administrative and structural metadata. Although Hodge (2001) attempted to discuss metadata in a broader context, beyond the digital library context, he listed the same three (descriptive, administrative and structural) but added an additional type of metadata namely right management metadata.

Both the articles written by Bowes (2006) and Prahallad et al. (2006) provided a similar definition of descriptive, administrative and structural metadata.

According to Bowes (2006) and Prahallad et al. (2006) descriptive metadata is used to discover and identify digital content, for example, things such as author name or the title of the digital library content. In order to facilitate the discovery and identification, descriptive metadata describes the content, quality and context of data resources (Patridge & Namulanda 2008). Patridge and Namulunda (2008) define descriptive metadata as the data which describes data in terms of its context, quality, content of a given data source. The purpose of descriptive metadata is to identify resources, facilitate operations on data resources and discover data resources.

Administrative metadata relates more to the technical details of digital objects such as creation dates, file types (Dushay 2002; Bowes 2006; Prahallad et al., 2006). It is also noted in the literature that administrative metadata is responsible for access rules as well (Bowes 2006; Prahallad et al., 2006). Hodge (2001) however identified Right Management metadata and assigned it the responsibility of managing who had the necessary rights to access and manage a given piece of content, information such as intellectual property rights or distribution rights. Seemingly Right Management metadata is more of a subset of administrative metadata than a separate type.

Structural metadata related to how files were organized or put together to form the digital objects within the digital library, a given library object might consist of text, video or audio that needs to be structured and organised in a specific way to be most useful (Bowes, 2006; Prahallad et al., 2006).

While describing metadata for digital objects Dushay (2002) provides a definition of structural metadata as data which describes the relationships amongst digital components using hierarchies, labels or both. This definition of structural metadata strongly resembles the definition provided by the digital library literature; more than likely because digital libraries contain digital objects to which Dushay (2002) is most likely referring. The definition of structural metadata also has a resemblance to Sun's (2005) reporting metadata, which defines which data is to be displayed on a report and how it is to be formatted and presented, although only at a far higher level of abstraction.

Bowe's (2006) definition of structural metadata does however not relate directly to digital objects but rather to data itself. Bowes (2006) defines structural metadata as the data which describes the logical and physical structure of data and ideally how these data-elements are used and perceived as part of the user interface.

From the sourced literature it seems that structural metadata can be on both the data-level or on the digital object level, defining the relationships between data or digital object. Whether structural metadata only considers the relationships or whether it contains also encompasses behaviour to related digital objects or pieces of data is seemingly unclear from the literature.

While discussing clinical applications of metadata Brandt et al. (2003) classified and also defined two types of metadata: descriptive metadata and process-related/technical/active metadata.

The descriptive metadata definition given by Brandt et al. (2003) was notably similar to the digital library context's descriptive metadata in that it is used to describe content although Brandt et al. (2003) describe metadata in a broader sense. They argue that the definition of descriptive metadata goes beyond simply describing content and identifying content, which caters for and supports user interaction and usage, but also includes semantic descriptions and controlled vocabularies, which gives additional computer process-able meaning.

The second type of metadata that was classified by Brandt el al (2003) is Process-related metadata also known as Technical or Active metadata. Technical metadata was defined as being the data which is intended to support the operations of software applications, for example, definitions of what constitutes valid input user input.

In terms of statist creation and management, statistical metadata exists. Gillman and Appel (1997) defines two types, namely: microdata and macrodata. The microdata metadata type is defined as metadata that relates to the micro-level of the population for which statistics are being gathered, the individuals or households etc. The macrodata was described as metadata that was derived and aggregated from the microdata. In general Gillman and Appel (1997) describes statistical metadata as the information which is required to use statistical data. More than likely this use refers to identifying and aggregating.

Statistical metadata seems to be a subset of the descriptive metadata definition from the digital library discipline, especially the microdata which seem to describe the units of measurement for the statistics. Although not explicitly stated in the literature, a general definition provided of statistical metadata as being data/information required to use statistical data might also imply that the micro- and macrodata described falls partly under administrative metadata.

As such it's very likely that the defined statistical metadata types are more than likely specialization, primarily those defined in the digital library discipline, and not new types to the catalogue of metadata types.

Bassiliades et al. (2003) describe educational metadata as data which has the ability to describe the entities involved in the education. This again in part seems to link back the descriptive type of metadata and might be, much like the statistical metadata, simply a specialization and not a new type.

Within the database context Kim (2005) defined eight types of metadata: (a) system catalogues metadata, data descriptors that include columns names involved in database-table relationships, column attributes, and information on when compiled code is invalidated, (b) relationship metadata, data on the relationships between data entities, the types of relationships and semantics of the relationships, (c) content metadata, which describes the content in the database (database-columns, database-tables etc.) ranging from keywords to complex business rules, (d), data lineage metadata, which contains information on the previous changes made to a data entity, (e), technical metadata such as format, encryption and similar technical information, (f), data usage/business metadata which describe the usage and purpose of data by the user or application, (g), system metadata, describes system environment such as the hardware and software, (h), process metadata which describes processes, process steps and process outputs.

Of the eight types of database metadata that Kim (2005) listed, it was noted that system metadata and process metadata have the weakest case for being defined as metadata and in actuality system and process metadata can more likely be seen as data in their own right.

The database metadata has many similarities with the core three digital library metadata types (administrative, descriptive, and structural) but are less generalizable, as many relate specifically to aspects of database also by using the digital library types to categorise the database metadata types. It is clear that many of the database types fall within more than one of the digital library types, for example data usage metadata contains administrative qualities as well as descriptive qualities and relationship metadata has structural and administrative qualities.

Some of the sourced literature discusses more generic types of metadata. Two types of metadata, namely guide and control metadata is indicated by Bretherton et al. (2002). Guide metadata is described as being intended for use by humans and is usually stored in a natural language or free text format. Control metadata on the other hand is defined as being the opposite of guide metadata, intended for usage by computer systems. Bretherton, et al. (2002) states the differences between guide and control metadata are not fixed, and that it is possible for them to overlap.

Kerr (2008) distinguishes between two types of metadata namely: back room metadata, which provides data related to process usage (extraction, cleaning, loading), and front room metadata which helps with querying data and is more descriptive in nature.

Seemingly back room metadata and control metadata is equivalent as both are intended for use by the computer, and not a human user, and are intended to enable some form of computer process. Front room metadata and guide metadata are seemingly equivalent as both are intended to guide and aid human users and both are fairly descriptive in nature.

Guide metadata has strong inclinations to descriptive metadata, while control metadata seems to encompass administrative, structural and perhaps even the right management metadata.

It follows from the literature reviewed that most of the defined types of metadata relate to each other in some way. Furthermore the metadata types defined in the digital library discipline seem to play a central role in defining many of the additional types of metadata, especially the core three, namely: descriptive, administrative and structural. But on a higher level these different types (descriptive, administrative and structural) can be further divided into the categories of guide and control metadata.

Without attempting to label the interpretations of metadata Bretherton et al. (2002) gave several examples of user roles and how these relate to metadata. For the 'intelligent novice' metadata can simply be a data which points to more detailed information. For the database manager metadata relates to the database schema i.e. names of database-tables and naming conventions and relationships amongst database-tables. For the computer system engineer physical level information allows him or her to manage a system, things such as file types and formats are considered metadata. These usages however seem to be related to guide metadata.

For the intelligent novice the relevant metadata is seemingly descriptive metadata, providing more information and insight. For the database administrator it is descriptive (such as database-table names and descriptions) and structural metadata (the database-table structures and data relationships). For the computer system engineer the relevant metadata seems to be descriptive and technical metadata and possibly administrative metadata.

Beyond the metadata types identified by Hodge (2001), Al-Khalifa and Davis (2006) propose an additional characteristic that can be applied to metadata namely, semantics. Al-Khalifa and Davis (2006) define semantic metadata as data which relates web resources to a number of classes and properties defined in an ontology.

Hodge (2001) proposed two views that could be applied to the semantics of metadata namely, a “minimalist” view and a “structuralist” view. The minimalist view is aimed at keeping the elements of the metadata to a minimum and the semantics and syntax of the metadata simple. The structuralist view on the other hand attempts to create a finer semantic distinction which can easily be extended for a particular community.

Al-Khalifa and Davis (2006) define the types of metadata semantics as: implicit, formal and powerful. The implicit metadata semantics appears in text documents and is not rigidly defined. Within the formal type the metadata representation is more structured and definable. Finally for the powerful type of metadata semantics the meaning of complex data and metadata structures is represented by a number of related simple syntactic structures.

Much in the same way that data will always have metadata, metadata will always have some form of semantics. The implicit metadata semantic types indicate this fact, but in the case of implicit metadata semantics these semantics are not officially captured and worked into the metadata. The remaining two metadata semantics can easily be worked into the view proposed by Hodge. The Minimalist view encompasses the formal metadata semantics while the Structuralist view encompasses the powerful metadata semantics.

From the sourced literature it seems clear that metadata has found a wide range of usages in a number of very different contexts and fields. Metadata within the digital library context is fairly well-developed defining three core types of metadata: descriptive, administrative and structural. The digital library core types can easily be generalised and can serve as a means to categorise metadata types within other contexts. There are however notable exceptions since after all descriptive, administrative and structural were originally intended to be used in digital libraries and as a global standard or means of categorisation.

The different types of metadata can however be categorised in terms of their purpose and usage, specifically in terms of guide metadata, which is intended to aid users and to be user understandable, and control metadata, which is intended to be used by computers and to be computer process-able. Overlap does occur when using guide and control metadata to categorise the types of metadata but these are far rarer then when using descriptive, administrative and structural as a means of categorisation, or is at least less likely based-on the types identified above. This idea of categorising the different types of metadata based-on the intended use and by who/what (user or computer) was also voiced by Kim (2005).

This section looked at some of the different types of metadata defined within the academic literature, but was not intended to be a detailed and thorough look thus the possibility does exist that even more types of metadata exist in the greater body of academic knowledge. Because of this the section attempted to create some sort of categorisation of different types of metadata to at least help better define the intended type of metadata to be used during the research.

The following section considers some of the standards which govern the different types of metadata.

2.3.3 Metadata Standards

As discussed in the previous section there are a number of metadata types found in the academic literature and because of the vague definition of metadata these types are defined in accordance with the context in which the metadata is used as well as the intended use of the metadata. The sourced literature also pointed to a number of different standards which governs the usage of metadata-elements in their given context.

This section considers the metadata standards found in the sourced literature as part of the technological due diligent in order to understand which standards are available and possibly relevant to the research.

Metadata standards normally take the form of a metadata schema. A Metadata Schema is a specification for an object or a piece of data (Godby et al., 2003). An example would be the 15 predefined data-elements of the Dublin Core used within the digital library context (Thirion et al., 2006) or the specification for the name/value pairs found in the On-line Public Access Catalogues standard (Marshall, 1998).

According to Al-Khalifa & Davis, (2006) a number of organisations were, and still are, involved in the production of metadata standards. Gillman and Appel, (1997) as well as Hatala and Rochards, (2002) also refer to the active role that both private and public sector organisations have had in the production of standards for metadata. The development and adaptation of metadata standards is a continuing as a result of the ever changing contexts in which metadata is being applied (Greenberg, 2004).

Because of the abundance of metadata types it is important for a given standard to be: "clear, comprehensible, consistent, complete, flexible and simple to use" (Timpf, et al., 1996). As such a given standard must include specification for the modelling and exchange of metadata and the knowledge ontology and vocabulary (Kim 2005)

Table 2.7: *Metadata standards found within the literature.*

Metadata Standard	Used In	Literature Sources
Dublin Core	Digital Library: indexing, retrieval, storage of library objects	Kerherve (1997) Onyancha, et al. (2001) Paivarinta, et al. (2002) Thirion et a (2003) Wagner, Weibel (2005) Prahallad, et al. (2006) Patridge (2008)
Metadata Object Description Schema (MODS)	Digital Libraries, descriptive	Greenberg, (2005) Prahallad et al. (2006)
Metadata Encoding and Transmission Standard (METS)	Digital Libraries	Dushay (2001) Gartner (2002) Cundiff (2004) Bowes (2006) Prahallad, et al. (2006)

Table 2.7: *Metadata standards found within the literature (continued).*

Metadata Standard	Used In	Literature Sources
Catalogue et Index des Sites Medicaux Francophones (CISMef)	Describe and index French-Language health resources	Thirion et al. (2003)
Machine Readable Cataloguing (MARC)	Digital libraries, stores bibliographic information in predefined tags and subfields	Allen & Schalow, (1999)
FGDC	geographic information systems	Kerherve (1997)
CDIF	modelling tools	Kerherve (1997)
MDIS	CASE systems	Kerherve (1997)
Learning Object Model (LOM)	E-Learning	Thirion et al. (2003) Rehman et al. (2005) Al-Khalifa & Davis (2006)
AttentionXML	User activity on Web Sites	Najjar et al. (2006)
Metadata for Images in XML (MIX)	Define the structure of Images	Chou & Goethals (2009)
TextMD (Text Metadata)	Define structure of Plain-text.	Chou & Goethals (2009)
FOXML	Fedaro open Source digital repository software	Bowes (2006)
MOA2	Digital object modelling and data access	Dushay (2002)
HIDDEL	Trust and Quality of Health information services	Thirion et al. (2003)
Medical Core Metadata (MCM)	Health Resources, enhanced retrieval over the internet	Rehman et al. (2005)
Cooperative Online Resource Catalogue (CORC)	Digital Libraries. Helps build databases of shared cataloguing for web access. Supports MARC and Dublin Core, exports and imports.	Hodge (2001)
Common Data Standards metadata descriptors	Health Informatics. Used to describe health data, primarily cancer data, to exchange aggregation and use. Developed by National Cancer Institute.	Patridge & Namulanda (2008)

Within digital library systems one of the more common standards is the Dublin Core. The Dublin Core was created in 1995 during a workshop held in Dublin, which the standard is named after (Prahalad, et al., 2006). The Dublin Core standard was created with the intent of it being a simple and concise standard which authors of web resources could use to describe their own web-based documents/resources (Hodge 2001; Patridge, Namulanda, 2008; Prahallad et al., 2006)

The Dublin Core defines 15 elements for the identification and management of digital library objects being: an author/creator, data, description, format, identifier, language, publisher, resource type, rights, subject/keywords, title, contributor, coverage, and relation, source (Thirion, et al., 2003). These core elements are all optional and repeatable, allowing the Dublin core to be as simple and flexible as needed, for this reason the Dublin Core is also sometimes used outside of the library domain for describing and cataloguing (Prahallad et al., 2006)

The Dublin Core is primarily aimed at document-like objects, which seems only natural seeing how it was intended for web-based documents stored with a digital library. As such the ability of the Dublin Core to support non-document-based resources is directly proportional to how closely the intended resource resembles a standard Dublin Core compliant document (Rehman, 2005).

CISMeF (Catalogue et Index des Sites Medicaux Francophones or translated in English the, Catalogue and Index of French) is an specific implementation of the Dublin Core, containing all the Dublin Core elements with the addition of: contributor, coverage, relational, source. Unlike the Dublin Core, CISMeF is used to for the purpose of indexing and cataloguing health resources in French (Thirion et al., 2003). The Dublin Core and CISMeF are fairly similar only differing in the domain of implementation.

Machine Readable Cataloguing (MARC) is another standard used in library systems. MARC uses predefined tags and subfields to store bibliographic information relating to the object within the library system. MARC intent is to enable automated manipulation for the use of compared applications (Allen & Schalow, 1999).

Metadata Encoding and Transmission Standard (METS) is a standard used to create XML documents which show the hierarchical structure of the objects within a digital library. The METS contains information about the names and locations of the actual files which make up the digital objects contained within the library systems (Cundiff, 2004). METS is primarily a structural metadata standard although it does make provision for administrative and descriptive metadata (Bowes, 2006; Cundiff, 2004), meaning that METS handles all three of the major types of metadata used to describe digital library content (Gartner, 2002).

Learning Object Model (LOM) defines a number of categories with 80 elements in total for storing information relating to e-learning object. The categories defined by LOM include: generalised, lifecycle, meta-metadata, technical education, rights, relations and annotation. LOM also contains all of the Dublin Core elements (Thirion et al., 2003).

An International Standard Organisation (ISO) standard for metadata is also available, named ISO 11179, a standard that is intended to define metadata semantics (Park, Kim, 2006).

This section looked at some of the more prominent standards that govern metadata application in order to see which standards are available and applicable to the research. Most of the articles sourced while looking for metadata standards related to standards in digital libraries, primarily Dublin Core. Because of the abundance of digital library standards it is obvious that a lot of standardisation work has gone into metadata within the digital library context. Even applications of metadata standards in other context such as the healthcare context are also based-on the Dublin Core standard, an example of which was CISMeF.

This seems to indicate that there is room for more research in the realm of metadata standards beyond the Dublin Core. Although this also means that if the Dublin Core is inapplicable for research, such is the case with this research, then little formal well-developed standards can be drawn on.

The following section looks at the practical applications.

2.3.4 Metadata usage

The previous sections attempted to define metadata, the different types of metadata as well as metadata standards in order to gage the relevance to the research. It was previously touched on but this section looks specifically at the applications and uses of metadata and how it could be relevant to the research. This section focuses mostly on the usages of structural metadata, since it is intended to be used in the research, but does touch on other usages beyond those of structural metadata.

From the academic literature it is clear that metadata and the problems it aims to solve, has been around for more than 40 years (Bernstein, 2000). As seen from the sourced literature, metadata has a fairly vague definition, because of this metadata can be categorised into a number of different types. As a direct result of having so many different types of metadata there are also a number of different usages and a number of different standards guiding these different applications.

Much as all data does not have to be universally available, metadata itself does not need to be universally available and can easily be local or private (Marshall, 1998) serving only the need of a given system of process, for example control metadata (Bretherton et al., 2002) / Back Room Metadata (Kerr, 2008) contains the data needed to enable certain processes or computer activities. This information does not need to be made available to the user or any none related processes and activities.

Metadata exists within most knowledge management systems, either implicitly or explicitly and plays an important role in the systems that utilise it since the metadata codes the system's knowledge and directly relates to the system's integrity (Brandt et al., 2003).

Metadata is essentially data about data as has been defined in multiple academic sources (Timpf et al., 1996; Eyenbach, 2003; Smail-Tabbone, 2005; Kim, 2005). Additional information allows the data to be more usable for a specific purpose or in a specific context by either users or computers (Kim, 2005).

It was stated by Rizvi & Imdadi (2008) that the representation of data is metadata. Srivastava & Velegrakis (2007) saw metadata as a means of clarifying semantics. Schuurman & Leszcynski (2006) went further than that and said that metadata can be used as a mechanism to convey both semantics and ontological information about data. According to Aumeller (2005) data can only be semantically rich if it was annotated with metadata.

Semantic annotation is especially attractive since most relational databases do not normally contain text annotations for the relationships amongst the database-tables which described the semantics of the relationship and the database-tables (Marenco et al., 2002).

Although metadata does have the ability to resolve issues around semantics, Van Ginneken (2003) noted a major issue when using metadata convey semantics. In order to detect semantic equivalents Van Ginneken (2003) noted that an extensive and complete definition of all possible synonyms was needed, which would not be feasible in a large complex domain such as healthcare. Van Ginneken (2003) proposed that metadata rather be used to limit what can and can't be said in a specific context.

Various advantages of using metadata to annotate semantics was detailed by Al-Khalifa & Davis (2006): semantic metadata is machine process-able, because its semantics is derived from an ontology guaranteeing well-formed meaning. Semantic metadata is flexible and extensible, meaning that it can be easily annotated with additional metadata and even allows mixing of ontologies. Semantic metadata has the advantage of added interoperability between systems with a common ontology.

A number of management systems such as database management systems or knowledge management systems in fact use metadata without explicitly referring to it as metadata or identifying it by any other name (Kerherve 1997, Kim, 2005). Within a database management system there is clearly a distinction made between the data and the structures in which the data resides as well as the relationships between these data-elements, clearly an example of structural metadata although it is rarely explicitly called metadata (Kerherve 1997; Bowes, 2006)

A number of examples of metadata within a database is provided by Srivastava and Velegrakis (2007) including: schema, integration constraints, comments about data, ontologies, quality parameters, annotation, provenance and security policies

Whereas Bowes and Kerherve identified the metadata within the relational database system, Marenco et al. (2002) rather referred to it as being meta-knowledge, but he also identified it as being implicitly captured within the schema of a database, within the structure of the database-tables and within the relationships amongst the database-tables.

Al-khalifa and Davis (2006) identified a very similar purpose of metadata as the ones found within the database management system, but did not specifically make reference to it in the context of the database management system but rather in the context of an e-learning system. Al-khalifa and Davis (2006) noted the following uses: to describe the content, format, purpose and structure of data.

The above usages identified from the sourced literature are all consistent with structural metadata. But structural metadata doesn't simply exist in database and knowledge management systems also contain a number of usages. HTML tags for-example can be defined as structural metadata since the HTML tags define the physical appearance and formatting of data (text, images etc.) within a web page (Dushay 2002)

Structural metadata can also be used in a number of other contexts including: information warehouses, relational databases and operating system file tables (Bowes, 2006)

One of the more common explicit usages of none structural metadata is for the identification, management and navigation of data according to Kerherye (1997). Smail-Tabbone (2005) assigns similar roles to metadata being: indexing, documentation and retrieval. Srivastava and Velegrakis (2007) also add the related functionality of understanding, maintaining, querying, integrating and evolving data, although they primarily look at its usage within a database.

Meta-tagging is an extension of the indexing and retrieval functionality, which according to Hatala and Rochards (2002) can be used to share knowledge. Although Hatala and Rochards (2002) primarily looked at community generated meta-tags being used to share knowledge between communities, this differs from the indexing and retrieval functionality identified by Kerhery (1997) and Smail-Tabbone (2005) only in its scope and focus.

Although meta-tagging is a fairly easy way to add metadata to a piece of data or an object, it does have some shortcomings. For the most part most meta-tags are generated by humans, which leads to problems of subjective interpretation of content as well as deceit, ignorance and sloth of meta-taggers, all of which can lead to poor quality meta-tags. (Hatala, Rochards, 2002)

Also echoing the previous usages of metadata Thirion et at. (2003) noted that metadata on the internet aided in the discovery and control of web resources as well as to enhance the retrieval and navigation of and between these web resources.

While looking at the usages of metadata within library systems, Park (2009) identified several uses which were very similar to the above uses of indexing, retrieval and documentation. Park noted that metadata serves the bibliographic needs within the digital library discipline such as discovery, use, provenance, currency, authentication, administration of digital library object.

Specifically within database systems Patridge and Namulanda (2008) identified the following usages: Discovering datasets, identifying datasets, assembling similar datasets, distinguishing dissimilar datasets and providing access to datasets.

The above usages all have to do with the management of object (be it data object or digital object) in term of usage. Clearly these usages refer to descriptive metadata used in library systems (Dushay, 2002) which describes data for human usage (as discussed above), and technical metadata, which contains technical details for machines operations on data (as discussed previously). Both technical and descriptive metadata play an important role in enabling usage of data and digital objects.

For each of the interpretations of metadata given by Bretherton et al. (2002), and discussed previously in Section 2.3.2, a different use can be ascribed. For the intelligent novice, the descriptive and guide metadata is used to find relevant information. For the database manager the technical, descriptive and structural metadata provides information relating to the structure of the database schema as well as details on how to manage the database. For the system engineer the guide, descriptive and technical metadata serves as a means to allow him to manage and create the system.

Hodge (2001) also provides three uses for metadata: as machine understandable information (technical metadata), to describe records which in turn describe electronic resources (descriptive metadata), and as a formal schema for resource description, applying to all objects whether digital or not (descriptive metadata).

Within the context of relational data relationships Srivastava & Velegrakis (2007) identified the following uses of metadata: Tag data with quality parameters; Annotate schema structures with textual descriptions to communicate semantics. Provenance can be provided by metadata as annotation, can store schema and mapping information in special structures, clarify semantics.

This section identified two of the more apparent uses of metadata provided in the literature, specifically the ability of metadata to define the structure and relations that exist between data-elements, and secondly the ability of metadata to help with the management and indexing of electronic content.

The section considers the criteria with quality metadata.

2.3.5 Metadata Quality

The previous sections looked at the different types of metadata and contexts in which metadata occurs (Section 2.3.1), the standards for metadata (Section 2.3.3) and the usages of metadata (Section 2.3.4). This section follows from the previous section and considers what constitutes quality metadata, which is an important part in the management and usage of metadata.

Literature was sourced on the topic of metadata quality for a number of reasons namely: uncertainty that exists about the definition of metadata within the literature. The research considers structural metadata which is not prevalent in the academic literature and finally to create criteria by which the application of metadata in the research can be measures.

The quality characteristics of metadata can be identified in terms of the functional requirements and goals which the metadata must support, specifically the quality of metadata can be measured in its ability to enable or perform a specific purpose (Park, 2009).

In clinical research Chong et al. (2003) defined the following quality requirements for metadata: firstly, it should describe content independent information (location of patient records); domain independent information (structure of patient record); domain specific information (treatment guidelines) at the level of abstraction of the given information. The second quality requirement is that the metadata should contain mappings between the source data attributes and the domain specific terms. The third quality requirement is that the metadata should describe rules for interpreting context specific information. The fourth quality requirement is that the metadata should store data in a machine-readable format. The fourth and final requirement for quality metadata is that it should enable quick retrieval of structured information.

The criteria for good metadata listed by Chong et al. (2003) is in line with the idea of metadata quality criteria being based-on the intended goals of the metadata and the functionality the metadata is intended to support. After all the criteria for quality clinical research metadata isn't necessarily appropriate in other contexts.

Park (2009) defined that issues of poor metadata can be traced back as far as the creation stage. During creation Park (2009) noted two issues that caused poor metadata to be created: first reason was inaccurate data entry, miss typing or poor extraction from source materials and the second reason was an inconsistency in the subject vocabulary, the same object being referred to by different names or usage of a vocabulary that is foreign to a given domain.

Requirements for quality metadata were defined by National Information Standards Organisation (NISO). NISO requirements for good metadata are: the metadata must be appropriate for intended use, support interoperability, conform to community standards, include a clear statement of the intended condition and terms of use, must support long term management, metadata records as objects which must conform to the requirements of quality objects, namely: authority, authenticity, achievability, persistence, and unique identification (NISO Press, 2004; Park 2009).

This section identified some of the criteria for quality metadata in order to create a framework for identifying and measuring the quality of the metadata used within the research.

The following section considers the ways in which metadata can be created.

2.3.6 Metadata creation

This section considers the academic literature on the topic of metadata identification and metadata creation. This section was included in order to understand the origin and sources of metadata and is the logical progression from the previous sections that looked at the types of metadata (Section 2.3.1), metadata standards (Section 2.3.3) and quality criteria for metadata (2.3.5).

Because metadata is data about data it can be difficult to distinguish between what is data and metadata. Marshall (1998) purposed the use of an ethnographic analysis in order to identify new types and sources of metadata, although identifying all the possible metadata-elements and sources of metadata usually occurs over a period of time.

The academic literature primarily looked at the creation of metadata from two angles, manual metadata creation where someone physically inputted the metadata and automatic metadata creation where it is generated by mostly autonomous processes. The sources articles primarily focused on the pros and cons of these methods of metadata creation, and the possibility of using a combination of the two metadata creation methods.

While discussing automatic metadata creation Greenberg (2004) looked at two methods namely: metadata extraction and metadata harvesting. Metadata extraction was the process of using an algorithm to automatically detect and structure metadata from richer (such as narrative) sources, such as the content displayed via a web browser. The process of metadata harvesting is simply the process of automatically collecting and categorising already existing metadata such as the meta-tags in the Hyper Text Mark-up Language (HTML).

In order to test these two different methods of automatic metadata creation, Greenberg (2004) primarily used web resources to demonstrate the effectiveness and usage. Metadata extraction is discussed in several other sourced articles but metadata harvesting is rarely mentioned, more than likely because it is rare to find resources which already have metadata (such as web pages) that can be harvested. It was however noted by Greenberg (2004) that even with resources that already contained metadata such as web-pages, automatic generated metadata was not always ideal especially for the process of resource discovery.

The sourced academic literature also pointed to problems with the process of automatically generating metadata, Ingenerf (2001) noted that there existed well-known problems when it came to the utilization and interpretation of unstructured free text sources. The concern of automatically extracting data from free text sources is because of the complexity of natural language and the difficulty of computer to interpret it. A possible solution is to involve humans at least partly in the metadata extraction process (Park, 2009).

Another issue with the automatic extraction of metadata was the in defining what constitutes a workable structure for the metadata. Even with well-defined metadata standards it can be difficult to find the correct structure for a specific utilisation of metadata especially when dealing with unstructured source material (Marshall, 1998).

There are issues however when collecting metadata at a later stage especially after the data has already been collected. If metadata is generated after the relevant data has been collected, then more effort is required and in some cases the process can be hampered by a lack of information. By extension the same issues can be applied to the automatic extraction of metadata from existing sources (Timpf et al. 1996).

A possible solution to the listed issues of metadata creation Schuurman & Leszcynski (2006) propose the use of ontology-based metadata where metadata is automatically extracted from a formal ontology.

As automatic metadata creation has its advantages and disadvantages so too does the process of manual metadata creation. There are numerous issues surrounding the creation of metadata base-on older systems including: the existence of rigidly structured files which contained the technical metadata needed for the system to operate, and the lack of documentation. Management of the files within order systems proved difficult and semantics could not always be ensured (Brand et al., 2003).

From the sourced material it seems clear that automatic metadata extraction is not yet perfect. A lack of deeper understanding of a given situation and the inability to interpret the complexity of narrative sources means that automatic metadata extraction is not always the most efficient means of creating metadata. A manual metadata generation process itself can become difficult if the maintenance process is overly complex.

Ideally the structures of the metadata would be defined before the data itself has been collected and that the metadata and the data would be collected in unison ensuring that no metadata or data is missing.

Unfortunately this isn't always the case. Normally the metadata exists implicitly and can be difficult to discover through automation. Marshall's (1998) proposition of using ethnographic analysis to identify the relevant metadata seems most logical, but it might also be more time-consuming than desired. Metadata creation can be difficult especially if metadata is being extracted or harvested from some form of pre-existing systems which might be poorly documented, or if free text narrative sources are used.

2.4 Repository

This section considers literature on the topic of repositories as part of the technology due-diligence of the research in order to gain a better understanding of what is available on the topic of repositories in the academic literature. This section doesn't cover the academic literature on the topic of repositories in detail but focuses rather on the success criteria in order to aid in the evaluation of the artefact developed as part of the research process.

A repository, whether it contains digital or non-digital objects, must contain metadata which relates to the objects or to the services which the repository provides. An example is the importance of descriptive metadata like the Dublin Core for digital library repositories (Dushay, 2002).

Because of the many different types of content that can be stored within a digital repository, Thibodeau (2006) noted that the requirements for what constitutes a successful repository can only be derived from: the intended purpose, the context and environment in which it must function and from the intended content to be stored within it. Specifically it is sometimes necessary to add contextual information, such as how, where and when the content was created, to the digital content in order to ensure that the content is useful. Any criterion for success thus is relative to the type of repository, its usage, and the type of content to be stored in the repository.

It was however proposed by Thibodeau (2006) that a successful repository can be measured along 3 axes: orientation, coverage, collaboration.

Coverage related to the objects that the repository is intended to hold and included the acquisition and the usage of those objects. Relevance is the consideration for the coverage, which objects should and shouldn't be acquired (Thibodeau 2006)

Collaboration refers to the environment in which the repository functions, whether it is an isolated system or does it need to collaborate with other systems in order to be able to perform its purpose, essentially where the repository falls in the isolated vs. collaborative spectrum (Thibodeau 2006).

The orientation of most repositories falls into one of two categories: retrospective or prospective. Prospective indicates that the purpose of the repository is to meet the needs of a user community, to provide services and access to the content. Retrospective the focus is on preservation, the content is intended to be stored, with minimal access and infrequent alteration of the content (Thibodeau 2006).

This section looked at the literature on the success criteria for a digital repository in order to create criteria for defining the characteristics of a repository artefact as well as the criteria for the measurement of the success criteria of a digital repository. The following section follows on looking more specifically at the literature on the topic of semantic repositories and some of the requirements for a semantic repository.

2.4.1 Semantic Repository

This section looks specifically at the literature relating to semantic repositories and follows the previous section that focused on repositories in general. This section further considers some of the requirements of a digital repository but narrows to scope of the inquiry specifically to semantic repositories.

A semantic repository was defined by Davies (2005) as a: “Combination of a ‘knowledge base’ and an ontology”. Pantel and Pennancchiotti (2008) defined a semantic repository as a resource that organises knowledge in a highly structured fashion at either a conceptual level (such as concept relations) or a sense level (worked senses and its relations).

The possibility exists of adding semantics to a given repository through the use of ontologies, which would enhance access to knowledge and the management but providing an explicit representation of a repositories conceptual model, according to which knowledge within the repository can be defined (Palmonari, 2008).

However some issues do exist around semantic repositories. Adding and linking data in semantic repository is both time-consuming and error prone especially when a large volume of data was to be inputted or a large number of people was involved in inputting data. However most of these usability issues, around adding and linking data, arose from a poor user interface design or having to work with overly complex ontologies (Keller et al., 2004).

Usability is notable concern within semantic repositories, although it can be argued that the reason for this was because repositories were handmade and limited in their size and scope. Essentially the usability issues within a repository arose because the semantic repository was created for a specific person or persons by people who might not have been able to envision every single possible use or requirement for the repository (Pantel & Pennancchiotti, 2008).

One way that the literature purposed to solve the usability issues of semantic repositories was ensuring a more visually rich interface that allowed users to better understand the overall topology of a given ontology and furthermore allowing users to easily navigate between their local and a global perspective (Keller et al., 2004). Ideally the more visually rich interface would allow users to better understand how elements of an ontology fit together and improve their users' ability to input the correct data and correctly relate it.

Palmonori (2008) also emphasises the importance of easy navigation within the semantic repository. Palmonori (2008) stated that navigation within repository allowed the user to become more aware of the knowledge within the repository which in turn enabled the user to retrieve more accurate information.

Clearly from the literature there are several requirements that a semantic repository needs to meet: the ability to provide users with a comprehensive yet understandable view of the data it contains (possibly using a more visually rich user interface), some form of explicitly captured ontology, and the ability to track and correct errors with added and linked data.

The following section looks more at the topic of semantics, in order to provide further clarification of some of the topics raised in the sourced literature and also continues on from the literature on HI, specifically HI integration detailed in Section 2.2.4.

2.5 Semantics

As discussed under HI interoperability semantics plays an important role in enabling system integration, and ensuring that semantics is maintained while transferring data. But the research does not focus directly on interoperability in the healthcare domain but rather at the underlying issues that cause the issues.

The field of semantics in IT is however a complex and multifaceted one with a lot of different branches of research and inquiry. Thus this section covers the topic in more detail in order to gain a better overall understanding, before attempting to relate it back to the current research.

Semantics in IT as we know it today owes much to the semantic web concept proposed by Tim Berners-Lee, although the semantic web still has to find mainstream usage or implementation (Keller et al. 2004). The concept of the semantic web is an extension to the current World Wide Web, adding an additional layer of semantics to it. The idea was to take the masses of information currently available on the internet and transform it from human readable to machine readable (Euzenat, 2001; Eysenbach 2003) using a critical mass of metadata to structure the information (Kiryakov, et al., 2003)

The semantic layer was the key to achieving the goals of the semantic web. In a none-IT based context semantics is a study of the meaning of word and their relations (Inmon et al. 2008). The semantic web would thus be able to pull together semantically related information sources, and distinguish the meaning of them. 'ABC' would no longer simply be 3 letters but could mean the 'alphabet' or 'basics of…' or a host of other things.

As has already been stated, semantics also appears as an important part of interoperability on the system level. In order to ensure effective and efficient interoperability between systems the systems need to do more than simply send and receive data, they need to understand the semantics of the data being transferred. A means of ensuring that data is semantically rich is to annotate the data with metadata (Aumeller 2005).

Semantics with the field of IT is still a very vague concept used in several different contexts over the years (Schorlemmer & Kalfoglou, 2003, Al-Khalifa & Davis, 2006) having different interpretations for the different areas of computer science in which it appears (Al-Khalifa & Davis, 2006).

Ginneken (2003) identified two levels of semantic redundancy, on the conceptual level where two or more concepts are equal but treated as separate, and at the composition/context level, where a given concept can be different depending on the context in which it occurs.

Redundancy at the conceptual level was fairly easy to identify and resolve with sufficient definition of synonyms, although at a composition/context level such redundancies can be far more difficult to identify and manage. It was also noted in the literature that the medical domain is particularly vulnerable to the issues surrounding semantic equivalents because of the numerous context and interrelated aspects, such as symptoms having many causes (Ginneken 2003).

Mead (2006) states simply that semantics is meaning. And as discussed previously in, Section 2.2.4 under the heading Integration of Health informatics Systems, it is important for two systems to be able to not only share information but also to understand and utilise the shared information.

Semantics is an important part of interoperability in that for two systems to fully interoperate they need to not simply be able to share information, but they need to both understand what is being shared and interpret the information the same way. In order to be able to use the information correctly (Ingenerf et al., 2001), they need to be able to communicate at a semantic level (Euzenat, 2001).

Semantics has become a far wider field and has become more synonymous with ontologies and through the usage of standard ontology definition languages (Smail-Tabbone, 2005).

An ontology describes and models a given domain. Semantics helps in this regard since it allows the same element in the ontology to be known by multiple names. A customer and a patient as an example from the healthcare domain is the same thing and has the same constraints and attributes but if semantics are ignored the same thing might be treated as two separate entities (Dogac, 2004).

Semantics was viewed by Schuurman & Leszcynshi (2006) as being seeped in context with no two contexts necessarily having the same levels and types of semantics. They saw that semantics was inadvertently linked to the ontology of object classification and the concepts of object definitions.

The lack of an adequate representation of semantics, including the relationships which exist amongst concepts, was viewed by Ingenerf (2001) as one of issues that prevented traditional systems from being adequate for more 'ambitious' usages.

As discussed previously in Section 2.2.4 under the topic of Integration of Health Informatics Systems, semantics which is simply a common understanding of data, is the key to achieving interoperability, but that most heterogeneous systems clearly fall short of this ideal. This problem might be as a result of the numerous differences in the way that medical concepts can be expressed (ingenerf et al., 2001), no particular one being superior to the other.

This issue was also raised by van Ginneken (2003) while discussing semantic equivalents. Van Ginneken (2003) wondered if it was truly possible to avoid these semantic equivalents in the healthcare domain where concepts can play many different roles in many different contexts, leading to a number of issues surrounding semantic misunderstanding.

It was noted by Marenco et al. (2002) that users (presumable with little experience in data management) have the tendency to think of the connections between concepts within their domain in terms of semantics and not as the conventional primary key and foreign key relationships common to relational database systems.

It was proposed by Davies (2005) that well-defined semantics allowed for the relationships between elements and concepts to be represented in a more meaningful way; furthermore Bird (2003) spoke of the advantages that a visualisation approach offered in the management of complex data.

Of course the visual representation of semantics and semantic relations between concepts strongly relates to the concept of ontologies which essentially do the same. Schorlemmer and Kalfoglou (2003) identified that semantics can easily be expressed in a common ontology which would allow all concepts to be easily assimilated in other ontologies.

The idea of representing semantics visual through ontologies is also supported by other academic literature. Rizvi and Imdadi (2008), view the semantic web as a collection of corresponding ontologies originating in different domains and from different communities, using metadata descriptions to construct such ontologies.

The issue of semantic is a notable issue in terms of interoperability. Semantics is simply the meaning of data relating to different meanings that can be placed on a single data-element of a group of data-elements. As has been noted in this section semantics is important in interoperability since systems that share data need to assign the same meaning and interpretation to the data being shared. As mentioned metadata can aid in creating semantically rich data but the overall management of semantics can be difficult, but the literature points to the usage of visually rich user interfaces and the application of ontologies in order to aid in the management of data semantics.

The following section discusses the topic of ontologies in more detail.

2.6 Ontology

Ontologies have a strong link to semantics, as discussed in Section 2.4, and it further plays a role in interoperability, as is discussed in this section. In IT ontologies play an important role to enable interoperability as discussed above, along with semantics. Ontologies also arose as a topic of interest in an attempt to better understand the discrepancies between the definition and usage of data. Because ontologies are a representation of the data and data relations within a given domain (Farquhar, 1995; Chandrasekaran, 1999) it was assumed that ontologies would offer a means of understanding the data and usage. Ontologies also provided an interesting insight into how domain specific data was organised and used.

Within this section the literature on the topic of ontologies is reviewed and discussed. This discussion looks at what ontologies are and different usages and advantages using ontologies and finally at the usage of metadata within ontologies.

In philosophy the word ontology denotes the study of the nature of items in existence. In some disciplines of IT such as Artificial-Intelligence however numerous contradicting definitions exist (Noy & McGuinness, 2004). A general definition of an ontology within the IT literature refers to the body of knowledge that describes objects within a specific domain such as healthcare or business and the relationships amongst these objects (Chandrasekaran, 1999). Farquhar (1995) provided a simpler definition stating simply that an ontology is “a combination of terminology and definitions”.

An ontology thus describes the general objects, terms and concepts for a given domain but does not specify the instances and values of these items, i.e. it specifies that medication exists within the healthcare domain but not a specific instance of medication such as aspirin (Musen, 1998).

According to Devedzic (2002) an ontology can provide a number of useful functions: firstly, a design rational for a knowledge base; secondly, a definition of the essential concepts within the domain of interest and thirdly a tool to enable more disciplined knowledge base design and fourth a tool to enable knowledge accumulation.

These useful functions of an ontology provided by Devedzic (2002) most likely arise from the fact that ontologies explicitly define the concepts and terminology within a given domain, or domain specific solution, such as a software program.

It is possible for an ontology to range in level of abstraction but no ontology can define the entirety of a given domain, or more accurately the amount of effort required to define all the objectives and terms within a given domain is simply not feasible in most situations. Therefore ontologies are more suited for specific uses such as modelling a subset of a domain or specific processes and activities (Chandrasekaran, 1999; Devedzic, 2002).

According to the literature ontologies can include the following objects: classes; instances of something; relationships amongst things; the properties/values of a thing; functions and processes of a thing and the constraints and rules involving the object (Obrst, 2003). These items are partly agreed upon by Kuhn and Guise (2000).

However these objects listed by Orbst (2003) seem to be more closely related to the terms and concepts involved in application and database development, and seem to serve as an example of a fairly specific context, that of a domain model.

An ontology attempt to provide meaningful relationships between the entities that it defines to gain additional benefit and usefulness (Chen, et al. 2003). An ontology helps to structure knowledge and aids in the analysis and comprehension of this knowledge (Chandrasekaran, 1999). An ontology can thus help to not only generate semantic data but is also an important tool for the representation of semantic data (Rizvi, 2008).

A number of reasons to develop an ontology exist (Noy & McGuinness, 2004): first, it allows for a common understanding of information; enables sharing of sharing knowledge (Jones, et al., 1998); second, enables reuse of domain knowledge (Jones, et al); third, makes assumption of the domain explicit, fourth, a means of separating domain knowledge from operational knowledge and finally a means to analyse domain knowledge.

It is widely recognized that an ontology plays an important role in the development of knowledge based systems (Jones, et al., 1998) and all software for that matter, since software needs models of the domain in which they function (Devedzic, 2002)

Whether explicit or implicit when a user interacts with a system he is in fact interacting with the ontology of that system. The ontology of user interface defines the terms and images used on the interface and the meaning of those terms and images (Stead, 2000).

An ontology allows a software developer to formalize the concepts of the application domain (Kuhn & Giuse, 2000) conceptualising facts about a domain such as the structures in which data is to be stored (Stead, 2000). Ontologies are described by Musen (1998) as being the core of any working HI system.

This explicit conceptualization of domain is especially useful for the development of intelligent systems because ontologies have the advantage of being both human and computer interpretable, thus making it an effective tool for building knowledge bases (Musen, 1998).

By defining common elements within a domain such as patient, medication or caregiver, an ontology serves as a communication medium helping both sides of the conversation (be it human or IT) to know what constitutes a patient, medication or a caregiver.

Ontologies reuse is also important in the development of intelligent systems (Musen, 1998) since two systems working in the same domain can conceivable use the same ontology, or at the least an ontology will only require minimal change.

Ontologies appear in software systems and solutions, as discussed in this section, and explicitly define the relevant concepts and terminologies that the solution encompasses. Logically then it follows that ontologies do not only have to be applied to IT-based solutions. In a domain such as rural HBHC, discussed in more detail in Section 2.1.2, where IT penetration is low and paper based systems are primarily used there also exists a number of ontologies. After all HBHC is a subdomain itself of healthcare, it has its own terms, concepts and relations between these concepts. These paper-based systems used in rural HBHC also then contain ontologies, just as any IT-based solution would have were it to attempt to fill the same role.

Because ontologies define concepts within a given domain it allows for concepts to be mapped across different ontologies allowing heterogeneous systems to effectively interoperate (Farquhar, et al., 1995; Hartmann 2005; Martenson & Horndahl, 2005). The ontology can provide a common data representation amongst the different ontologies and can deal with the semantics involved in mapping objects between ontologies (Davies et al., 2005).

Sharing between ontologies is simply the process of aligning the concepts defined in the different ontologies (Kiryakov, et al., 2003). Alignment involves finding commonalities within the different ontologies, or transforming one or both of the ontologies in order to create commonalities. When these commonalities exist data can successfully be transferred between these ontologies.

Ontologies are not static, or more correctly the domain it represents is rarely ever static, thus the ontology needs to be updated to represent changes in the represented domain (Rizvi, 2008).

The process of ontology evolution or adaptation can be difficult because the consequences of a change can be hard to envision and the sources for the change might be difficult to interpret but an ontology has to evolve over time in order to ensure that it remains usable (Davies et al., 2005)

Davies et al. (2005) defined two types of ontology changes, usage-driven, which is based-on changes in how the data is being used and data-driven, where changes to an ontology are done as a result of changes to the underlying data-sets. Davies et al. (2005) noted the usage-driven changes were harder to identify most likely because changes to the underlying data-sets tend to be a planned and explicit process.

An ontology can be developed in collaboration between individuals and organizations although experts and practitioners in a given domain might have a number of different needs, and view a problem from different angles. If an ontology is to be successfully developed or evolved it is necessary that the different viewpoints and needs are communicated (Devedzic, 2002)

Ontologies have both a present use and potential future uses on the internet. One of the driving forces behind the adoption of formal ontologies is the semantic web initiatives (Dogac et al., 2004), which was touched on in more detail in Section 2.5 under Semantics. Ontologies are also commonly used in search engines and e-commerce websites, in those systems the ontology helps a user to find the information he or she is looking for (Stead, 2000)

There are a number of languages that exist to help the construction of ontologies such as ontolingua (Chandrasekaran, 1999), RDF, OWL (Rizvi & Imdadi, 2008) and several others.

As seen in Section 2.3.1 and Section 2.3.2 which looked at and discussed different types of metadata, metadata can be applied to a number of different objects not simply to documents, Hartmann et al. (2005) argues this same fact and purposes that metadata can be applied to ontologies as well. It was argued by Hartmann, et al. (2005) that metadata, which was machine process-able, would improve the access to ontologies, aided in the reuse of ontologies and the maintenance of ontologies.

There already exists a number of large standardized and structures vocabularies. A lot of work has been done in the field of terminologies and vocabularies (Kuhn & Giuse, 2000), in medical domain for example there exists Systematized Nomenclature of Medicine (SNOMED) and Medical Subject Headings (MeSH) which contains an elaborate lists of medical terms ordered in a hierarchical format (Ingenerf et al., 2001; Noy & McGuinness, 2004)

From the literature previously touched on a number of uses of ontologies, primarily looking at the ability of an ontology to document and explicitly define the concepts within a domain, which in turn aids software development in a number of ways, previously listed. This section also discussed an ontologies ability to handle semantics and how metadata can be used alongside with ontologies.

2.7 Design Science Research

In order to better describe the intended research design it was clear that more information needed to be provided on the topic of DSR. This section considers some of the academic literature sourced on the topic of DSR in order to provide a more holistic view of the research and provide a more detailed description of the various concepts that appear within DSR.

This section starts off by looking at what DSR is according to the sourced literature and then focuses on the justification provided in the literature for why design can also be considered an acceptable research process. Finally important topics raised in the literature such as the nature of artefacts in the DSR process and comparing DSR to similar research methods such as Action Research is presented and discussed.

The next section considers how the academic literature perceives DSR.

2.7.1 What is Design Science Research?

Design-based research has been given a number of different names within the IT and IS academic literature over the years. Some of the previous names given include: Engineering type research in 1994, Design Science in 1995, System Development approach in 1997, Constructive type research and prototyping in 1998 (Gregor 2002), Software Design Methodology in 2003 (Hasan, 2004) although it is more commonly known as Design Science Research (Hevner et al., 2004).

This section discusses the literature on the topic of DSR and attempt to better explain what DSR is and how it is positioned within the academic literature, thus providing some clarification on how it is used within this research.

The characteristics of research within the IT and IS field which sets it apart from the numerous contributing fields, such as mathematics, philosophy, psychology, sociology, economics and management (Gregor, 2002), is that IT and IS research is orientated not towards the study of natural phenomena but the study of artefacts and artificial phenomena (March & Smith, 1995). IT and IS research specifically focuses on artefacts that are used in human-machine systems and the arising phenomena born from the interaction and usage of the artefact (Gregor, 2002). The reason for this is because IS as well as the organisations that the IS are intended to support are purposefully designed, artificial in nature and inherently complex (Hevner et al., 2004).

As previously touched on there are two major paradigms for research within the IS and IT field namely: design-based research and behavioural-based research (Hevner et al., 2004).

According to Hevner et al. (2004) behavioural-science is a theory based research method that seeks to create and validate theories which explain or predict the behaviour of humans or organizations in relation to some form of IS based solution or artefact. Behavioural-science research attempts to understand reality or a given research problem (March & Smith, 1995; March & Storey, 2008).

DSR on the other hand is a form of research in the IS and IT field which seeks to solve IT related problems by using and applying knowledge generated not only by BSR but also by kernel-theories (knowledge originating outside the IT and IS field) (Hevner et al., 2004; Niehaves, 2007).

According to Hevner et al. (2004) DSR seeks to address "wicked problems", problems which exist or arise from the complex interaction where the problem constraints and environment are ill-defined and multiple different solutions exist each of which depends heavily on human creativity and teamwork.

It was noted by Hevner et al. (2004) given these wicked problems it is inherently difficult to describe the processes and laws relevant to the solution and problem, furthermore the size and complexity of the solution space makes the problems "computationally infeasible".

One way that design-based research seeks to solve research problems is by creating new and innovative artefacts, that serve human purposes and needs (March & Smith, 1995; March & Storey, 2008) and expand the capabilities of humans and the organisation (Hevner et al., 2004). Wang and Wang (2010) further called for the innovative use of design science, and noted that standard applications in standard environments are no longer of interest to the design-science discipline.

DSR is focused on the problem being addressed and the creation of an IT or IS artefact in order to solve the problem (March & Storey, 2008). DSR is therefore inherently a problem solving process (Hevner et al., 2004), which involves the representation and presentation of a design related problem or issue and the subsequent generation and evaluation of a design-based solution (the artefacts produced in the DSR process) (March & Storey, 2008).

Essentially where behavioural science research considers how things are, DSR considers how things ought to be to attain a desired outcome (Hasan, 2004).

DSR originated from the science of the artificial and from the engineering field. Design science seeks to generate knowledge by creating and evaluating research artefacts. The research artefacts produced by DSR is conceived to be inherently knowledge containing objects, containing ideas, concepts, assumptions, practices and technical capacities. Because these artefacts are viewed as containing knowledge they can be used to effectively accomplish the processes of analysis, design, implementation and management in the IS discipline (Hevner et al., 2004).

Numerous reasons exist for undertaking the construction of an artefact such as: attaining a predefined goal, in order to address a design issue or extend the boundaries of known application of IT and IS (March & Smith, 1995; March & Storey, 2008). But just as important as the construction of artefacts is the evaluation of these artefacts because the evaluation produces the desired research findings and provides the necessary proof that the designed artefacts are able to perform their intended function or address the research problem. For this reason DSR is conducted in two complementary phases that are: build and evaluate (Hevner et al., 2004; March & Storey, 2008). These two phases iterate throughout the research process.

The evaluation of artefacts within DSR provides a better understanding of the problem the artefact seeks to address, allowing the quality of the artefact and the process used to create the artefact to be evaluated and improved (Hevner et al., 2004). Artefacts are purposely designed and created, with the intent of performing a specific purpose. These artefacts inherently contain and represent these underlying design considerations and assumptions and proof that these artefacts successfully resolve or address a given problem means that their design and assumptions is correct in some regards.

It was Hasan (2004) who referred to DSR as a "disciplined investigation" which is conducted within the development process and is aimed at improving the "thing" being developed (called the artefact by Henver et al. (2004), the developer, development process or even the research methodology.

DSR for the most part does not contain hypotheses, experimental design or data analysis, according to Hasan (2004) who also identified that this was usually a problem for reviewers of the DSR processes and findings. Instead of hypotheses, DSR strives to create artefacts that are innovative and valuable (Gregor, 2002). The development process involved in creating the different artefacts that make up a solution can be long and complex. Because of this DSR normally simplifies the research process by only looking at or decomposing a problem into smaller components which can be used to look at a relevant means, end or law that is in play. The artefact can thus be used to look at one a specific aspect of the research (Hevner et al., 2004).

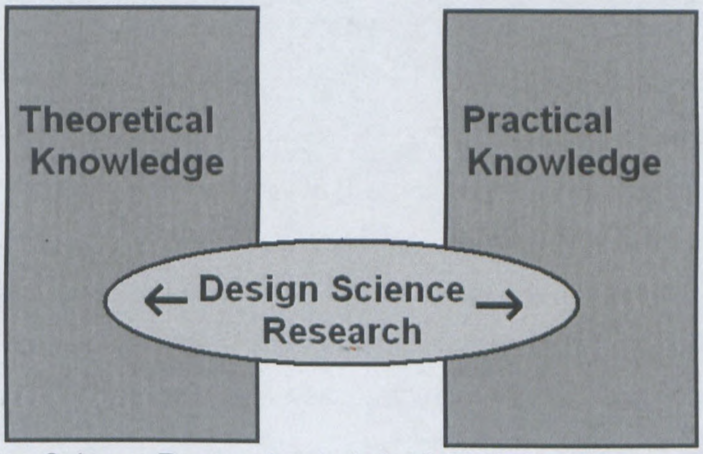


Figure 2.3: *Design Science Research role in bridging Theoretical and Practical Knowledge.*

The attempt of DSR is to bridge the gap between theories and practice (Ellis & Levy 2010). The researcher within DSR does not only produce theoretical knowledge but applies practical knowledge of a situation or a task to create an artefact (March & Smith, 1995). After applying knowledge to create the artefacts, the aim of DSR becomes to explain why the artefact (constructs, models, methods and initiations) works, or possibly why it doesn't work (Hasan, 2004) by means of evaluation, the artefacts against criteria such as value and utility (March & Smith, 1995).

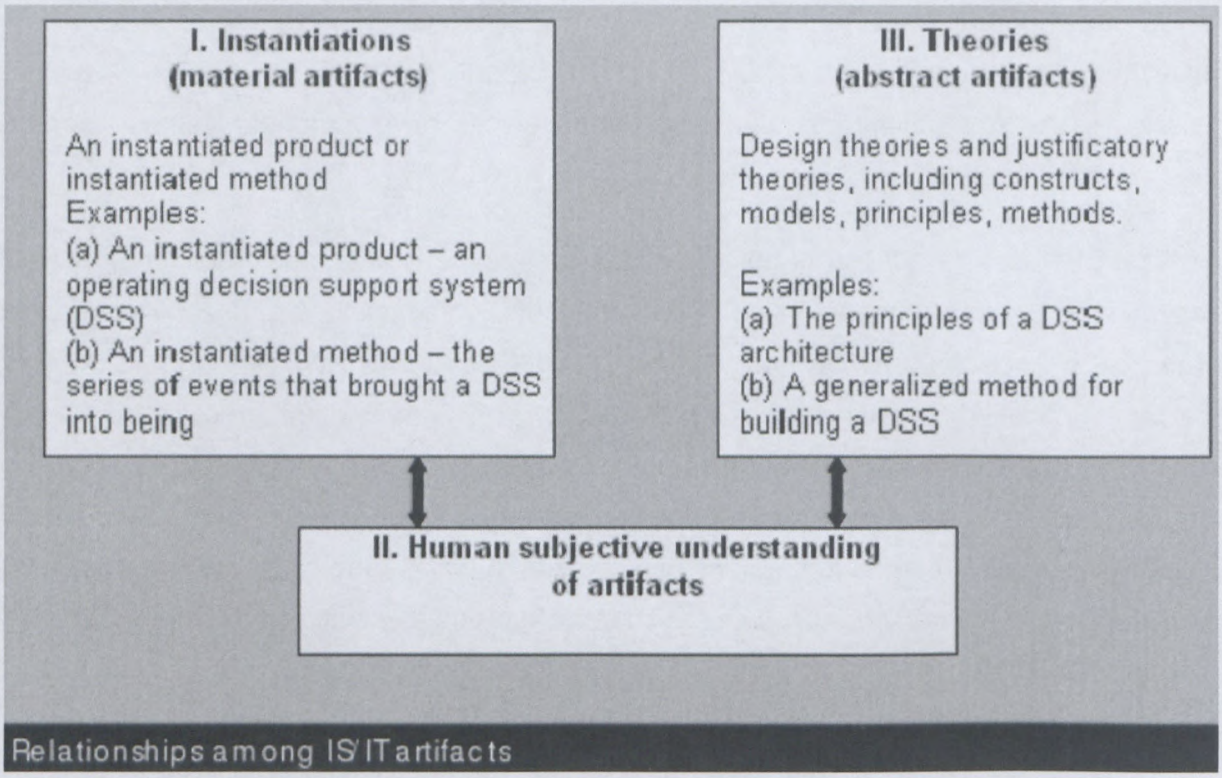


Figure 2.4: *Relationship amongst IS/IT artefacts (Gregor & Jones, 2007).*

Figure 2.4 from Gregor & Jones (2007) shows the interplay between instantiations and theories. Here the human researcher is the tool which interprets the instantiations and the theories and is the means by which information moves from between the two. The human researcher finds and interprets the theories and uses it during the instantiation and development but through the process of development the human gains new insight which can be applied to refine existing theories or to create new ones.

When a DSR approach is used for the initial research within a fairly new or unknown research area, the artefacts predominately take the form of prototypes with the goal of presenting the feasibility or addressing an identified research problem. However when DSR is conducted in a less obscured research area which is fairly well known, the researcher must identify and show that the existing IT artefacts, if they exist at all, are not adequate to address a given problem (March & Storey, 2008). After all there would be little reason to undertake research if there is nothing to learn.

This section looked at the academic literature on the topic of DSR, in order to provide a better understanding of a possibly obscure research methodology. The following section considers the arguments in the literature that supported the idea of design as a means of research.

2.7.2 Design as Research

This section considers the justification found within the literature for using design as a credible form of research.

A number of attempts have been made within the IT and IS discipline especially within the academic literature to define and justify the usage of DSR. In spite of these attempts DSR has been slow to spread within the mainstream IS and IT research community. DSR penetration into the IS and IT research community was so low that only a small minority of the papers were being published within prominent IS research journals. For the most part the majority of DSR findings and research tends to be published within engineering journals (March & Storey, 2008).

According to Ellis and Levy (2010) research involves: "addressing an acknowledged problem building upon existing literature and making an original contribution to the body of knowledge."

Research was defined by Hasan (2004) as being a: "diligent and systematic enquiry or investigation into a subject in order to discover facts or principles which are accepted or professed rules of actions. Hasan (2004) went on to define the outcome of research as being to: "adds to the body of knowledge of a discipline and it is appropriate to emphasise the strong conceptual link between research and the process of knowledge discovery and creation."

From the previous definitions it is clear that research seeks to address problems through the application of systematic process of investigation or enquiry in order to generate knowledge, knowledge which can be used to add to the current body of academic knowledge.

Hevner et al. (2004) gave the definition of design as being the: "purposeful organization of resources to accomplish a goal". Hever et al. (20004) went on to further identify that design is both a process which consists of a number of expert activities and a product/artefact produced by these expert activities.

The basic ideal underlying DSR is voiced by Archer who came from the industrial design discipline, where design-based research is a more accepted form of research than in the IS and IT discipline. Archer believed that design was a creative process and that it could be codified, captured and shared (March & Storey, 2008).

Hasan (2004) purposed that people who participate in the process of systematic analysis, design and production are in fact taking part in higher metal activities (creative and innovative thinking) which provides new insights and knowledge in relation to the problem being addressed. Hasan (2004) further stated that in such a case where knowledge is created during the development life cycle of an IT or IS solution, then the activities can be seen as being a legitimate research method.

The design and development of IT and IS solutions is itself a creative process based-on application of systematic and diligent methods. IT and IS solutions are created in order to perform a specific function, to addressing a certain need or identified problem. If the process of design and development is based in and builds on current academic knowledge and effort is made to identify the research contribution and prove an acceptable among of rigour throughout the development process, then it is possible for the process of design and development to be legitimate research.

Research in natural and behavioural-science is based-on discovery and justification; while research in DSR is based-on build and evaluate (March & Smith, 1995) with the eventual goal of DSR being the utility of knowledge generated through the build and the evaluation process (Hevner et al., 2004).

The following section looks at the literature on the artefacts produced by DSR in more detail.

2.7.3 Artefact in Design Science Research

Briefly touched on in the previous discussion is the concept of Artefact within DSR. Artefacts are created and evaluated as part of the DSR process in order to produce an acceptable research contribution. Within this section Artefacts in the DSR context are discussed in more detail.

As previously mentioned IT and IS artefacts and the organisation in which they appear are artificial and purposely created to perform some role or function (March & Smith, 1995; Gregor, 2002) and are inter-dependent on the users and the context of use in meeting specific needs or performing specific functions (Wang & Wang, 2010).

IT and IS in general can be considered more than simply things or tools to perform some activity or get some sort of a job done. Artefacts in IS and IT (and thus DSR) incorporate inherent logic, assumptions and knowledge, but these artefacts also interact with people in ways which has previously not been done (Hasan, 2004).

DSR artefacts can be broadly defined as: models, methods, constructs, instantiations and theories. Not all sources agree on the validity of theories as an acceptable type of DSR artefact. These artefacts not only help in the creation and analysis of data but also its presentation (March & Smith, 1995; Gregor 2002; March & Storey, 2008).

A construct is the vocabulary and symbols of a domain, the conceptualisation of a problem with the domain. A model is a set of statements expressing the relationship amongst constructs, the abstraction and the representations. A method is a set of steps (algorithm, guidelines or practices) used to perform a task although it's noteworthy that not all methods can constitute artefacts (Iivari, 2007); a method is based-on constructs (language) and a representation (model). An instantiation is the application of an artefact in its environment, the implemented and prototyped system (March & Smith, 1995; Hevner et al., 2004)

Although these artefacts are commonly accepted and presented within the literature a number of other artefacts also exists. Artefact can include: social innovations, new or previously unknown properties of technical/social/informational resources (March, Storey, 2008), new theories explaining the causes of a problem, new design and developments models and finally new methods and process for implementing models or tools (Ellis & Levy 2010).

According to Ellis and Levy (2010) there are three factors involved in the creation of a DSR artefact: the creation of a conceptual framework, creation of system architecture and the creation of a prototype to be tested and evaluated (Ellis & Levy 2010).

Within DSR, artefacts are perceived as embodying knowledge and one of the primary sources of research data. Gregor (2002) noted that an artefact can reflect the tools and methodologies used in its construction, while certain design theories link the construction methods with the particular type of artefact. An addition was noted by March and Smith (1995) that theorising in IT research should relate to IT artefacts within the environment that they function, based-on the characteristics which makes it unique, and therefore requiring a unique explanation.

As previously mentioned artefacts are purposefully constructed, artificial in nature and used in human-machine systems. Artefacts can be studied and constructed to solve an identified problem and these artefacts in turn can give rise to artificial phenomena which can itself be studied (March & Smith, 1995).

A requirement for DSR artefacts to be able to provide credible research data, the utility of the artefacts must be demonstrated, the artefact must be proven to address an identified issue in some form (Hevner et al., 2004). Artefacts contain assumptions, design decisions and logic. If the artefact successfully addresses a given problem, then the underlying knowledge can be assumed to be credible. The evaluation helps to show the utility of the artefact and aid in explaining how the artefact addresses the identified problem as well as identifying the underlying knowledge contribution.

Both the artefact and the phenomenon are artificial but neither can ignore the natural laws (Hevner et al., 2004). Natural and behavioural-sciences can provide a meaningful contribution to the design-science field by providing knowledge and information relating to the natural world in which the artefact must function (March & Smith, 1995).

This leads to a situation where Design-Science can exploit knowledge provided by natural and behavioural science to produce artefacts which in turn give rise to phenomena which can be studied by behavioural science (March & Smith, 1995).

DSR artefacts are rarely full-grown IS solutions used in practice (Hevner et al., 2004). This might be because of the wicked nature of the issues DSR seeks to address. If the problem domain is ill conceived or ill-defined then the DSR artefact might be little more than a proof of concept, never intended to be used in practice. Another possibility might be that most commercial IS projects have a monetary component to it and no research component.

Artefact are technical in nature but are not independent of the social context and artefacts are interdependent and coequal to the people and the organization in which they are used, and both need to work together to meet the desired outcomes (Hevner et al., 2004; Wang & Wang, 2010). However Hevner et al. (2004) argued that people and the organisation are no more important than the artefacts themselves and that both are equally crucial.

Artefact however are not static, they can evolve through human action or because of new technology that has been developed. (Gregor, 2002)

2.8 Literature Review Conclusion

Initially the literature review focused on the academic literature that related to the primary context of the research, HBHC. In order to provide a more holistic view on the topic of HBHC the literature touched on HBHC in both the developed and developing world but with focused placed on highlighting the challenges faced by developing countries to provide healthcare services and the ability of HBHC to aid in the delivery of quality care services. The literature showed that HBHC can aid in the delivery of care services. Despite many of the issues faced by developing countries, HBHC itself had to face a number of challenges.

After contextualising the overarching research context the literature review looked at the topic of HI, and the various advantages it brings to healthcare and ability of HI to help overcome some of the issues faced by rural HBHC. The reasons why HI became more prominent in healthcare was discussed but also the issues involved with the implementation of an HI solution. One of the primary requirements for achieving many of the promised advantages of HI was interoperability, but interoperability itself was also a major issue within the healthcare field, with decades of work and research having gone into solving the issue.

The literature thus also looked at some of the HI standards that attempt to overcome in the interoperability challenge. It was notable that these standards such as the HL7 proposed means for defining the meaning (semantics) and structuring the data used in the healthcare domain, leading to the conclusion that the interoperability challenges were as a result of differing data structures and an inability of heterogeneous systems to covey the adequate level of semantics when sharing information.

The literature review thus focused less on the issues of interoperability but more on the underlying issues that cause the interoperability challenges, and looked at the topics of semantics and ontologies.

The literature review also contained a reasonable amount of technological due-diligence, and looked at technologies that could be used to address some of the underlying issues of interoperability. Specifically the technological due diligence of the literature looked at the topics of metadata and repositories.

The topic of metadata was covered in some detail, since it became clear from the literature that it was a fairly wide field with numerous interpretation and possible misinterpretation.

The literature review thus provided a good insight into the context of the research, the underlying issues that the research attempts to address as well as technologies which could be useful in aiding the research and possibly lessening some of the issues identified in the literature relating to interpretation of data semantics and structures.

Chapter 3 Research Design

This chapter considers the specifications of the research, detailing the various factors that led to the research undertaken, the research process and how the research produced viable research findings.

The first section in this chapter looks at the research problem that gave rise to the need for the research. During the discussion of the research problem special attention is paid to identifying the underlying issues which gave cause to the overall problem of this research. This is important because the research is undertaken in a research area which is characterised by its complexity and with little pre-existing research. Detailing the research area and research problem is important because the research attempts to use DSR which is not widely used or reported in the academic literature.

The second section lists and discusses the research question and sub-questions that arose from the research problem as well as the objectives that the research attempts to achieve. This section relies heavily on the preceding section, the discussion of the research problem and area, as the research question originates from these areas and directly attempts to address a given aspect of the identified research area / problem.

The third section considers the research epistemology and ontology. Before the research methodology can be discussed it is important to understand the epistemological and ontological stances adopted for this research about how this research identifies “the truth” and the view of “the world”. It is important to define these concepts in terms of the research as it directly guides how the research questions is answered and also influences the methodology that is implemented in order to answer these research questions.

The fourth section of this chapter considers the research methodology used to answer the research questions within the adopted epistemology and ontology. The research follows an amalgamated DSR methodology, based-on several of the more prominent design science research methodologies detailed in the academic literature. In order to provide the necessary insight into the research area and research problem (essentially the kernel theory required to perform the DSR) an ethnographic study was undertaken which is also discussed under the methodology section.

In order to better understand the intended DSR approach the academic literature on the topic of DSR is discussed. Because DSR as a means of research is not heavily prevalent in the current body of academic literature it was deemed important to provide sufficient discussion of DSR and several relevant topics to better contextualise, not only the term itself, but also the literature as these relate to this research study. The academic literature on DSR was deemed more relevant to the research process and was thus included within this chapter and not within Chapter 2 (the literature review). The section on DSR attempts to describe design and development based research in more detail and touches on several of the more prominent topics relating to DSR found within the literature.

After the research methodology is discussed the next section focuses on the research justifications. The justification section looks at the reasons why the research effort was undertaken in the first place. The justification section goes on to detail the reasons why a DSR approach was chosen to conduct the research and its relevance to addressing the research problem. Finally the it details the reasons why a derived methodology from a combination of the various DSR approaches presented in the academic literature was used instead of simply selecting the most applicable of the DSR methodologies.

Rigour and Relevance within the research are considered and discussed as well as the context/environment in which the research was conducted. A delineation of the research is presented as well as the assumption for the research study.

The research is being conducted within a subset of healthcare and the ethical considerations of the research are discussed in a following section.

The next section considers the research problem.

3.1 Statement of Research Problem

This section discusses the research problem that gave rise to the need for the research and in order to do so this section also touches on some aspects of the research area. This section will draw on some of the literature findings detailed in the Chapter 2.

South African NGOs and other similar organisations that attempt to deliver HBHC services to rural and under-resourced communities have a tendency to utilise common data-elements differently within their unique contexts. As detailed within Chapter 3 in Section 2.1, HBHC in under-resourced communities is responsible for the delivery of care at the homes of the patients. HBHC attempts to overcome many of the issues faced by these communities which prevent the provisioning of formal healthcare. These HBHC initiatives face numerous issues of their own, which was detailed in Section 2.1.2. Since no two communities are exactly the same the HBHC initiatives that cater for these communities are just as diverse in not only their organisational structure, or types of healthcare services they provide but also in their data requirements, gathering, storage and utilisations.

These differences in the understanding and utilisation of data-elements appear in both the HBHC initiatives paper-based and IT based-system in the few occasions within the rural South African HBHC context that IT based solutions are actually available.

HBHC initiatives have a tendency to utilise predominately paper-based systems which are not intrinsically difficult to use or understand and simply implementing a paper-based systems does not necessarily lead to issues of data inconsistencies. However because the paper documents are difficult to interpret and analyse, especially by outside parties such as IT solution developers and policy makers, it makes it difficult for external parties to create relevant solutions to data-related issues being faced by HBHC initiatives. For example, what constitutes a name can be difficult to fully discern when it has many representations and functions that are spread across different papers. The same issue occurs for a patient record, since the record is for the most part spread across a number of different paper documents and the same data is potentially used and represented differently based-on the goal of the specific paper document, i.e., patient diagnoses documents, patient care documents, patient detail sheets etc. The predominate usage of the paper-based systems within HBHC thus obfuscate these underlying issues of data inconsistencies because these inconsistencies are spread across a number of different documents and can only easily be identified if all these documents are collectively studied.

In the rare occasion that an IT based system is being utilised in HBHC the data and knowledge collected and managed by HBHC initiatives are stored in heterogeneous systems, using different formats and within different data structures thus leading to further issues of data-element inconsistencies. Some HBHC initiatives are governed by different controlling bodies, usually the funding organisations, while other initiatives form part of national group. In South Africa HBHC falls under the National Department of Health if they provide a formal home care health service. However, in practice there are many other forms of care provision that are provided to persons at their homes and in some cases these care activities are outside the scope of healthcare activities, e.g. cleaning houses, cooking food etc. In the case where care services relate to poverty and other socio-economic conditions the Department of Social Development also becomes involved. All these bodies can dictate which systems the initiative must use and since these governing bodies are not necessarily in direct communication or collaboration. It is possible then for each governing bodies to require that the HBHC initiative implement new systems which differs largely from any other systems that are currently in place.

Furthermore because some of these controlling bodies have authority over a number of HBHC initiatives, all of whom are to some extent unique, any solution they implement might not take these unique characteristics into account. A HBHC initiative which has its own set of data requirements necessary for it to function properly, might then decide to not fully utilise the system mandated by the governing body but rather a more relevant IT or paper-based system.

This fundamental misunderstanding of the meaning and usage of the data-elements in HBHC further leads to a number of issues amongst which includes: the inability to implement more advanced IT based systems; issues when attempting to implement interoperability amongst these systems and an inability to effectively analyse the data and knowledge usage of the many different HBHC initiatives.

Without being able to fully understand the definition and usage of data and knowledge across the field of HBHC most people who attempt to improve the state of HBHC as a whole by implementing some form of IT-based solution will most likely fail to achieve make a notable improvement. This is because any system that is being implemented will most likely be based on misassumptions about HBHC such as incorrect assumptions about data definitions and usages. It can then be easily assumed the implementation of non-relevant solutions will continue to appear within HBHC because of the lack of understanding about what is really going on at the ground level of HBHC.

HI holds a number of advantages and can aid HBHC in a number of ways as detailed in Chapter 2 within Section 2.2.1 under the heading of Advantages of Health Informatics. HI solutions could greatly aid HBHC and thus indirectly benefit those who rely on HBHC for their care and treatment needs. But without a fundamental understanding of the HBHC and the usage of data these advantages can be difficult to achieve.

Very few IT based solutions are currently implemented or widely used within the rural South African HBHC context. Not only are tools used to understand the utilisation of data-elements is none-existent within the research context but even rudimentary HI systems are not implemented or if they are not effectively being utilised. The literature pointed to the need for better and more relevant overall design of solutions and also for the creation of systems that meet the needs of the users and which are also perceived to be effective and easy to use by the relevant stakeholders.

A lot of research has been done on localising successful foreign solutions and on utilising specific technologies, such as mobile devices, to overcome specific problems. Little literature however was found however on analysing and designing solutions intended to function within the South African HBHC context.

Without more research being done and more importantly disseminated on the process of designing for the rural South African context, any new projects possibly face a number of unknowns which might negatively impact the overall successful of these projects. It is argued that design plays an important part in ensuring the success of a new IT-based solution and by extent poor understanding of design considerations and the design process can play a large role in the failure of a given solution (Purgathofer, 2006).

3.2 Research Question

What is required to design and develop an appropriate repository for the HBHC context to adequately store and represent care data-elements at an ontological level?

3.3 Research Sub-Questions

Table 3.1: *Research Sub Questions.*

Sub-questions	Research method(s)	Objectives
What are the contextual implications for designing a repository of HBHC data-elements?	Literature review. Ethnographic Study of the HBHC context.	To gain additional insight into the rural HBHC context of South Africa. To gain the necessary design requirements to successfully create a system that can be used in the South African home-based care context. To gain an understanding of how the HBHC characteristics impact design and development.
How are the care data-elements and the relationships between them appropriately represented	Design and Development based research. Literature review. Ethnographic Study of the HBHC context.	To determine an appropriate means representing care data-elements. To gain an understanding of the types of care data-elements and how they relate to each other.
How can the appropriate design considerations for a semantic metadata repository be identified within a research process.	Literature review. Design and Development based research.	To understand the best way to use the process of design and development as a means of research. To understand how, if at all, design knowledge and innovative thinking can be captured.

3.4 Objectives

The aim of the research was to first explore how data-elements are used at facility level within the HBHC context in South Africa. Once the different uses of data-elements were established a semantic metadata repository was designed and developed to represent the care data-elements used in HBHC at an ontology level. The research thus attempts to create viable research results that are applicable to similar development or research undertaking within the same or a very similar context.

Overall the research also seeks to gain a deeper understanding of HBHC specifically within the under-resourced and underdeveloped South African communities that the research focuses on, but also potentially providing a basis from which a deeper understanding can be gained into similar contexts. The research attempts to primarily understand the data usage and data flows within the research context.

DSR is purposed by the academic literature as a means of using IS design and development as a means of conducting research. This research uses and adapted DSR methodology which combines many of the common elements of several DSR methodologies found in the literature. An additional objective then is to gain an insight into the effectiveness of such a research approach which combines design, development and research. Furthermore the research also seeks to understand such a methodologies ability to represent design and development insights / knowledge. Yet another objective considers how a DSR methodology can be utilised in research by providing a practical example of DSR. Collective it is the intend of the research to add to the body of DSR knowledge

Following a DSR approach a by-product of the research is the creation of the first iteration (prototype) of a semantic structural-metadata repository which could potentially be used within the research context. The Repository is intended to be used to gain an understanding of the various data-elements used within the diverse HBHC contexts found in South Africa. But the repository is only viewed as a possible first step in solving the issue of inconsistent data usage in rural HBHC of South Africa. It is hoped that by using the repository these data-elements can be effectively captured and represented, leading to a better understanding of the HBHC data-elements. This understanding is intended to offer a basis from which further development of relevant and successful systems and data standards can be undertaken. As such the research does not attempt to define and present these HBHC data-elements in any detail but uses only a handful of data-elements to present the usage of the repository.

By producing IT-based artefacts and going through the process of development from design to initial prototype, the research also attempts to detail some of the steps and activities that would go into the development of an IT-based solution intended to be within a context similar to that of the research.

3.5 Scope of Study

This section details the scope of the research and attempts to provide a clear understanding of what the research focuses on and what the research process entails. This section details the scope of the research in terms of which communities the research focuses on and which parts of the IS design and development process the research focuses on.

Limiting the scope of the DSR undertaking is common practice, because the design and development aspects of DSR can be a complex activity involving numerous individuals and potentially involving a number of complex topics. For this reason the research usually only focuses on a specific subset of the overall process / problem / solution (Hevner et al., 2004).

The research looks at the design considerations for a semantic metadata repository for a HBHC initiative which caters for the healthcare needs of under-resourced and underdeveloped rural communities in South Africa. The research is thus primarily focused on the rural South African HBHC context, with very little focus placed on formalised context such as hospitals, clinics or other formalised healthcare institutions. The research also looks at both the Western Cape and Eastern Cape provinces of South Africa but the majority of the research data is produced from the Stellenbosch Hospice which caters for the Kayamandi Community outside Stellenbosch in the Western Cape province of South Africa.

The development effort undertaken as part of the research follows an iterative agile development approach but the research itself only focuses on the first iteration of development, looking mostly at the analysis and fundamental design and development activities necessary to create a semantic metadata repository. The human computer interaction, aesthetic and usability elements of the solution are touched on by the research. However because the research only looks at the initial iterations the UI and HCI elements are not refined and developed enough to contain sufficient depth to produce data containing sufficient rigour and relevance to constitute a contribution to academic body of knowledge. As such the design considerations purposed by the research do not stretch to encompass these human-factor components and only lightly touches on these were most applicable.

Because the research only focusing on the first iteration of development of the semantic metadata repository the research is focused mostly on technical design requirements and not on the usability and user interface design requirements.

The research also focuses on the development of a system to be used at an institutional level, with the necessary infrastructure in place (computers, electricity etc.). The research however touches on the overall HBHC context.

3.6 Epistemology and Ontology

3.6.1 The research approach

The research was conducted using a qualitative-pragmatic approach following a design science research approach.

The DSR approach was chosen because it was one of the primary means found in the academic literature for conducting "practical" research in IS, basically DSR offered a means of combining design and development. Development is conventionally associated with practical / industry component of IS and IT and rarely is it associated with research directly. Using both design and development within the research was seen as the most effective means of addressing the research problem as well as identifying and understanding what is required to design and develop an appropriate repository for the HBHC context to adequately store and represent care data-elements at an ontological level.

The literature supports the usage of a pragmatic approach, which is advantageous for the research as DSR is intended to apply theoretical knowledge in order to achieve some goal or objective (which is not necessarily the research objective or goal). If a pragmatic approach was not followed then it would be difficult for the researcher to say with any certainty that the correct theories were being applied.

The research ontology is based-on a nominalism stance from a subjective viewpoint where it is assumed that knowledge is constructed using social constructs such as language etc. This chosen ontology allows for knowledge / understanding and innovation created during the design and development processes to be narratively captured and presented.

A DSR approach is based-on the concepts of design, construction and evaluation. More general information on the topic of DSR is provided in a following section as literature on the topic is presented and discussed. The concepts of design, construction and evaluation within DSR bare a strong resemblance to grounded research approaches such as pragmatism, which is used in most constructive and developmental research methodologies used in the creation of systems in practice (Hasan, 2004).

Thus the overall epistemology to a lesser extent and the evaluation to a greater extent have an interpretative influence.

The following section focuses on the chosen epistemology and ontological stance used during this research. This section attempts to briefly provide clear arguments and justifications for the choices made as part of the research, specifically the choices made in regards to the ontology and epistemology.

3.6.2 Research Paradigm

The word epistemology relates to the discussion and debate around how a human can come to know or achieve what is considered "true knowledge". Epistemology specifically attempts to clarify the relations (the process) between the object or source of knowledge (the thing being examined or evaluated) and the derived knowledge (product) (Niehayes, 2007). For this reason epistemologies are an important aspect of research because it defines in part how knowledge is to be gained from the object that the research is focused on. Further an epistemology also guides how this knowledge is to be verified in order for it to constitute "true knowledge" in the academic sense.

As touched on in Section 2.6, under the heading of Ontologies, an ontology in philosophy denotes the study of the nature of items in existence (Noy & McGuinness, 2004). Ontologies form the basis by which epistemological discussions about the nature of the "real world" can take place. These epistemological discussions tend to focus on the nature of the objects from which knowledge and the "truth" can be derived (Niehaves, 2007).

A research paradigm pulls together epistemologies and ontologies; it's a worldview based-on a specific ontological and epistemological stance (Niehaves, 2007).

Hevner et al. (2004) attempted to argue for DSR as being an additional epistemological stance along with those of positivism and interpretivism. Although DSR has found some support within the academic community there is still some debate as to whether DSR is in fact a third major paradigm for research. Andriessen (2006) for example saw DSR as neither paradigm nor a methodology. According to Iivari (2007) it is also required that DSR be based-on a sound ontology thus for the purpose of this research DSR is considered to be a research paradigm (encompassing a number of potential epistemology and ontology combinations) for which a number of guidelines and methodologies exists within the academic literature. In this research however a more hesitant stance in relation to DSR is taken, DSR is not seen as being an outright new epistemology but this research does not go so far as to dismiss the value of DSR as a research methodology..

As all research undertakings attempt to produce some form of knowledge with "truth" being the measure of the quality of this knowledge, after all what use is research if it results in knowledge which is misrepresentative of the truth. Furthermore as an epistemology defines how "truth" can be known it obviously follows that explicitly defining and detailing the epistemology is of use and importance when conducting research.

The methodology implemented during this research undertaken is based within a Design Science-based approach. Bertelsen (2000) however noted that epistemologies do not explicitly inform us of how to conduct the process of producing design orientated knowledge, which is obvious of interest to DSR. Niehaves (2007) went further and noted that although work has been done to explicitly theorise an epistemology for DSR, no widely accept stance has yet been taken. Clearly some work has been done in terms of defining an epistemology for DSR but no consensus exists as to what is and is not an acceptable epistemological stance for DSR, leaving room open for debate.

Before a given epistemology can be chosen first the topic of research epistemologies needs to be discussed. Some differences do exist as to the exact definition of certain epistemological stances. Niehaves (2007) defined positivism as an epistemological stance that has as one of its fundamental tenants that a real objective world exists and that it is possible to know of the world and the items in it as something separate from the researcher. Furthermore Niehaves (2007) defined interpretivism as a epistemological stance that also assumed the existence of a real world but did not assume that someone could achieve an objective knowledge. All knowledge was thus in part influenced by the researcher. Even though a number of other definitions exist, it is clear enough that pure interpretivism and pure positivism lie at polar opposites of each other and that most research falls somewhere in-between.

According to Niehayes (2007) the academic literature on the subject of DSR (specifically on the topic of evaluation) is mostly dominated by a positivist epistemology where the proof of "true knowledge" (or more exactly the evaluation process which provides evidence of this proof) is done through mathematical formalisms or experimentation. However Niehayes (2007) made the argument for an interpretive epistemology and even demonstrated the application of an interpretive epistemology using the requirements for interpretivist research defined by Klein and Myers in 1999 on the DSR guidelines purposed by Hevner et al. (2004).

3.6.3 Pragmatism and Design Science

A number of other sources also refer to the pragmatic stance and its application in DSR (Iivari, 2007). This most likely is due in part because of DSR desire to bridge practice and theory (discussed in more detail later in this section) by practically implementing theories and creating theories from practical implementations. Pragmatism is the philosophy that states that truth is that which works when used in practice (Hasan, 2004; March & Smith, 1995), further stating that truth and utility (effective and efficient artefacts) are closely related concepts (Hevner et al., 2004), essentially that if something works or provides the needed answers then it also contains some truth.

According to Hevner et al. (2004) DSR is one of two research paradigms in the field of IT and IS along with Behaviour Science Research (BSR), Niehaves (2007) however saw BSR and DSR as simply two complementary perspectives in IS research rather than being two paradigms. However both agreed that BSR and DSR are distinguished by BSR being knowledge-producing or problem understanding while DSR is knowledge-using or problem solving (Hevner et al., 2004; Niehaves, 2007).

This knowledge using and producing views towards DSR and BSR are closely related to the previous mentioned bridging of theory and practice. A question does then arise about the application of DSR in a research area where very little work has previously been done (little contextual knowledge exists that can be reapplied to some form of research). In such a situation a researcher can attempt to apply knowledge from a different research area but the relevance is questionable. A second option of course is for the researcher as part of the DSR to learn about the research area first, even if only at a basic level, in order for this baseline knowledge to feedback into the research and ensure that any knowledge / theory that are applied as part of the research are in fact relevant and apt.

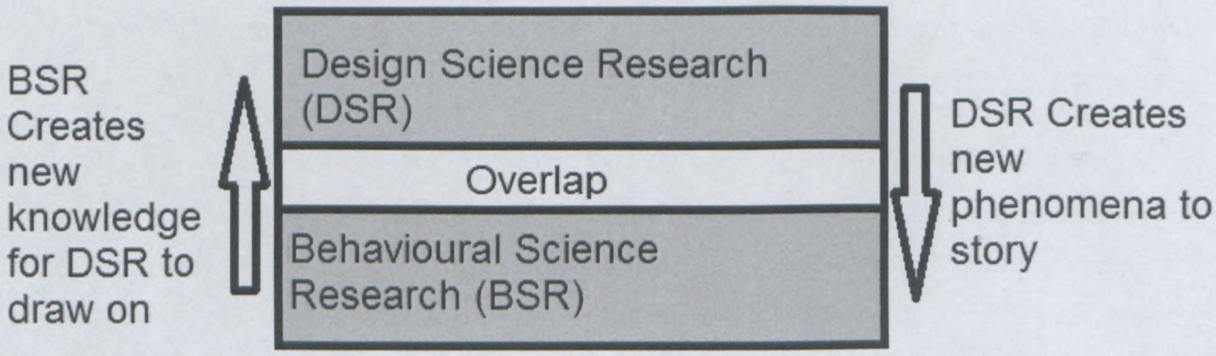


Figure 3.1: *Interplay between Design and Behavioural Science Research.*

From the literature, specifically that of Niehaves (2007), this research views DSR and BSR as complementary types of research conducted within the IT and IS field. Furthermore because BSR can produce knowledge by studying artefacts and the context in which these artefacts are used and DSR can consume and apply knowledge to create artefacts. It is further assumed that there does well exist an overlap between the two, that DSR can pose aspects of BSR and BSR aspects of DSR.

This overlap is also supported by the academic literature, when discussion design methods Purgathofer (2006) discussed analysis and synthesis as being a fundamental concept of the design process, essentially stating that before an attempt can be made to undertake solution development the problem first needs to be understood or analysed. Furthermore understanding the problem at hand is crucial if any attempt is to be made at collecting any sort of solution requirements (Purgathofer, 2006).

The above seems to cement the importance of understanding the problem and further Ellis and Levy (2010) called for DSR to have strong literature groundings especially the design and construction aspects of the research. However research context would also then exist where the academic literature is insufficient or outright missing and thus cannot provide the literature grounding necessary for requirements to be defined and the solution construction to be undertaken. In such a research context it is necessary to first use BSR methods to create this necessary level of understanding and insight. As has been shown previously there is a clear overlap between DSR and BSR, it is thus argued that it is acceptable for DSR to use research methods that are focused towards "understanding" in order to apply research methods focused towards "solving". In this research which faces a similar issue, that of limited pre-existing academic knowledge relating to the rural South African HBHC context, this "understanding to solve" takes the form of an initial ethnographic study undertaken in the rural South African HBHC context.

An example of "meaning" and "sense" in terms of design artefacts was provided by Bertelsen (2000) as: "The meaning of a hammer is the assemblage of shaped iron and wood. The sense of the hammer is that it is used by a carpenter for driving nails."

3.6.4 Design and Development

Because of the design and construction orientated nature of DSR and the contradiction between meaning and sense in regards to design artefacts, Bertelsen (2000) stated that design-orientated epistemology (or some form of DSR epistemology) should attempt to explicitly address this contradiction between representation and construction (the meaning and the sense of artefacts). The reason for this was that such an epistemology would be able to bridge the theory and practice gap and would allow for the transition between the interpretation (the source knowledge or knowledge originating in the analysis of the problem) and the eventual construction of the artefacts and instantiation (the running software).

Furthermore Bertelsen (2000) suggested that an epistemology for design-based research would also be required to support the creation of knowledge based-on the social aspects of a given technical issue, such as the result of implementing new technology. This however seems somewhat at odds when taking the BSR's "knowledge producing / problem understanding" and DSR's "knowledge using / problem solving" nature into account.

However this too can be resolved by acknowledging that it serves to strengthen the existence of an overlap between BSR and DSR.

Because pragmatism states that what works in practice is truth, this helps to bridge the gap between representation (sense) and construction (meaning). If a specific construction is able to solve the problem to achieve a given sense, then the construction has a measure of truth and success.

The rigorous construction of DSR artefacts is, according to Iivari (2007); one of the key characteristics which distinguishes design-based research from simple practical applications within the IT and IS fields such as software construction. Because of the importance which Iivari (2007) placed on the rigorous artefact construction, he also called for the process of construction to be transparent, to show where the various DSR artefacts originate from. It is thus important for the construction process, and the design process as well, to be clearly detailed and captured implying that the type of data gathered to be qualitative (textual, narrative etc.) as quantitative (numbers, statistics etc.) data is ill fitted to meet this "transparency" requirement.

Nominalist ontology states that knowledge can be constructed using social constructs like language; this ontology fits perfectly with the requirement for transparency and qualitatively capturing the design and construction process.

According to Iivari (2007) there are two opinions of how DSR and practical applications of IT and IS can be differentiated. Firstly that the essence of DSR lies in the application of scientific evaluation methods of artefacts, this focuses DSR not on the construction of new artefacts but on the evaluation of current existing artefacts. The second opinion is that the essence of DSR lies in the application of sufficiently rigorous construction methods for the creation of the various artefacts; this allows DSR to be applicable to the construction of new artefacts.

This research leans towards identifying DSR in terms of the latter opinion, the application of sufficiently rigorous construction methods and the application of existing knowledge to create new innovations in the IT and IS field, but this research also identifies the need and importance of implementing scientific evaluation methods.

A pragmatic construction approach is undertaken in the research. Pragmatism state that what works is "truth" and further pragmatism prioritises practical truth over objective truth. The pragmatism construction approach along with a nominalist ontology used in the research produces a number of textual and graphic representations of the design and development process and an interpretative evaluation was thus chosen as it was seen as the best way of interpreting these methods.

With this section providing the research epistemology and ontology (the research paradigm), the following section details the research methodology undertaken as part of the research.

3.7 Methodology

This section presents the research methodology employed during the research process. This section considers and discusses several DSR methodologies found in the academic literature, these methodologies are analysed and combined to create a DSR methodology that combine the most relevant aspects of the various discussed DSR methodologies.

From the research problem detailed previously it was clear the solution was fairly complex and contextual in nature. The detailed literature analysis detailed in Chapter 2 provided a large amount of useful and insightful knowledge relating to HBHC and to the developing world but few sources were of direct relevance to the research area, meaning that although the literature provided good insights into the research area there were still many uncertainties.

In order to supplement the literature analysis and provide additional information for the design-science component of the research, an initial exploratory ethnographic study was conducted in order to better contextualise the research and align the research findings with the academic literature. The ethnographic study did not form the primary means of data collection and analysis but it still formed an important contribution of the overall research.

Ethnographic research and methods are being used increasingly outside of the field of anthropology (Bharwani, 2006). Research methods such as an ethnographic study is usually not associated with DSR, which seeks to solve problems by utilising knowledge and creating design artefacts, but is more closely associated with BSR which seeks to gain insight and understand the problem since an ethnographic study seeks to understand human and human, human and institution, human and machine interactions (Bharwani, 2006). As has been discussed in Section 3.6 and visualised in Figure 3.1 it is possible for DSR to make use of BSR, as an overlap does exist between the two IT research paradigms.

Within this research the ethnographic study forms part of the kernel theory of the DSR helping to guide the design and development component of the research. Kuechler and Vaishnavi (2008) defined kernel theories as being behavioural / natural science theories originating outside of the IS discipline (much as ethnography originates within the anthropology field (Bharwani, 2006)) which has the potential to provide novel means of addressing IS design issues. Furthermore Kuechler and Vaishnavi (2008) also recognised the interplay between BSR and DSR and they saw kernel theories as being able to inform the DSR efforts.

The ethnographic study is but a small part in the larger research. The larger part of the ethnographic study is conducted primarily at the start of the research and where necessary during the design and development process, further ethnographic activities are undertaken to provide the necessary insights to inform and guide these efforts.

The remainder of this section details the specifics of the DSR methodology implemented during the research.

During the literature analysis several methodologies and approaches for conducting DSR was discovered. These methodologies ranged in depth and complexity. However none of these methodologies seemed to fit with the characteristics of the research characteristics including: “understand to solve”-nature of the research as detailed in the Section 3.6, the development of a technical prototype solution and not a commercial one, the use of an ethnographic study. These methodologies did provide a basis and guidelines for constructing a more adequate DSR methodology. However before this chosen methodology can be presented it is necessary to analyse the various methodologies found within the literature and to detail the process of creating the chosen methodology.

The requirements for general academic research was listed by Ellis and Levy (2010) and includes the following: that the research needs to be driven by a problem appropriate to the type of research, that the research needs to be based around questions that can be answered by the type of research being conducted, the assumptions, limitations and delimitations of the research must be acknowledged, that the results of the research can only be produced by the applied methods and finally that the conclusions must be based-on the results produced.

These six requirements for generic research above provide good criteria and guidelines for what can be considered to be a good generic research methodology and thus can serve to ensure that the intended created DSR methodology meets the requirements for not only the current research but for general research.

Ellis and Levy (2010) also provided a methodology for conducting DSR.

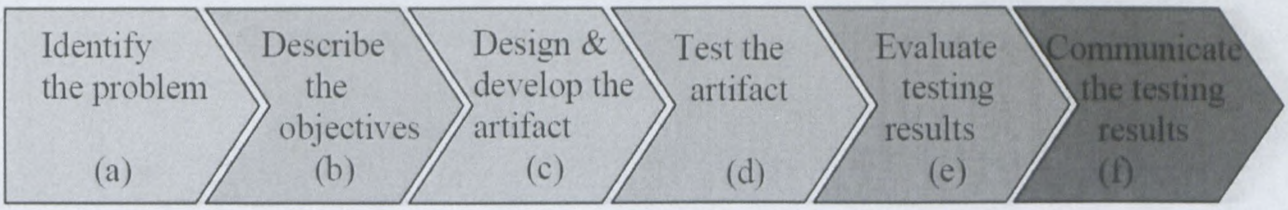


Figure 3.2: *Ellis and Levy (2010)'s six phased design and development research approach.*

While discussing the first phase "identifying the problem" of their DSR approach, Ellis and Levy (2010) made reference to Hevner et al.'s (2004) factors that underlie all design-based research namely: environmental factors, inherent complexity of problems and solutions; the capricious nature of possible solutions; solutions that were partly dependent on the human creativity and dependence on collaborative efforts to solve the problems.

Within the design and development phase Ellis and Levy (2010) purposed three important factors, namely: constructing a conceptual framework, creating the architecture and creation of a prototype. Conceptual framework includes the functionality and requirements, and serves to connect the research problem with the intended solution artefact. Creating the architecture involves analysing the alternatives and the rational in a given architecture. The prototype is created in order for evaluation and to ensure the prototype is applicable in the intended context.

Three essential considerations during the test and evaluation phases were mentioned by Ellis and Levy (2010): identifying the developed artefacts ability and inability to meet the intended functionalities and requirements; evaluation must use processes supported by the literature and finally evaluation must ensure the value of the artefacts.

Most of the phases of Ellis and Levy's (2010) purposed approach to DSR is in line with their own guidelines for general academic research. Their proposal for an approach to DSR is then seemingly just a specification of a more general approach to research. The approach also seems to neglect explicitly mentioning the need for grounding the research in existing knowledge or reviewing the current academic or practical knowledge in order to reapply it in the research.

The DSR methodology purposed by Ellis and levy (2010) is one of the newest and seemingly most complete research approaches, although possibly leaning more towards conventional general research than towards DSR. Ellis and Levy (2010) make reference to a number of older approaches but do not seem to directly refine any of previous approaches. The methodology is also fairly clear but Ellis and Levy (2010) focus on conveying DSR to the novice programmer thus focus on fairly low level concepts and many aspects relating to general research. The article unfortunately does not provide too many practical examples of the methodology in use.

Table 3.2: *Hasan (2004) four stages of Design Science Research.*

Stage:	**Description:**
1. concept design	This phase involves an in-depth literature review and interview and communication with knowledgeable practitioners. In this phase the researcher finds existing knowledge in order to apply it to the problem area, making adapting and amalgamating the existing knowledge were needed.
2. constructing the architecture of the system	This is the new knowledge creation phase, where researcher defines components models, algorithms and data structures using the previously designed concept.
3. prototype	Proof of concept phase, single prototype or evolving prototype. In this phase the prototype provides new insights about the problem it addresses and the about the system.
4. product development	Formalise system specification to create and test a robust system. In this phase a product is made for a sponsor or for commercialization.

Within Hasan's (2004) stages for DSR there is a strong call for the research to be anchored in existing knowledge, originating both inside and outside of the academic body of knowledge, and there are well-defined phases in which knowledge can be created. No evaluation phase is defined as the third stage is intended to be the proof of concept and proof of the validity of the overall research.

The methodology purposed by Hasan (2004), seems to be more specifically related to the development of actual commercial products and as such the purposed framework bares a strong resemblance to a software development methodology much more so than it resembles a research methodology (when compared it to the general research requirements purposed by Ellis and Levy (2010). Hasan's (2004) work was a notable attempt to marry research and software development, but did not seem sufficiently developed as a research methodology. The article by Hasan (2004) however does provide a good practical example of his methodology being implemented.

Table 3.3: *Hevner et al. (2004) seven Guidelines for DSR.*

Guideline:	**Description:**
1. Design as an artefact	The research must produce viable artefacts such as: constructs, models, methods and instantiations
2. Problem relevance	The research must solve problems using technology based solutions. It must thus be possible to address the research problem using the design, development and implementation of adequate technology solutions.
3. Design evaluation	Using established evaluation methods the efficiency, efficacy and quality of a designed artefact must be demonstrated, to ensure valid academic knowledge can be derived.
4. research contribution	The overall research must be shown to contribute in the areas of the design artefact, the design foundation or the design methodology being utilised.
5. research rigour	Rigour is required in both design and evaluation of artefacts.
6. design as a search process	Desired artefact which satisfied the requirements of the environment and addresses the research problem is developed using available means and knowledge, which the researcher identifies using the current available knowledge.
7. communication of research	Research results need to be communicated to technical and non-technical audiences.

The guidelines provided by Hevner et al. (2004), as shown in Table 3.2, are fairly detailed and because of this it is conceived as being fairly easy to implement and use, if somewhat dated compared to the other DSR methodologies sourced in the literature. It is somewhat interesting to note that Hevner et al.'s (2004) guidelines do not directly involve problem identification but they do call for the researcher to prove validity of the DSR by emphasising the problem relevance, scientific evaluation, the need to proof the research rigour and clear identification of the research contribution.

These guidelines provide a detailed set of requirements for any methodology, but they do not provide a comprehensive and detailed "step-by-step" for conducting DSR. In fact these guidelines do not seem to be sequential since it calls for problem relevance after artefact design. The guidelines also lack certain aspects that are common to most research, such as the role that the existing body of knowledge plays in the research.

Another DSR methodology sourced in the academic literature was that of March and Storey (2007). The methodology purpose by March and Storey (2007) contained the following six steps: the first step was the identification and description of relevant problems that can be addressed with DSR; second step involves a demonstration of no current adequate solution; third steps is the development and presentation of DSR artefacts; fourth step is the rigorous evaluation of the DSR artefacts in terms of utility, five step is the identification of the value the artefacts in regards to the IT knowledgebase and practice; the sixth and final step is an explanation of the implication that the artefacts have on the intended context of used.

What is of interest to the methodology of March and Storey (2007) is that it is necessary for the researcher to explain the implications of the artefact in the context of use. The most adequate means to do so would be to implement the solution and to conduct a long term study in order to gage these implications. This limits the application of this methodology, since it can seemingly only be applied to situations where the entire software process will be undertaken and the software will be implemented. This however is not the focus of the current research. Alternatively it could mean that it is necessary to predict these implications, but these predictions could potentially not include unknown emergent properties and phenomena of the software in practice.

The methodology of March and Storey (2007) does not make the same strong argument for using the existing knowledge (academic literature and expert experience) as some of the other methodologies. Potentially however showing that no adequate solution exists would involve a component of literature analysis or review, or even evaluation of the research area. However the design and development methodology does not explicitly refer to the design or development decisions to be grounded in the available body of knowledge.

Overall the methodology of March and Storey (2007) is remarkably similar to the other methodologies found within the academic literature.

The methodology for DSR purposed by Peffers et al. (2007) contains a number of activities: the first activity is the identification and motivation of a relevant research problem; the second activity involves the definition of the solution objectives; third activity is the design and development of DSR artefacts; fourth step is the demonstration of the artefacts ability to solve one or more problems; fifth step is the evaluation of the created artefacts and the sixth and final step is the communication of the research.

The methodology of Peffers et al. (2007) differs to most of the other methodologies in only a few regards, most notable is the need to explicitly show that produced research artefacts address the research problem in some regards. Showing the applicability of the artefacts to solve research problems occurs before the artefacts are evaluated, usually these two activities would conceivably be done at the same time.

Presented with a multitude of different guidelines and frameworks, most of which only partly applicable to the current research, it was decided that combining in the most relevant steps and guidelines into an amalgamated approach could be successful. A lot of commonality existed between the different frameworks but none of the individual frameworks seemingly fit exactly with the research.

Table 3.4: *Categorisation and grouping of the phases of the sources Design Science Research approaches.*

Category	Activity	Source
Research problem	Identify the Research Problem	Ellis and Levy (2010)
	Problem Relevance	Hevner et al. (2004)
	Identification and description of relevant problem	March and Storey (2007)
	Demonstration of no current relevant solutions	March and Storey (2007)
	Identification and motivation of relevant problem	Peffers et al. (2007)
Objectives	Describe the Objectives	Ellis and Levy (2010)
	Identification and motivation of objectives	Peffers et al. (2007)
Knowledge Search	Design as a Search Process	Hevner et al. (2004)
Conceptual Design	Concept design	Hasan (2004)
Design / development	Design and Develop the Artefacts	Ellis and Levy (2010)
	Construction of the system Architecture	Hasan (2004)
	Prototype	Hasan (2004)
	Design as an Artefact	Hevner et al. (2004)
	Development and Presentation of research Artefacts	March and Storey (2007)
	Design and Development of research Artefacts	Peffers et al. (2007)
Test / Evaluation	Test the Artefacts	Ellis and Levy (2010)
	Evaluate test results	Ellis and Levy (2010)
	Design Evaluation	Hevner et al. (2004)
	Rigorous artefact evaluation in terms of usability	March and Storey (2007)
	Demonstration of artefacts ability to solve research problem	Peffers et al. (2007)
	Evaluation of research artefacts	Peffers et al. (2007)
Research Rigour	Research Rigour	Hevner et al. (2004)
Communication	Communicate the results	Ellis and Levy (2010)
	Communication of Research	Hevner et al. (2004)
	Communication of research	Peffers et al. (2007)
Research Contribution	Contribution	Hevner et al. (2004)
	Identification of contributions (practical and academic)	March and Storey (2007)
Practical use	Practical implication of artefact usage	March and Storey (2007)
Other	Product Development	Hasan (2004)

By broadly grouping the various activities in the identified DSR approaches it became clear that there are several strong themes amongst the sourced approaches: namely the identification of problems relevant to DSR-type research; the construction of DSR artefacts and finally the evaluation of these constructed DSR artefacts. These themes are strongly related most likely because DSR seeks to address relevant research problems by creating artefacts which need to be evaluated in order to not only produce results but also show the relevance of the artefact in relation to addressing some aspect of the problem area.

Several secondary themes were also identified from Table 3.3: the creation/design of objectives for the research and the need to communicate the research. Communication of the research and its findings seems almost implicit when conducting academic research.

Less common to the sourced DSR methodologies were the topics of: Research Rigour, conceptual design distinct from solution design, the implication of artefacts in practice and commercial product development.

Although rigour in any form of research is important it is odd to see so few of DSR approaches containing it. This could be because rigour is required throughout the research process and thus these approaches assume that explicitly detailing rigour as a set phase is both unnecessary and potentially confusing because of the pervasive nature of rigour in the research process.

It seems that both the implication of the artefacts in practice and the commercialisation of the created solution are long term phases, since the process of accurately gauging the implications of an artefact can take years and potentially involve a number of different individuals. Commercialisation can also be time-consuming and involves a number of other factors beyond the initial research such as what is necessary to make a given software design/prototype ready for use in practice by real users or for commercial sale and not to mention the myriad of advertising and marketing related considerations that go hand in hand with commercialisation.

For this reason a combined framework was created using the various frameworks presented within the academic body of knowledge. In total the combined framework has seven distinct phases that for the most part sequentially follows from each other. Some exceptions do occur such as the design and development process can be repeated and additional literature review / contextualisation can also be conducted later (for example as part of the design and development).

The following framework was created and used for this research:

1. **Problem identification, understanding and motivation** – The identification and motivation of a relevant problem which can be solved using a DSR approach.

   This phase involves initial research (literature review) in order to better understand and motivate why the problem is important and why DSR is an acceptable research approach to be used to address the identified problem.

   This phase involves ensuring that no current solutions are available or if they are, identifying any gaps in the research, which could be regarded as research worthy (a problem, or even an opportunity).

2. **Identifying the objectives/focus of the research and solution** – Identification of the objectives, both of the research and of the intended artefacts. The artefact objectives are only identified in a situation similar to this research where the artefact/prototype's objectives and the research objectives are not the same but still related and both are relevant.

3. **Concept design** – In this stage knowledge is gathered (from literature in the case of this research) with the goal of conceptualising the intended problem space, generating the requirement of the solution and developing the solution concepts. This phase also involves selecting between possible solutions if multiple solutions are presented and if the focus of the research demands it.

   This phase thus reviews (gathers) current academic knowledge and can possibly create knowledge if solution concepts are evaluated (according to some devised criteria), compared to each other and chosen from.

4. **DSR artefact design and development** – A solution is designed using the previously identified concepts (step 3) to meet the identified objectives (step 2) to solve the identified problems (step 1). In this phase the DSR artefact are created.

   The artefacts are objects intended to both fill and intended practical purpose but are also knowledge containing objects (containing logic, design decisions and assumptions). This is a knowledge application and creation phase, but this phase might also give rise to the need for more knowledge gathering in order to clarify and new or previously unknown issues or concepts.

5. **Artefact evaluation –** The design and the prototype and or artefact developed from the design are evaluated to ensure they solve the problem and meet the objectives, research rigour will be included.

   Artefacts are knowledge containing objects, and tools used to conduct DSR. By evaluating the merit and demerits of the artefacts the researcher can show if the underlying logic, design choices and assumptions were correct or incorrect.

   This phase might also involve additional knowledge gathering in order to define the criteria used for evaluation. This phase also involves some data analysis.

6. **Research contribution** – The research contribution needs to be identified from the design and artefacts developed. By evaluating the artefacts a researcher can generate research data, but data can also be created based-on experiences or findings during the previous phases, such as the findings of the conceptualisation phase and findings and experience during the design and development of the solution.

   This phase also contains data analysis (depending on the type of data produced in the previous phases) and the creation of research results and findings.

7. **Communication** – The results of the research is communicated to the intended audience, in this case within a thesis. This phase involves making choices around data presentation.

This framework which was used in this research is a combination of the previous discussed DSR approaches sourced in the academic literature. Some of the phases of the various sourced approaches to DSR were similar and thus combined.

The figure 3.3 graphically represents the research methodology used in this research; this figure is used throughout the remaining chapter to demonstrate where the content of the chapter falls in relation to the methodology, as certain chapters may contain information relevant to one or more steps.

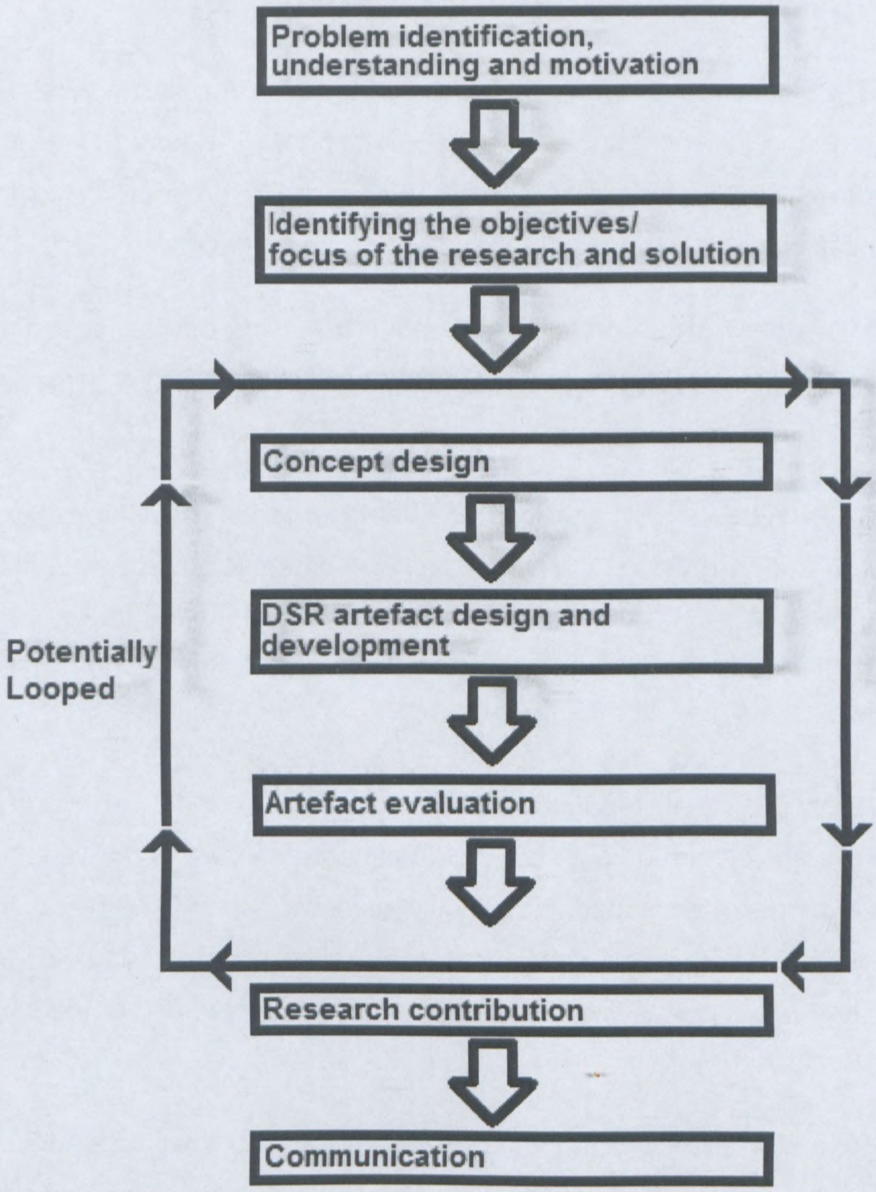


Figure 3.3: *Graphical representation of the research methodology.*

In the graph a potential loop or iteration is shown (denoted by the square with arrows). These steps follow on each other high-level conceptualisation is created for the solution. These high level conceptualisations are then used to create more concrete designs which are developed in the forms of artefacts and finally these artefacts are evaluated. This iteration can be done for the development of each artefact in the research. However in the case of this research the development effort follows an agile-scrum approach consisting of a number of iterations in which an iteration is focused on the development of a set of related functionality.

The next section looks in more detail at some of the justification for decisions made in the research.

3.8 Ethnographic study

As part of the Concept Design phase of the DSR methodology used in this research an exploratory ethnographic study was conducted at the Stellenbosch Hospice and the Motherwell community both in South Africa. This ethnographic study is intended to provide the kernel theories which aids in guiding the research and development efforts.

This section gives a brief overview of the ethnographic study focussing on the reasoning, orientation and purpose of the study as well as the products produced by the ethnographic study.

The Stellenbosch Hospice provides HBHC services to the Kayamandi Township outside Stellenbosch in the Western Cape Province. The Olive Leaf Foundation (OLF) is responsible for providing support and assistance toward community development efforts, amongst which enabling the provisioning of HBHC services to the Motherwell community located in the Eastern Cape Province.

The ethnographic study involved a detailed documentation analysis performed on several of the care documents from both initiatives. These care documents served as the primary source for identifying the usage of data and the identification of the metadata and the necessary data structures. However the document analysis only looked at the field and labels on the documents and not at any patient's personal information, because of the clear and obvious ethical issues involved.

Additionally information from a pre-existing ethnographic study on both of the healthcare initiatives was available to support the primary findings. The secondary ethnographic data was produced as part of the overarching project in which this research was conducted by another member of the overarching project, namely van Zyl (2011). Further secondary information was available from another project member, namely Delen (2011), who created a number of profiles and personas for the caregivers and patients, alleviating the need to involve real patients and caregivers.

The ethnographic study was conducted as part of the Concept Design phase of the research and a landscape model was created from the ethnographic date for two reasons: firstly it helped to contextualise the research and secondly it helped to create a better idea of where information originated, what information was being used and by whom and the flow of information.

The ethnographic study was thus an important part in creating the solution, requirements and concepts but also provided a detailed view of the HBHC initiatives.

3.9 Research Justification

This section considers the justification for specific aspects and choices made within the research.

Specifically three aspects are covered in this section: first is the reasons and justification for why the research was undertaken; second is the reasons and justifications for the using a DSR approach to conduct the research and the third is the justification for using a combination of the various DSR approaches found in the academic literature.

The following section considers the justification for the research.

3.9.1 Research

This section considers the justification for both why the research was conducted and for why the given research problem was addressed.

The research was conducted as part of a South African Finland Partnership (SAFIPA) in Information Communication Technology funded project conducted at the Cape Peninsula University of Technology (CPUT). The project's overall focus was on relieving tension in under-resourced and underdeveloped communities, specifically focussing on access to and the provisioning of healthcare within these communities. The research problem came to light as the project started to look at these communities and at the healthcare facilities that attempt to address the issues surrounding the provisioning of healthcare.

These initial problems identified during the projects were further elaborated using the academic literature.

The literature review, detailed in Chapter 2, pointed out issues faced by developing countries such as South Africa in providing quality health services and the challenges of delivering quality care in developing countries, especially within rural and poorer areas. The literature also showed that HBHC promised to address the lack of quality care services in rural areas of developing countries but that HBHC also faces numerous issues of its own.

One of the identified issues that inhibited HBHC was the inconsistent usage of data and data-elements across different HBHC initiatives and the lack of standardisation or attempts to standardise these elements. This issue was identified in both the literature and by the research project in which the literature was conducted.

The literature showed that most work on standardising data in healthcare was either not focused on the rural HBHC context or involved conforming to a given standard without first understanding the current situation of usage and understanding of the various data-elements. These inconsistencies and misinterpretations lead to numerous issues when developing IT based solutions for HBHC which if resolved could potentially greatly benefit the rural South African HBHC domain. From this the justification arose for developing an IT based solution that attempted to better conceptualise and understand the issues surrounding data-elements arose.

The semantic metadata repository as a potential solution arose from the literature review. A repository-based solution offers a means of storing and sharing information, the metadata and semantics provides a more abstract (at the ontology level) representation of the data-elements used in HBHC. Because of this abstracted representation it limits the need to have potentially confidential medical information involved in the research process.

However simply creating an IT based artefact doesn't necessarily constitute research. The above provided the motivation for developing the semantic metadata repository, but without being able to sufficiently show rigour and relevance, the artefact itself could not constitute research, even if a DSR approach was being used.

During the literature review and from the overarching research project it became apparent that HBHC in rural South Africa had a low ICT penetration and that many of the ICT undertakings are unsuccessful. This showed that there was room and a need for further research in terms of analysis and design of IT-based solution intended to function with the HBHC context.

While this section showed the justification for the research and for the research problem, the following section considers the reason why a DSR approach was chosen.

3.9.2 Research Approach

This section considers the justification for why a DSR approach was used to conduct the research.

The desire to follow a DSR approach was born from the requirement of the overarching research to produce a useable IT artefact/prototype and by the researcher's personal preference to conduct a more "technological" and "practical" orientated research that was still rigorous and could contribute to overall academic knowledge.

Because DSR is still immature in relation to other forms of research within the IT and IS field (Wang & Wang, 2010); conducting the research using a DSR approach offered an opportunity to contribute to the academic knowledge on multiple levels, not only in terms of the research result but also being able to contribute knowledge to the field of DSR itself.

Because rural HBHC in South Africa has a very low penetration of IT and HI-based solution, conducting research using a DSR approach had an additional advantage. This advantage was born form the fact that DSR by its nature takes the practical issues into account which leads to more an accurately gaging of the feasibility, usefulness and design considerations involved. Further the practical aspect of DSR allowed for the issues and advantages of implementing IT based solutions to become more apparent. This is potentially of importance to others also wishing to design and develop solution for rural South African HBHC. The human components are not necessarily easier to gage using a DSR approach but within this research the human components did not form the primary focus of the research.

The artefacts developed as part of the DSR process also provided valuable tools that aided in examining the intended research context and answering the research questions. Some of the research questions within the research would not easily have been answered using behavioural research methods especially those relating to technical aspects. Thus a design and development based approach was most useful within this research.

Using a DSR approach also helped to manage the obvious ethical considerations involved when working within the healthcare space. The DSR approach that was used focussed more on the design and development of the research artefacts and lightly touched on the intended context of use and users, meaning that no ethically questionable data was collected and no patient information was needed.

The first section justified the research and the research problem while this section justified why a DSR approach was used to conduct the research and address the research problems. The next section looks specifically at why a DSR methodology was used that combined the various DSR approaches sourced within the literature.

3.9.3 Combined Methodology

The framework that was used in the research was a combination of the approaches proposed by Hasan (2004), Hevner et al. (2004), Peffers et al. (2007), March and Storey (2008) and Ellis and Levy (2010). Most of these frameworks have similar activities/steps and except for Hasan's (2004) and Hevner et al.'s (2004) approaches the remainder have a strong overall resemblance to each other. This meant that combining the various approaches was fairly easy. The combination process was detailed in Section 3.7 titled Methodology.

This section considers the reasoning why none of these sourced methodologies met the exact requirements for the research and why a combined methodology was used.

The four stages of DSR purposed by Hasan (2004), as shown in Table 3.1, is aimed towards the development of commercial software, or software that is intended to be used in practice so much so that the fourth stage is entitled 'product development. As part of the research conducted, only an initial prototype (proof of concept) was developed making only the first three stages relevant and the last however irrelevant. Looking only at the first three stages of Hasan (2004) DSR approach it becomes clear that it is less verbose than the other sourced approaches but that what was left was fairly detailed.

The DSR approach purposed by Hasan (2004) is thus not fully applicable to the current research as the developed artefacts are not intended to be commercialised. Furthermore the research focuses only on the analysis, design and development phases since the process of preparing the software for commercialisation brings to the forefront a number of different and unique requirements that would obfuscate the focus of the research.

The approach purposed by Hevner et al. (2004), as seen in Table 3.2, is fairly detailed but by closely examining these it becomes clear that these are merely guidelines, not a sequential list of steps needed to conduct in order to perform design-based research. The guidelines are important however and when comparing these guidelines to other DSR approaches, it is notable that some of these guidelines are distinct phases within the other DSR approaches, such as the Design Evaluation, Identification of research contribution and communication of the research.

The guidelines of Hevner et al. (2004) are also notable however for explicitly noting the need for rigour within the research, which is not explicitly stated in the other sourced DSR approaches. Although not a true methodology the guideline of Hevner et al. (2004) can be used in conduction with the other sourced approaches to help refine and clarify them.

The approach of Ellis & Levy's (2010) provides clear considerations and guidelines for conducting DSR. Although these guidelines do not clearly show how such research needs to be conducted as it omits both detailed steps and a practical example of their purposed approach. The article in which the approach appears however was aimed at the novice researcher and focused more on explaining the general research approach in tandem with the DSR approach.

In the article Ellis and Levy (2010) first purposes a generic research approach before detailing their DSR approach; however these two approaches bare a strong resemblance to each other. Seemingly Ellis and Levy (2010)'s DSR approach is a specification of a more generic research approach, but still general enough to be applicable to most forms of DSR, making the approach a good foundation to create the remaining approach.

The DSR approach of March and Storey (2007) is in line with the other DSR approaches in that it is based around the concepts of artefact development and artefact evaluation. A few notable phases do exist within the DSR approach of March and Storey (2007) specifically the demonstration of no current adequate solution and the need to define the implication of the artefact on the intended context of use. These two steps imply that the approach of March and Storey (2007) is focused not only on using the artefact as research tools but also focused on how these artefacts is used to resolve an as yet unresolved issue within practice. The approach of March and Storey (2007) thus attempts to generate knowledge not only surrounding the design and development process but also beyond that to generate knowledge about the practical application of the designed and developed artefacts.

March and Storey's (2007) DSR approach was not applicable to the current research, because the current research focused on using the research artefacts as tool to examine the HBHC context as well as the design and development of a software solution intended to be used in rural HBHC. The research however, unlike the approach of March and Storey (2007), does not on the impact of the IT since this would constitute a long term study over a prolonged period beyond the design and development phases which the current research focuses on.

The approach of Peffers et al. (2007) is fairly standard but is notable in how it ensures research artefact rigour, in that it is necessary for the researcher to show the utility of the artefact. The research artefacts are however created with the intent of solving a given problem and thus, although important, this step seems to be more fitting as a subset of the evaluation phase.

With so many different approaches to conduct design-based research within the IT field presented in the academic literature, it was clear that there was no standard DSR approach. The various approaches found in the literature did have their own merit and demerits with each seemingly being aimed at a specific subset of based-based research. The combined approach was thus decided on to make it more applicable to the current research with the intent of ensuring more valid research.

3.10 Data Collection

This section considers the data collection methods used in the research.

Data generated by DSR's "build" and "evaluate" needs little intricate analysis. The build and design phases created artefacts based-on concepts created in the concept design phase. These concepts contain their own assumptions about the environment, the context and the problem it seeks to address (Gregor, 2002). Evidence for the validity of the data generated is referred to as proof of concept (Hasan, 2004) which is produced during the Evaluation Phase of the DSR methodology.

Evaluation provides feedback for the constructed artefacts (Hevner et al., 2004) which are constructed in order to solve an identified problem or meet identified objectives. The artefacts are built with a number of assumptions and ideas at their core. The evaluation of the artefacts thus provides a means of validating or invalidating these core assumptions which in turn provides a possible deeper understanding of the problem or the context. These assumptions can furthermore be expanded if an underlying logic is based-on or can be associated with the current body of knowledge.

If an artefact is deemed to work within a given context then it can be assumed that the assumption about the context is in some form correct. It an artefact manages to solve a problem it can be assumed that the assumptions about a given problem are correct.

A researcher can however reflect on the creation of the artefacts, on the design and development process in order to analyse the results further.

Ellis and Levy (2010) identified three important considerations that needed to be taken into an account during the evaluation phase. These considerations included: identifying how an artefact does and does not meet the requirements; use accepted literature-supported process and finally ensuring that value of an artefact.

In the literature it was noted that the evaluation of artefacts required that the appropriate metrics and measurements be developed first (Hevner et al., 2004), Hevner et al. (2004) purposed that general measurements with which artefacts could be measured included: functionality, completeness, consistency, accuracy, performance, reliability, usability and finally comparability with the organisation intended to function it.

Generic metrics were provided by March and Smith (1995) that could be used to measure the different types of artefacts. It was purposed that constructs be evaluated according to their: completeness, simplicity, elegance, understanding and ease of use. Models are evaluated according to their: fidelity to real world, completeness, level of detail, robustness and internal consistency. Methods were purposed to be evaluated according to: operational, efficiency, generality and ease of use. Finally the development of instantiations according to: efficiency and effectiveness, impact on users and environment.

Evaluation is needed on two levels, the solution artefact and the research. The solution artefact needs to be evaluated to ensure that it is correct or meets the required outcomes, since this relates back to its ability to be a source of research data. The research (in this case the design and development process) needs to be evaluated as well, as this can also be a source of research data.

The design theory of Gregor and Jones (2007) has a number of components that can be used as a lens to look at and evaluate the research as well as the artefacts produced by DSR. The design theory of Gregor and Jones (2007) consists of eight components: purpose and scope, constructs, artefact mutability, testable propositions, justificatory knowledge, principles of implementation and finally expository instantiation.

From their literature the product specifically can be evaluated using the NISO standards for good metadata as well as Thibodeau's (2006) axes for evaluating a digital repository.

NISO purposes several requirements for good metadata which can be used to evaluate metadata, including: conform to community standards; support interoperability; use authority control and content standards; include a clear statement of the conditions and terms of use; support long-term curatorship and preservation; should have qualities of the good object: authority, authenticity, achievable, persistence and unique identification (Park, 2009).

The NISO purposed requirements have some merits, but some of the criteria are seemingly not universal, or at least do not apply to this specific research. Specifically conformation to community standards and support for interoperability is not seen as relevant to this research.

Conformation to community standards is not applicable. The research considers semantic structural metadata, as has hopefully been shown during the literature review in Chapter 2, there are little community standards for structural metadata and most metadata data standardisation work is based around descriptive metadata such as the Dublin Core.

The metric of support interoperability is a fairly grey area in this research because the purposed structural metadata model is not intended to support interoperability, in fact as has also hopefully been shown in the literature review not all metadata exists to facilitate interoperability. It would be possible to, with some work, adapt the purposed structural metadata standard to be used to translate between ontologies, since the metadata standard is intended to essentially capture ontologies.

The three axes by which a successful repository can be measured was stated by Thibodeau (2006) as being: orientation (retrospective/storage or prospective/usage), coverage (the content to be stored), collaboration (where it falls on the isolated versus collaborative spectrum). However Thibodeau (2006) did not specify any exact criteria by which to measure a repository.

The intended solution is a prospective, semantic metadata repository with limited collaboration and can be broadly measure along these axes.

This section considers the analysis of the data produced by the research, explained in more detail in the previous section. The following section looks at the presentation of the data.

3.12 Data Presentation

As noted several times before, DSR produced artefacts which are commonly identified as: models, methods, constructs and instantiations, most of which can in themselves be seen as some form of data presentation.

The prototype system itself (the instantiation) is presented as a proof of concept by showing the constructs that make up the prototype in the form of UI mock-ups and UI screenshots.

The ideal presentation of the models is in the form of: use case, conceptual diagrams, entity-relational diagrams, data flow etc.

Methods are also presented as to how certain aspects of the system were achieved, primarily in the conceptualisation.

3.13 Rigour

This section considers how rigour is insured within the research.

Hevner et al. (2004) noted that not only in DSR but also in behavioural-science research within the IT, filed rigour is a product of the effective utilisation of the knowledge base, the theoretical foundation and the research methodology. It is clear then that rigour in most research undertakings can be strengthened by drawing parallels to the current body of knowledge, such as academic works, and by applying commonly accepted methods during the research process.

But furthermore Hevner et al. (2004) noted the risk that overemphasis on rigour can lead to a lower relevance. Wang and Wang (2010) noted that traditionally relevance in DSR is built on need and that there is a difficulty, especially compared to behavioural research, when trying to apply rigorous tests to help ensure rigour. Rigorous tests can be difficult and time-consuming to create, possibly taking longer than the actual artefacts life cycle (Wang & Wang, 2010).

Specifically to DSR Hevner et al. (2004) noted that success is usually based-on the researchers' ability to select appropriate techniques for the construction of artefacts and the appropriate selection of means of evaluation, to generate the findings, and justification for the findings. To that extent Hevner et al. (2004) proposed a number of guidelines which are properly applied not only manage to contribute to the body of scientific knowledge but also to help ensure rigour in the research process.

The repository is evaluated according to Thibodeau (2006) axes for a successful repository, the metadata structure and implementation is measured according to the most relevant of the NISO metadata criteria and the design theory of Gregor and Jones (2007) is used to reflect on the research. Using criteria of evaluation based off findings within the academic literature is hoped that a more rigorous research was undertaken. In the Data Analysis Section (Section 3.11) the evaluation metrics are discussed in more detail.

Furthermore Wang and Wang (2010) provided four guidelines for conducting DSR namely: design as a new layer based-on new IT artefacts; design with the goal of improving existing IT artefacts or relevant phenomenon, design as a means of improving business practices by using new IT; design as an innovative instantiation.

It is hoped by using these guidelines provided by Wang and Wang (2010) and by using an amalgamated DSR methodology the overall research has a higher level rigorous.

By removing or minimizing the subjective influence of the designer, and in this case researcher, Purgathofer (2006) noted the scientific principles of research can be achieved, specifically repeatability, transparency and provability. However because design is considered to be a creative process this can be difficult. Rather the research attempts to make these biases and influences as visible and transparent as possible.

3.14 Assumptions

Research assumptions were described by Ellis and Levy (2010) as: "underlying all research efforts... things that researchers accept as true without concrete proof."

This section considers the assumption underlying the research.

It is assumed that South Africa is an example of a developing country, but that it has both characteristics of a first and third world country. It is further assumed that areas with first world characteristics are not far removed from areas with third world characteristics, separated by only a few kilometres at most.

Due to the complex nature of the problem that the IS-solution being developed as part of the DSR attempts to solve, it is assumed the intended user of the solution will be an expert user with some form of IT-training and a basic understanding of data and data relations. This does not mean however that the caregivers and nurses that form part of the HBHC environment will not be able to contribute but that the expert user will interact with the caregivers and nurses as needed to provide input for the system. This assumption directly impacts the process flow and UI design of the semantic metadata repository, and allows for less consideration to be given to the Human Computer Interaction component as well as the type of users who might be using the solution.

Hevner (2004) argued that the artefact and the context (the environment and the people) are both equally important from a research perspective. The assumption of the research is that Hevner's (2004) equality argument is in fact correct, that usability and functionality are both equally important. Furthermore it is assumed that the usability of a given solution is directly related to the context of use as well as the intended users, but that the underlying functionality is less reliant on the users. Therefore the context is taken into consideration; but that intense or deep user involvement will not be able to deliver generic research knowledge as it would potential cater for only the needs of the involved user(s).

3.15 Delineation of Research

Delineating the research and defining the intentional constrictions of the research is an important part of DSR (Ellis, Levy 2010) and of most other forms of research as well.

The analysis component of the research is focused primarily on HBHC within the Western Cape with supplementary findings form the Eastern Cape. The analysis component does however not exhaustively look at, the numerous care activities or at the patient involvement at the care process. The focus will be on the institution and on the institution-caregiver interactions, with limited focus on institution-governmental body and caregiver-patient interactions.

It is far too ambitious and time-consuming, to attempt to cover all aspects involved in designing and developing a semantic metadata repository, from the initial development through the implementation and eventually until the retirement of the solution.

The research is design and development based and thus looks primarily at the technical aspects involved in the development of the system. In this research attention is paid to the human related aspects only when needed and only when it relates to the design and development process. These human factors include amongst others the interactions of the healthcare workers with IT systems (the human computer interaction components) or required information and information usages. Limited attention is be paid to the user interface and user interactions with the system, or more correctly these components are not seen or treated as major contributions within the research.

A common issue with DSR research, specifically the products produced in it, is the fact that it is perishable, new technology developments can easily cause DSR research findings to become invalid or other innovative applications of technology can change the context in which the DSR research is conducted enough to negate any intended effect of the DSR product (Ellis & Levy, 2010)

Five factors were identified by Hevner et al. (2004) which are common to DSR including: poorly defined environmental factors such as requirements; an inherently complex problem and solutions; flexibility for chances to solutions; a solution dependent on human creativity or a solution dependent on collaboration.

The brief ethnographic study is used to address some of these issues, to help better understand the environmental factors, the problem and solution. The ethnographic study is not intended to be complete and comprehensive. The ethnographic data collected is simply meant to help define the environmental factors and to clarify the complex issues surrounding the problem domain.

The design considerations addressed by the research is those design considerations closely related to the technical aspects of the system and not the usability aspects.

3.16 Ethical Considerations

Ethics in terms of research was defined by Iivari (2007) as concerning: "the responsibility of a scientist for the consequences of his research and its results".

Although the research is being conducted in a field of HBHC there are little ethical issues to consider. Although the healthcare context, specifically HBHC, does play a role within the research, no personal or private information or details was gathered or looked at during the research, nor is it the intent of the research to reveal any personal or ethically unsound information.

DSR in IS and IT considers the technical aspects of developing of solution, using a number of different types of artefacts (as detailed previously). These are design artefacts which themselves do not need to in any way contain ethically questionable information.

There are only two possible areas of the research that might raise the need for ethical considerations. There is a need to define the context and requirements of the intended solution artefact, which requires an understanding of the problem domain and the variables therein. The second area is in the testing of the solution artefact, in this case the solution is intended for usage HBHC information.

These two ethical considerations can however easily be managed. Permission has been granted by the target HBHC initiatives to do an ethnographic study, which was conducted as part of the SAFIPA project, and helped to contextualise the solution. The institutions also granted access to empty copies of the paper forms they used to conduct their care processes. The ethnographic study and the document study were deemed sufficient to contextualise the problem and solution domain.

As for the second ethical issue, the testing of a solution artefact can easily be managed by using dummy data. The initiatives provided sample data to the SAFIPA project, which can be used for testing, although the focus of the solution artefact is not on the data used by HBHC but on the metadata, which is not person or deceased specific and thus should not have any ethical concerns associated with it.

3.17 Contributions of the Research

A number of possible contributions of DSR identified within academic literature.

Hevner et al. (2004) defined the contributions of DSR: the design artefact (what the researcher develops, models, ontologies, design algorithms etc.); which can be implemented to solve real life problems; improvements in the existing design-science knowledge base (by creating new methods, evaluation methods etc.); which can help refine future DSR and finally methodologies (by implementing metrics, evaluation methods etc.).

March and Storey (2008) identified the requirements that a DSR undertaking needs to meet, to be considered a valid contribution namely: identification and clear description of the research problem; demonstration that no adequate current solution exists; development and presentation of a novel IT solution; rigorous evaluation of the constructed artefacts and articulation of the value added to the knowledge-base and explanation of implication for management and practice.

The intended DSR framework being used is a combination of several other DSR frameworks found within the academic literature. It is possible that the reasoning behind using a combined framework might be of some contribution to the body of DSR knowledge.

The contribution of a successful research process is the first of Hevner et al. (2004) mentioned contributions, which is the design artefact. The design artefact within this research is a semantic knowledge-based tool, which in itself can serve as a common knowledge base which can serve as a common component in the development of healthcare systems, reducing cost and speeding up development.

The contribution to future development effort is also supported in the literature. Hasan (2004) noted that during DSR knowledge is not only developed about the process of developing an artefact but also of the context, a deeper understanding of the organisation and the environment in which the solution artefact seek to solve a problem.

The most likely contribution however is that the knowledge and insight gained during the design and development process is of use to others who want to continue research into the topic of semantic metadata repositories and also to those that want to develop their own knowledge base.

It was Hevner et al. (2004) who saw the goal of researchers in IS as creating knowledge about the usage and management of IT in human organisations and generating insight which could aid in the successful and practical application of IT solutions.

It is clear that industry and policy makers are interesting recipes to solve problems, and not the deeper understanding and logic which underlies the recipes (Gregor, 2002). March and Smith (1995) saw the information produced by IT research as of immediate or future perceived value on actions and decisions. The DSR research can therefore contribute to decreasing the uncertainty involved in future similar development efforts, by positively influencing actions and decisions and increase the possibility of success (Gregor, 2002).

The semantic metadata repository and care data ontology can be used by government to obtain a better understanding of how data is utilised on the ground level to inform their policies. Furthermore developers of HI systems will be better informed about how to design such systems to improve interoperability to better share data. Researchers will also be able to analyse the captured indigenous knowledge of caregivers to obtain a better understanding of home-based care provision in different communities.

Model/Artefact Development

Chapter 4 Conceptualisation

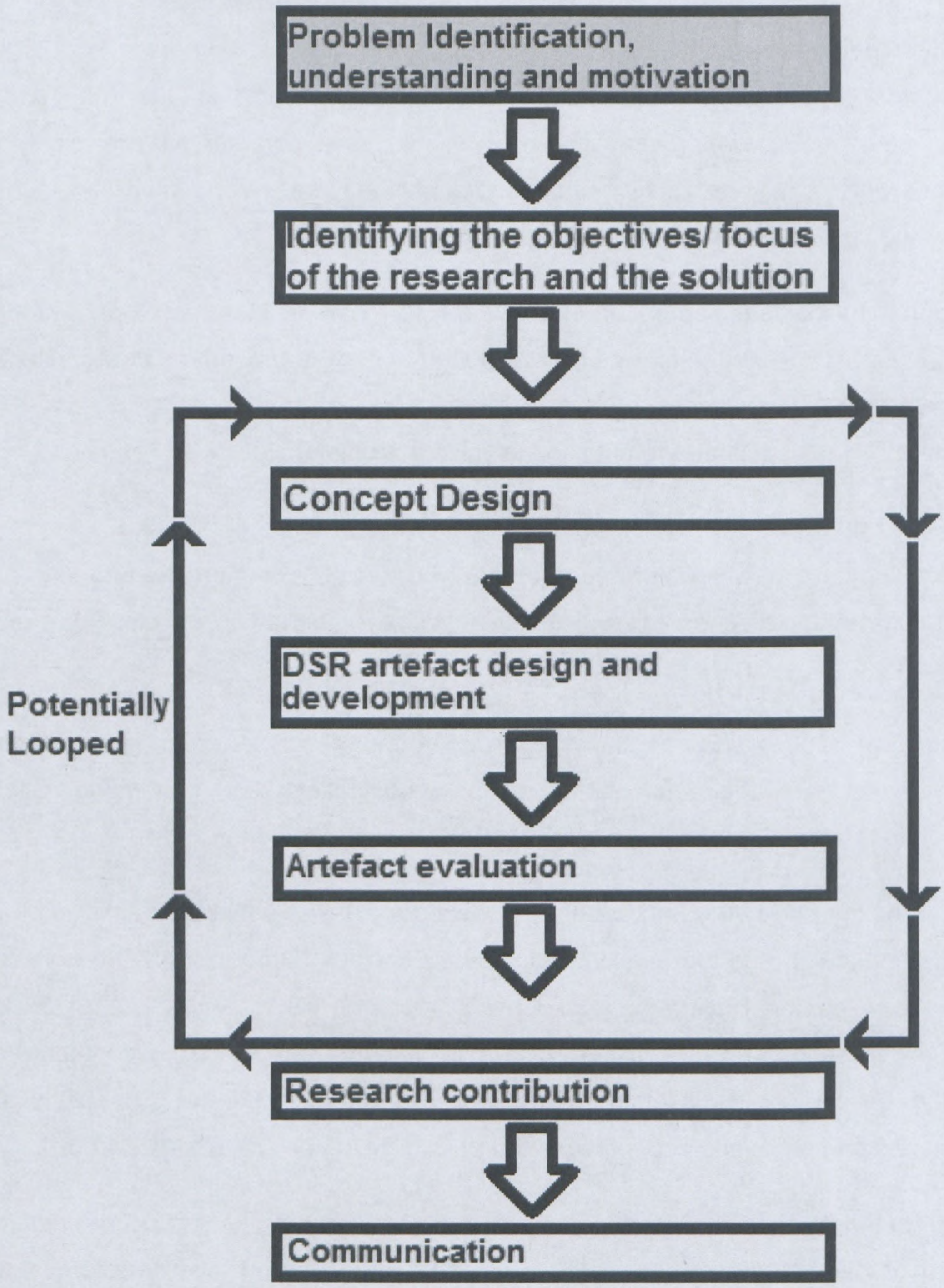


Figure 4.1: *Content of the Conceptualisation Chapter in relation to the research methodology.*

This chapter touches on the first phases in the DSR methodology utilised in this research shown in Figure 4.1, these phases are: problem identification, understanding and motivation. This research is highly contextual and in order to obtain sufficient insight in the problem the findings of an ethnographic study were considered to sufficiently understand the problem. The ethnographic data was also used in various other parts of the research and is not only relevant to the problem identification.

In the previous chapter the research problem was detailed and the objectives were given as part of the research methodology discussion. In this chapter the ethnographic results are discussed as part of refining the research problem. The research objectives are discussed in the following chapter in order to align the chapter structure with the methodology used for this study.

While the initial ethnographic study was conducted for the overall project the research problem for this study was identified and refined.

The ethnographic findings relating to this study are discussed next.

4.1 Home-Based Healthcare Context

The first section considers a general conceptualisation of the HBHC context; the second section looks specifically at the Stellenbosch Hospice and Kayamandi community in the Western Cape Province as well as the Olive Leaf Foundation and Motherwell community in the Eastern Cape Province.

4.1.1 General Home-Based Healthcare Context

This research considers the design considerations of a semantic metadata repository in HBHC while the underlying development attempts to create the semantic metadata repository. As a result of this various research artefacts were created to identify these design considerations. Essentially the development feeds into the research, providing the research data. The "research" and "development" attempt to address different but related problems and both are within the same context, that of rural HBHC.

This section specifically considers the HBHC context, in order to better understand the research area. The insights provided by the ethnographic study serve to guide the solution/development process, providing the problem, objectives and requirements whilst this context also provides a lens for identifying and placing the research findings to a specific context. The general HBHC context provided a good initial starting point for the research and the initial solution design / conceptualisation.

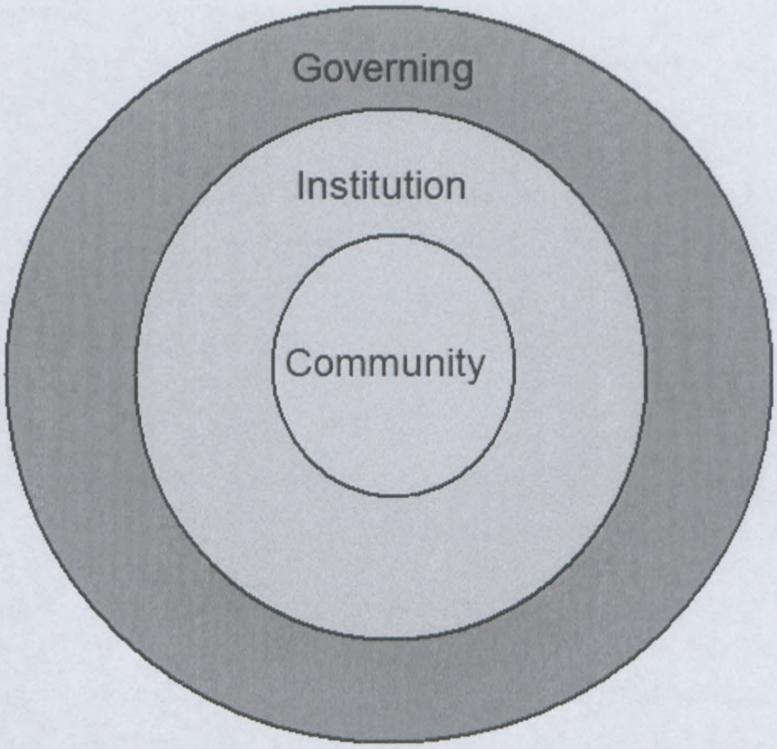


Figure 4.2: *onion model of the HBHC context.*

From the initial ethnographic research a simple onion model conceptualisation, Figure 4.2, of the interrelating layers involved in delivering home-based care services was created. The onion diagram provides a general overview of the different levels involved in the provisioning of healthcare to the communities.

The onion diagram was useful to represent the different layers involved in the delivery of HBHC services because it provides for the conceptualisation of these levels as follows: the community level is where the need for healthcare services originates; the institutional level provides the care services to meet the needs that originate in the community level; the governmental bodies provide the rules, regulations, some support and prerequisites that regulate the institution level in its attempt to provide care services to the community level.

The definition of a community within this research was kept as simple and inclusive as possible. A community was seen as a group of people who collectively associate themselves with each other, through geographic relations and similar social and economic situations. From the research there was a strong theme that the communities defined themselves as a community based-on these two criteria (location and situation). However the focus of the research was not to define what a community is or how those within Kayamandi and Motherwell perceived themselves or grouped themselves, or even how they defined communities. For that reason the research assumed a simple and inclusive definition.

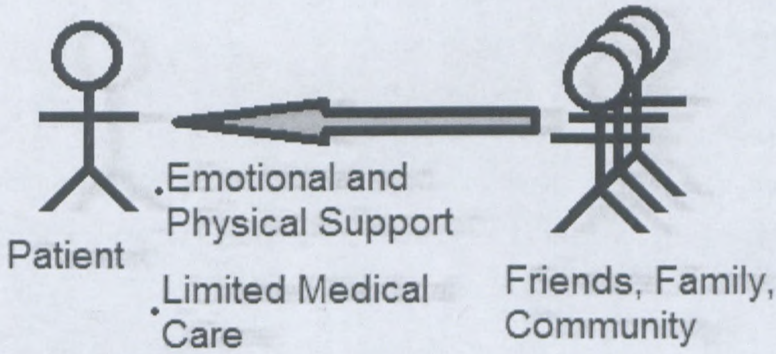


Figure 4.3: *Patients and their Community.*

From the beginning it became apparent that most of the patients in the rural communities have a strong supporting structure around them, consisting out of neighbours, friends and families who play an important role of supporting the patient and ensuring that the patients receive the necessary care. Individuals in the community also have a strong sense of community awareness and community responsibility, which is evident since most of the informal caregivers are volunteers from their own community.

The institution layer of the onion diagram above consists of a number of different types of institutions such as: HBHC NGOs, clinics, hospitals etc., ranging from formal to informal, public and privately run. All these institutions aim to provide care services to the community. Institutions within the institutional layer can fall within a varied spectrum between the government and the community. Some institutions having limited to no government involvement and relying mostly on community support while other institutions are government administered, funded and located outside the communities that they serve.

Within the onion model above the caregivers, who collectively form the bulk of the HBHC initiatives ability to deliver home-care, were most difficult to place. The ethnographic study showed that most of the caregivers were semi-trained volunteers from the communities where they provided care, most even still living within those same communities. The caregivers are a part of the community but they are also a part of the HBHC initiative, they receive basic training, support, management and limited payment from the HBHC initiative and serve as a link between the initiative and the community.

The caregivers provide a wide variety of services outside of simply basic nursing care. Caregivers provide emotional support and simple domestic services (such as cleaning the patient's home if they are physically unable to do so themselves) to their patients. In addition caregivers also provide education, warnings and information to the patients and other community members about the prevention and care of certain diseases and conditions. These services that the caregivers render are defined by the careplans. These careplans are patient-specific and created by the HBHC initiatives usually by formally trained healthcare workers such as a nurse if a nurse is available.

During the caregiving process the caregivers also take notes and record details of the patients, their families and their living environment and the care the patient received. These notes taken by the caregivers serve as a form of feedback allowing the HBHC initiatives to monitor the patients, their families and environment.

The government level consists of not only the Government itself but various government bodies, such as the Department of Health, Department of Social Development or sponsoring bodies. The sponsoring bodies are not directly involved in the care delivery but provide the support and means that enable the facilities the deliver care, this support is usually monetary. The South African Government Department of Health is not directly involved with providing care to the communities either, but enables and regulates a number of other initiatives which provide the community care.

The Stellenbosch Hospice in Kayamandi and Olive Leaf Health Clinic in Motherwell are both NGO supported and sponsored by parent organisations (who in turn are supported to some extent by government bodies). Having these sponsoring organisations is important because the patients to whom the HBHC initiatives provide the care service do not necessarily have the means to pay for the care services they receive. The sponsoring organisations do impose certain requirements on the HBHC initiatives, usually requiring the HBHC initiatives to provide statistical and aggregated feedback.

All the different layers above have their own information, data and ICT requirements, necessary to meet their varied purposes.

In order for the governmental bodies to be able to provide the regulations for the underlying institutions level, they need to have an understanding of the actions performed by the institutions as well as the needs of the communities. The governmental bodies need adequate information to be able to make correct and timely decisions required to optimise the functioning at the institution level and where possible provide some form of support where necessary.

The institutional level provides the care services to the community but in order to meet the community needs they must firstly be made aware of those in need and what services they need. It is also possible for the different types of institutes to also provide information to each other, a hospital, clinic and HBHC NGO when patients are referred allowing them to share information.

As previously mentioned the institutions also need to report to the governmental bodies and for this purpose they need accurate, up-to-date and complete information. This aggregated information is predominately used for statistical purposes. Governmental bodies also provide requirements for the institutions, such as the type and nature of the data they require as well as the types of care they support.

At the lowest level is the community. From the research it became evident that the communities for the most part are fairly distinctive but do share a number of characteristics which influence their healthcare requirements and context. It is the community that defines what care is required and the environment in which the care must be provided. In the case of HBHC the community is also the source of many of the caregivers and a source of support, such as a means of funding. The community for the most part helps to define the nature of HBHC initiatives, since the community provides the need which the HBHC initiatives attempt to satisfy.

4.1.2 Specific Home-Based Healthcare Context

In order to improve the conceptualisation of the HBHC context, the onion model above was replaced by a landscape model which was specific to the two communities in which the research was conducted.

The reasons for replacing the onion model with the landscape model included: the landscape model provides more detail on the interplay between the different levels (government, institution, community); the landscape model also provides a deeper level of understanding of the data flows between the different layers; the different stakeholders and participants can be named.

In DSR various artefacts are designed and developed as part of the research, but these previously mentioned unknowns all play an important role in defining the characteristics of the DSR artefacts. However because these are both unknown and very important it is necessary that they be known to insure the creation of successful artefacts (and thus successful research). The initial HBHC contextualisation was a good starting point for the research but it was not detailed or comprehensive enough to reveal all aspects, meaning that more detailed and specific ethnographic data was required.

The Kayamandi community was closer and this context was initially studied to understanding the HBHC services provided to the Kayamandi community by the Stellenbosch Hospice. The Motherwell community is located in the Eastern Cape and because of the distance only a single visit to Motherwell could be undertaken. Kayamandi and Stellenbosch Hospice thus provided the primary source of information while the Motherwell community provided, less detailed but still useful, comparative information.

In the following subsection the Kayamandi community and the Stellenbosch Hospice which provides HBHC to the community, are considered. The second subsection considers the Motherwell community and the Olive Leaf Foundation which provides a number of services to the Motherwell community including HBHC.

4.1.2.1 Kayamandi

This subsection provides a more detailed and specific discussion for the Kayamandi community just outside of Stellenbosch in the Western Cape province of South Africa.

The following figures (Figure 4.4 and Figure 4.5) appear in the reception area of the Stellenbosch Hospice's offices, providing both employees and staff with a reminder of the vision and mission of the Hospice.

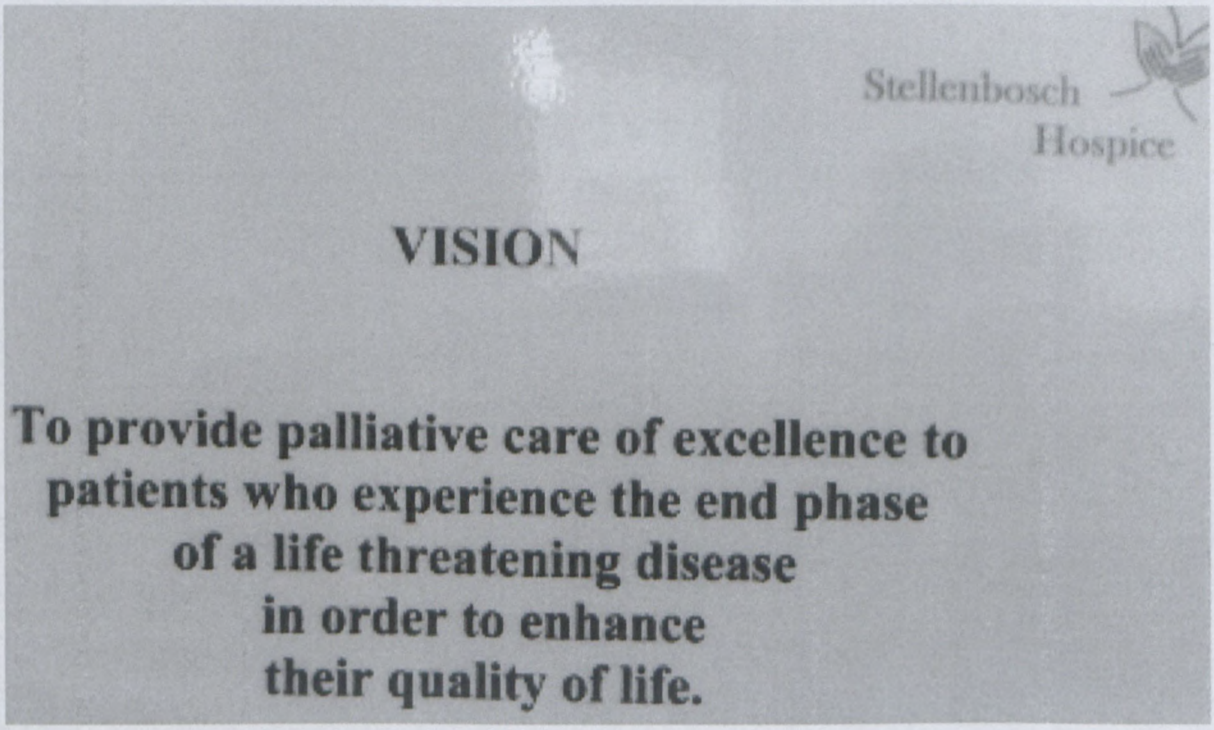


Figure 4.4: *The Stellenbosch Hospice Vision Statement.*

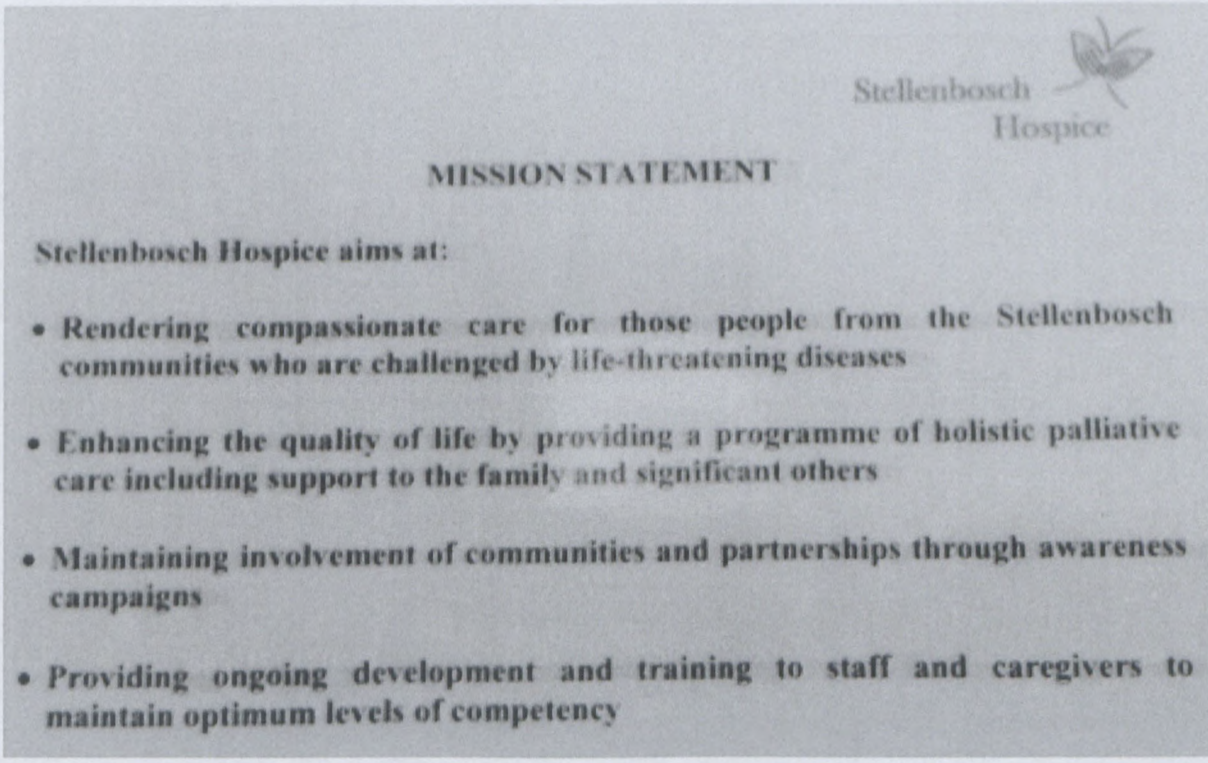


Figure 4.5: *the Stellenbosch Hospice Mission Statement.*

The following information was provided by the Stellenbosch Hospice:

Table 4.1: *Stellenbosch Hospice statistics for 2009 to 2011.*

	2009	%	2010	%	2011
Increase in patient numbers	426	60	685	30	888
Increase in staff numbers	64	14	73	10	81
Increase in operational costs	-	18	-	20	-
Current number of caregivers	-	-	-	-	40
Current number of nursing sister's	-	-	-	-	9
Current number of patients	-	-	-	-	888
Current number of patients in Kayamandi	-	-	-	-	202

Table 4.1 provides some statistics about the operations of the Stellenbosch Hospice, some information such as operational costs were included while other information was not available. The information does provide a clear insight into the current operations of the Stellenbosch Hospice.

The Stellenbosch Hospice is part of the South African Hospice group, with approximately 200 other hospices across South Africa.

Approximately 10,000 patient visits occur each month in the Stellenbosch Hospice, the government requires that caregivers visit a minimum of 10 patients a day during a 5 hour working day, 5 days a week.

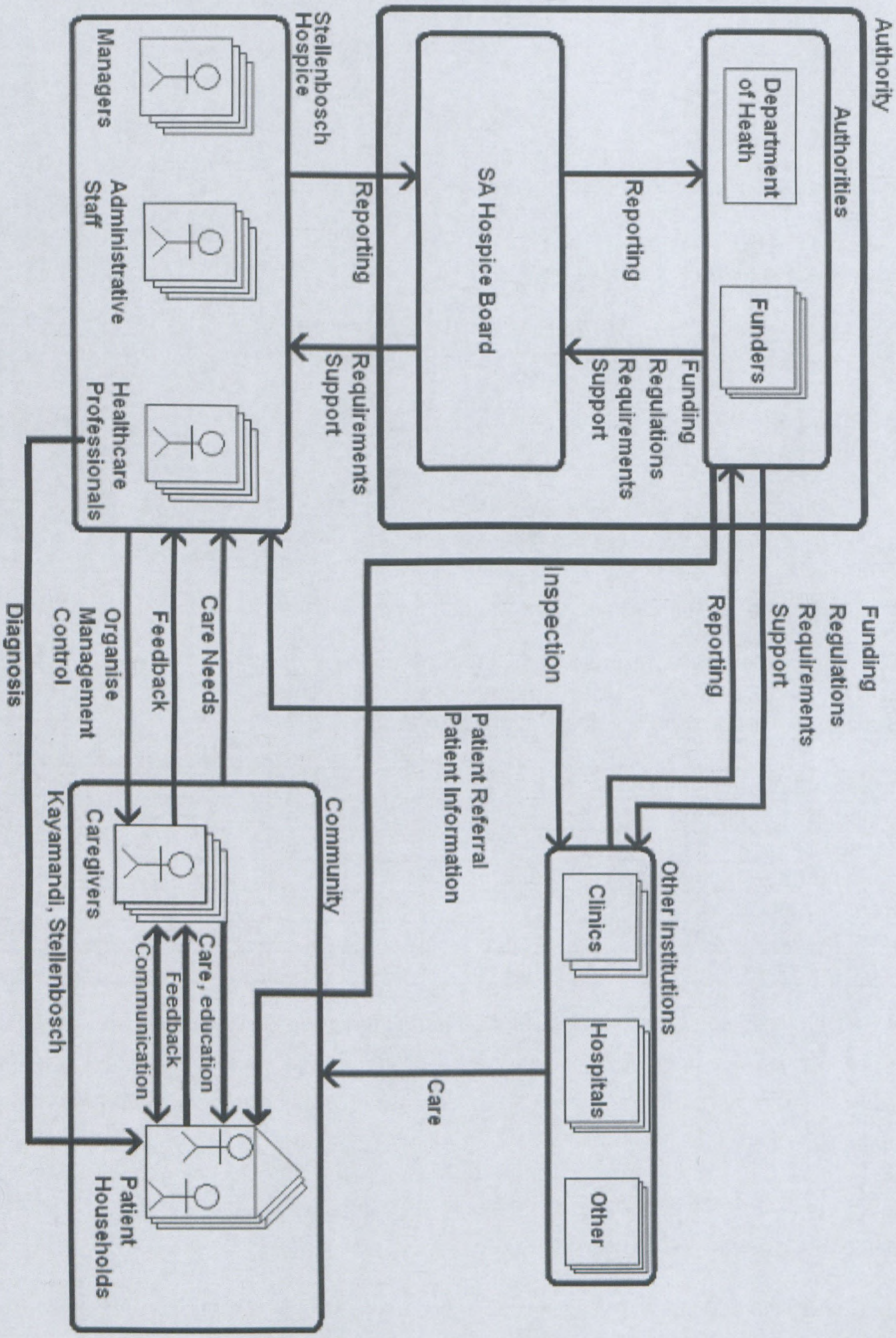


Figure 4.6: *Landscape Model for the Stellenbosch Hospice.*

Figure 4.6 shows the landscape model created during the ethnographic study for the Stellenbosch Hospice. The primary focus was on the Stellenbosch Hospice and on the Community, with limited focus on external governing bodies and external healthcare institutions.

The ethnographic study did have limitations, it was clear that there were several other facilities such as hospitals and clinics that also provide care services to the community, but these were located outside the community and were not considered in this study. From the ethnographic study it follows that several funding bodies existed, and that these funders provided not only funds but also required some form of feedback that the funds are being adequately utilised. Again this was outside the scope of this study with the focus on the care service provision from the Stellenbosch Hospice to the Kayamandi community.

The South African Department of Health does not directly interact with the Stellenbosch Hospice; it does however provide some funding to the SA Hospice Board, and thus indirectly to the Stellenbosch Hospice which is a subsidiary of the SA Hospice Board. The Department of Health does however shape the national healthcare policies, and thus indirectly how Stellenbosch Hospice should provide the care services. The Department of Health also requires statistical reports from the SA Hospice Board, which it generates from statistics gained from the numerous subsidiary hospices. The Department of Health furthermore also performs random inspections of the Kayamandi community to ensure that quality care is being provided. These inspections require that a paper document (or some other means of data storage) be present at the patient's home containing the information of the care that the patient has received.

The various other health care facilities, such as hospitals and clinics, are usually the first contact point between the community members and the healthcare system. A patient who is discharged from a hospital but who still needs some care will be referred to the Stellenbosch Hospice or a patient who was being treated by a clinic may also be referred to the Stellenbosch Hospice. The opposite is also true; if the Stellenbosch Hospice sees that a patient's condition worsens to beyond their ability to treat the patient they can refer the patient to a formal healthcare facility such as a hospital.

As seen in Table 4 there clearly exists a disproportionate number of patients to nurses (888 patients to 9 nurses), there simply are not enough nurses to provide the needed quantity and quality of care that the community demands. In order to help bridge this gap between supply and demand the Stellenbosch Hospice makes use of informal caregivers. The nurses' role for the most part is to provide support and oversight for the numerous caregivers, or lay workers, that are responsible for the bulk of the care provisioning. The nurse does the initial health assessment and compiles a care plan to be used by the care givers.

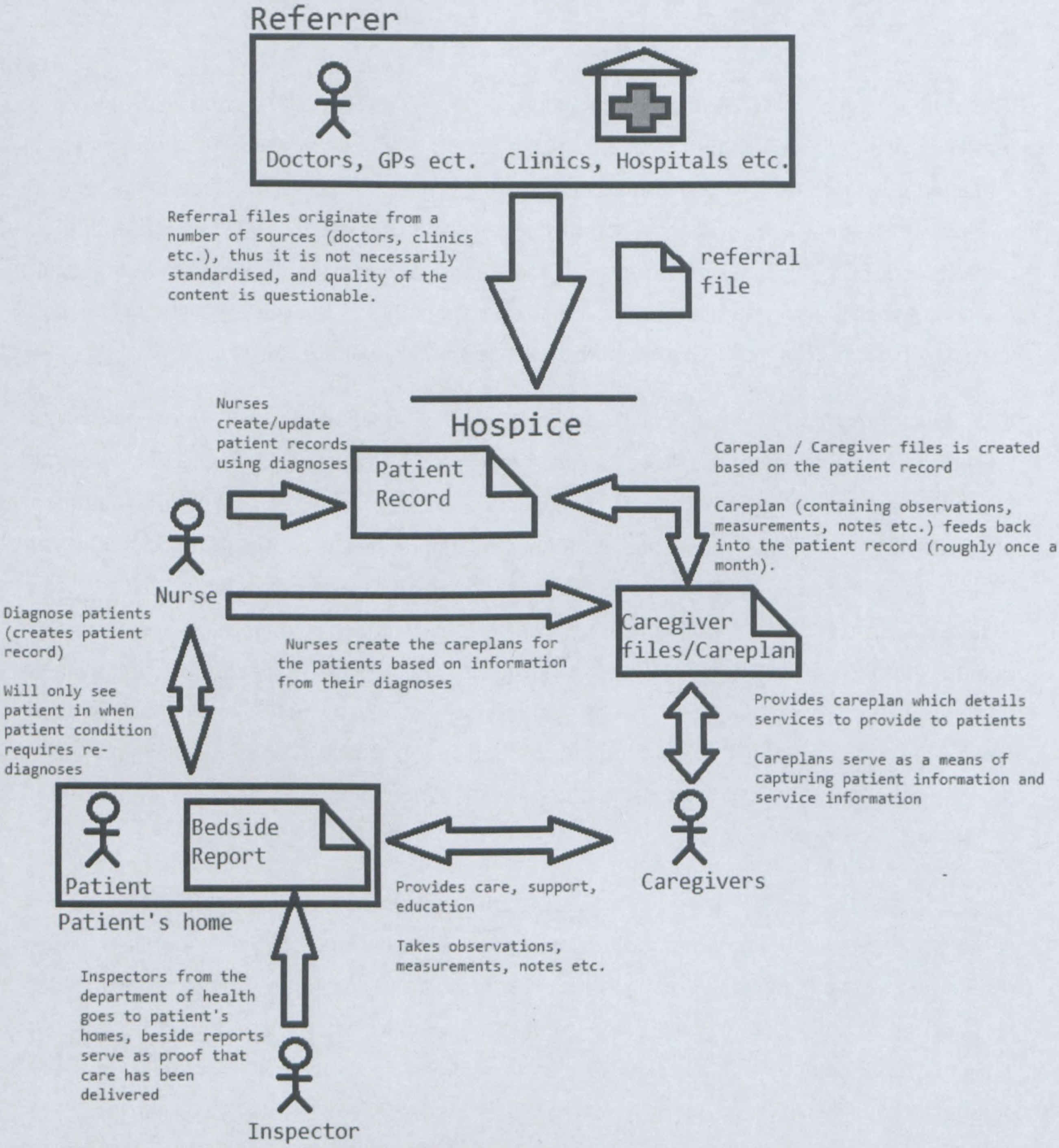


Figure 4.7: *the role of the Nurse in patient diagnosis and referral.*

The nurses are also responsible for the initial diagnoses of the patients; this is a legal requirement which semi-trained non-professional caregivers cannot meet. If a patient's condition worsens then it is also necessary for a nurse to re-diagnose the patient, and then possibly take additional action such as referring the patient to a formal institution such as a hospital. Figure 4.7 provides a more detailed representation of the referral and diagnosis process.

In the case of Stellenbosch there is a clear distinction between the facility (the Stellenbosch Hospice) and the Community. The Hospice is physically located outside the community it caters to as well the Hospice Board which provides the funding, support and requirements to the Stellenbosch Hospice. However the caregivers that work for the Stellenbosch Hospice themselves are volunteer members from the community and work as an intermediary between the facility and the community.

The caregivers are semi-trained volunteers, in Stellenbosch Hospice these caregivers are predominately female, most over the age of 30. Most of these caregivers come from the Kayamandi community providing an important link between the community and the facility. These caregivers aid in the provisioning of care but also play an important role in providing feedback to the facility about patient and the community in general. The caregivers use a number of paper documents, originating from the Stellenbosch Hospice, to provide this feedback. These paper forms are used to capture information about the patients, their living conditions and their immediate family. Once a month these paper forms are sent back to Stellenbosch Hospice, where the data is aggregated and reported to the SA Hospice Board.

There are four primary documents that are used by the Stellenbosch Hospice to capture and share data including: firstly the patient careplan, secondly the home file or bedside report and thirdly the patient record and finally the referral form.

The patient careplan is created by the nurses and defines the care and support activities that the caregiver is required to perform, but the careplan also acts as a feedback mechanism as notes, observation and measurements can be written down. The careplan is kept at the patient's home and provides the care giver with the guidelines of what care activities are required. These careplans are collected once a month and used to update the patient files at the Stellenbosch Hospice facility. A single careplan document can contain the details of multiple patients.

The Bedside report contains basic patient information and patient care information. The bedside report is used by a number of different parties for different purposes including: department of health inspectors to ensure that care is being delivered, other caregivers who are new or caring for a given patient can use the bedside report to "catch-up", family members can review the bedside report to see what care the patient is receiving and what care, if necessary, they can provide. The bedside report contains only the information of a single patient and the content of the report mirror some of the data contained in the careplan, as both are updated collectively while providing patient care.

The patient record contains detailed patient information for a single patient and for legal reasons is kept on site at the Stellenbosch Hospice facility. The patient record is used to create the careplans and in turn the careplans are used to update the patient records.

The referral form as is provided to the Stellenbosch Hospice by a referring body (a doctor, general practitioner, clinic, hospital etc.). The referral forms are not standardised meaning that different referral forms from different referrers for the same patient, contain similar information but presented differently and could even differ, e.g. different spellings for the patient name.

4.1.2.2 Motherwell

This section considers the Motherwell community and the Olive Leaf Foundation, which provides the Motherwell community with HBHC services.

The Stellenbosch Hospice is an example of a fairly well organised and structured HBHC initiative, as the Stellenbosch Hospice has the advantage of being part of a large network of hospices and being located in an area with a sufficient infrastructure (water, electricity, roads, cellphone reception etc.) to support the HBHC activities.

Motherwell is located in the Eastern Cape Province and therefore has a different context. Motherwell is also far more rural than Kayamandi and has even less access to necessary infrastructure.

Although the analysis and inquiry done on the Motherwell community is rather limited, the Motherwell community none the less provided a sufficiently different example of HBHC in rural South Africa and a good supplement to the ethnographic findings from Kayamandi.

Within Motherwell the Motherwell Health Clinic is responsible for delivering HBHC services and is primarily funded and supported by the Olive Leaf Foundation (OLF). OLF is a NGO which is primarily focused on supporting and aiding development initiatives within the Motherwell Community. The Health Clinic is not directly managed by the OLF but a memorandum of understanding exists between them stating both responsibilities and requirements towards each other.

The figure 4.8 shows the landscape of the Motherwell community, showing the data-flow interactions and relations between the various stakeholders in the HBHC service delivery of Motherwell.

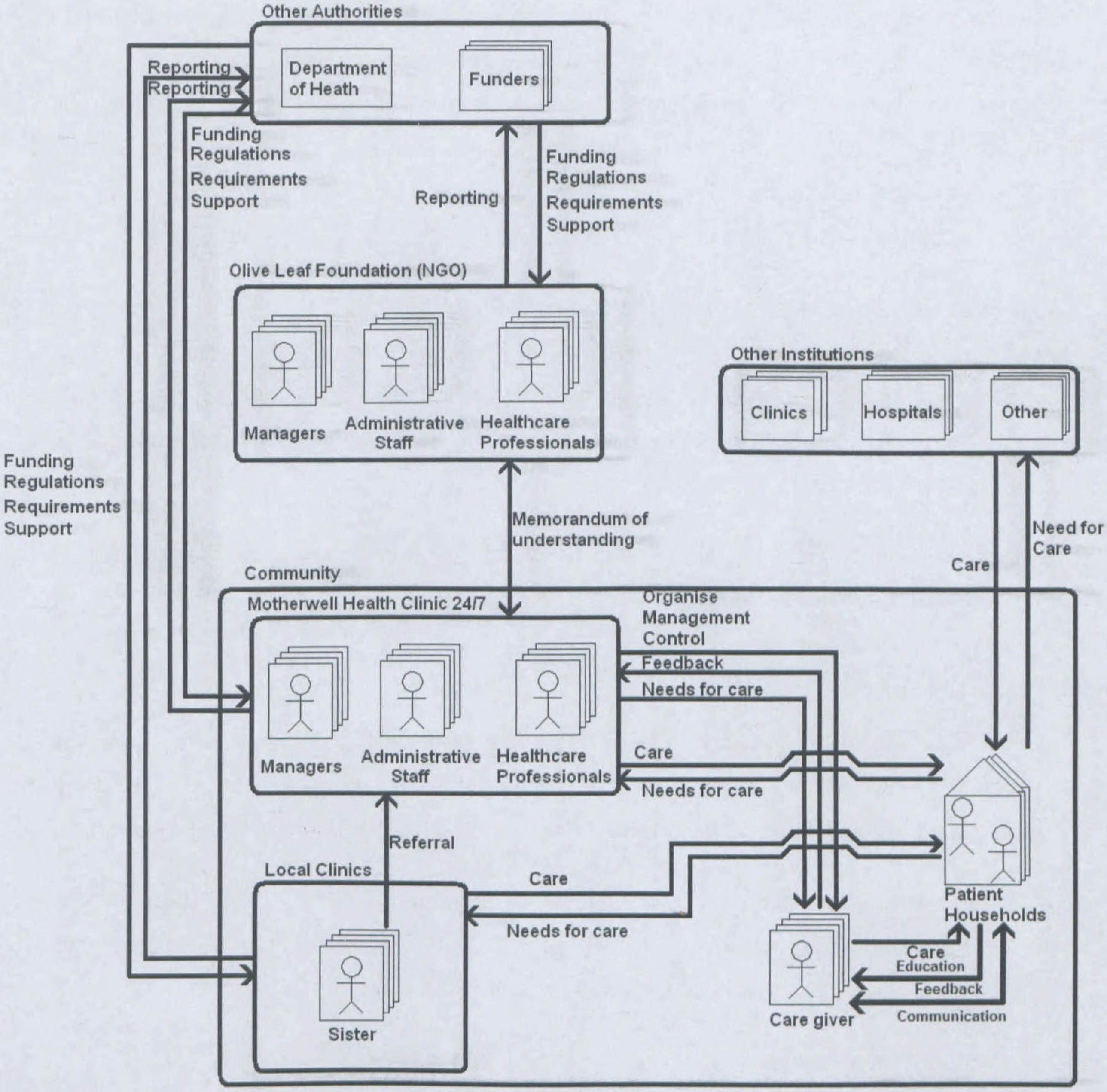


Figure 4.8, *Landscape Model for Motherwell.*

The number of semi-trained caregivers to healthcare professionals such as nurses is extremely disproportionate in the case of Motherwell. Where in Stellenbosch a nurse oversees the caregivers, in the Motherwell Community the caregivers are overseen and managed by yet another volunteer from the community. Because of the lack of available formally trained healthcare professionals, the caregivers who are responsible for providing the home care service are also required to take detailed notes and fill in comprehensive forms. The reason for why caregivers need to take these comprehensive notes is because they only semi-trained and thus do not have the training or legal-ability to make diagnoses. The detailed information they capture is necessary for the nurses who cannot also physically diagnose the patients to still be able to monitor the patients remotely.

A number of other facilities exist outside the community which also cater for the community healthcare needs, as well as those of several other communities. Because a number of local clinics and the Health Clinic exist within the community itself, the community members are far more likely to seek care from these facilities. There is thus a strong link between these local facilities and the Health Clinic, with patients usually being referred to the Health Clinic by these local healthcare facilities.

Because the Health Clinic is not a direct subsidiary of the OLF, the Health Clinic directly reports to the numerous authorities, such as funders or government bodies such as the Department of Health. Because the OLF itself has a number of funding and authoritative bodies, these also indirectly influence the Health Clinic via the memorandum of understanding between the Health Clinic and the OLF.

The following Section will compare the two HBHC contexts to try and identify commonalities and differences between them.

4.1.2 HBHC Commonalities and Differences

Both HBHC initiatives discussed in the previous section have a number of commonalities, since both are HBHC initiatives and both deliver home-based care to rural and poor communities in need. These communities, however, also have a number of differences.

This section lists some of the more notable of the commonalities and differences between HBHC in Kayamandi and Motherwell, but no attempt is made to explain these commonalities in any detail, since to do so effectively would be a research undertaking in itself.

The commonalities include:

- Catering for the needs of rural and poor communities.
- A low penetration of IT-based solutions and corresponding prevalence of paper-based systems, most likely because of cost issues or lack of available or quality supporting infrastructure needed for IT-based solutions.
- Majority of the caregivers are volunteers, semi-trained members from the same communities they care for.
- The HBHC initiatives provide a support and management role for the caregivers.
- HBHC initiatives work closely with other institutions (hospitals, clinics) to provide care.
- Caregivers are supervised by trained healthcare professionals.
- Both suffer from a lack of resources including human, funding and infrastructure to provide a quality of care.

- HBHC initiatives are closely linked to the communities and attempt to provide care that meets the specific needs of the community they cater for.
- Both have funders or supporting organisations that require specific feedback or help to influence how they perform their specific functions.

The differences include:

- In Kayamandi there are no healthcare institutes inside the community, while in Motherwell there is a community clinic physically located in the community and strongly linked to the community.
- Although caregivers are supervised by healthcare professionals, in Motherwell a healthcare professional has to supervise far more caregivers then in Kayamandi and also has to do this remotely, i.e., the healthcare professionals do not visit the homes of the patients and have to rely on the information provided by the caregivers to decide on the care needs.
- Kayamandi and Motherwell are geographically far apart, Kayamandi being in the Western Cape Province while Motherwell is in the Eastern Cape Province.
- The two communities have differing cultures with Motherwell primarily consisting of Xhosa speaking and African members, while Kayamandi consists of primarily Afrikaans and English speaking Coloured and Xhosa speaking African members.

The next section will discuss the paper based systems used by the Stellenbosch Hospice and the Motherwell Health Clinic.

4.2 Current Systems

This section considers the current systems that are used in HBHC. The first section considers the paper based systems that are predominately used before drilling down and detailing the various data-elements. The issues with the current paper-based system are detailed in the following section, in light of these issues the importance of addressing the identified problems shown.

4.2.1 Paper Systems

The previously Section 4.1 considered the HBHC context, specifically that of Kayamandi and Motherwell. It helped to define and identify the characteristics of the environment of the intended research.

This section considers the paper-based systems used to capture and manage data used by both the Stellenbosch Hospice and the Motherwell Olive Leaf Health Clinic. The paper-based systems play a key part in any attempt to understanding how these HBHC initiatives function. Because of the low penetration of IT based solution within rural HBHC these paper systems are the primary means of data collection, storage and processing. The paper-based systems can thus give a good indication of which data is collected, how the data is structured and what meaning is assigned to data. The patient careplans are also captured in these paper documents and thus form an important communication tool between the institution and the caregiver by acting as a tool for capturing caregiver observation and patient measurements for the institution's consumption. The careplan also helps define the caregiver's interaction with the patients by defining the care activities.

The ethnographic study found that very little IT-based solutions were implemented and utilised in terms of data management and acquisition within the HBHC initiatives in Motherwell and Kayamandi. IT-based systems that were used were spread-sheet based and used as a storage means or for generating reporting statistics. In most cases of the HBHC the paper-based systems were used as the primary means of capturing, transferring, processing and storing information.

The primary means for recording data across different HBHC initiatives was the usage of paper-based forms. In situations where some form of IT-based solution was being utilised it was kept to a minimum, in most cases only serving as an electronic backup of the paper documents. Stellenbosch Hospice used a number of paper forms such as a patient care plan, home-patient record, and home visit form and tally sheets. The patient careplan contains the various care activities and goals that the caregivers must undertake for a given patient, and contains fields for the caregiver to write notes and observations. The home visit form is used to keep track of which patient was visited, when, where and for how long. Patients are required to sign the patient visit form or their family members if the patient is not physically able to so. The tally form is ideally filled in at the end of each day, tallying the total number of patients visited and the number of specific care activities performed and this form is used to report the care visits.

In the case of Stellenbosch Hospice these paper forms, specifically the individual patient tally forms, are sent to an area coordinator to be used to create area specific totals which is in turn captured into Microsoft Excel spread-sheets. A different person plays a role at each step namely: the caregiver capturing and tallying their specific patient information, the area coordinator tallying the totals of the different caregivers of the different areas and finally the data-capturer who captures the data from the area coordinator into electronic spread-sheets. The spread-sheets however do not feed into any other system or serve any real purpose other than the creation and storage of aggregated statistical data.

In the case of Motherwell the forms were notably more detailed and comprehensive, because of the limited number of available nurses to regularly visit and diagnose the patients. The forms are required to capture more detailed information to enable the nurses to track the patients' conditions and take action accordingly, such as visiting the patient to check on their condition or referring a patient to a formal healthcare institution.

A number of issues were identified with the usage of these paper-based forms.

These forms were filled in by hand by a caregiver and interpreted by another who would use these forms to generate summary or statistical data. For most caregivers the care process took precedence over filling in forms, which led to most forms either being quickly filled in or filled in at a later time, meaning that interpreting the caregivers handwriting could prove error prone or the forms could be incomplete.

Because these forms are filled in by hand by the caregivers during the care process (while they are providing a care service to the patient), sometimes the text would be illegible or details would be left out because the caregiver were in a hurry, or more focused on the patients. It is also possible that a caregiver would not fill the forms in a timely manner and would put off filling in the forms until the end of the day or end of the week in which case the caregiver may have forgotten some of the details.

The forms are also subjected to the wear and tear of constant use. Caregivers would walk from house to house with the forms in their hands; these forms are then exposed to the weather conditions such as rain. The paper forms could also be damaged with repeated use or could easily get lost or stolen – this is an important ethical issue since confidential patient information can fall into the wrong hands.

After examining the forms from the different initiatives more closely, a number of other issues were noted. These issues included unnecessary repetition of information that needed to be filled in as well as some inconsistencies between similar fields spread across related care documents. Furthermore most fields were not accompanied by descriptions or notes. This along with seemingly similar wording and semantically similar terms being used opened the possibility for different interpretations to exist for some elements.

By examining the paper forms it became clear that the caregivers were required to fill in the same data in multiple locations on the same form. Most of the data required on the paper forms were narrative in nature but from the possible values were fairly repetitive, or 'standard'. Because the narrative values of certain fields will always be fairly similar the narrative field could be easily replaced by a list of possible values.

Furthermore when comparing the forms from Kayamandi and Motherwell it was found that the forms not only differed in complexity and the degree of comprehensiveness but also in their usage of data and definition of data itself. Paper forms from different initiatives intended for the same purpose used different names or completely different data.

This section looked at the paper documents used by Stellenbosch Hospice and Motherwell Health Clinic, the following section follows from the above by looking specifically at the data-elements used on the forms.

4.2.2 Data-elements

By looking at the paper forms it became quite clear that similar but different ontologies were used by different HBHC initiatives. What was constituted as the core data-elements and how the data-elements related to each other was also different form one initiative to the other. It is difficult to determine to what degree these data-elements differ.

The data-elements are considered in more detail in order to establish how the topic of data-elements relates to the intended solution artefact developed as part of the research.

Understanding how a HBHC initiative defines and identifies the different data-elements and the relationships amongst these data-elements within a HBHC initiative was a difficult task.

Some of the reasons for the difficulty in understanding the data-elements were: the lack of enforced standards for data-elements in HBHC since government required only statistical data but did not define or enforce standards for data storage or representation in HBHC; the dominant usage of paper-based systems; the lack of IT and IS-based systems; and a lack of formalised standard definitions of data and data-element by these HBHC initiatives themselves.

There was also a notable difference in the utilisation of data-elements by the HBHC initiatives. The reason for the different utilisations originates from the different data requirements amongst the different HBHC initiatives. In the Western Cape where the nurse does the assessment and diagnoses of the patients, the caregivers do not need to record assessment data yet again while providing care for the patients, although they do need to take note of new or worsening symptoms. In the rural Eastern Cape there is more of a need to take detailed notes of symptoms and conditions, in order to provide a rigorous means of ensuring that new health problems or complications do not appear.

It was also noted that data-elements were inconsistently referred to in different forms and across different initiatives. Although initially perceived as not such a large issue, it became more important as attempts were made to gain a collective understanding of the data usage across HBHC initiatives. Inconsistent references to the same or related data-elements occasionally led to misconceptions and incorrect assumptions being made.

The following section takes the data from the previous sections and attempts to define and identify the problems.

4.2.3 Identified Problems

This section uses the ethnographic data to identify some of the problems in HBHC, specifically problems the research attempts to address.

Part of the DSR methodology used in the research involves the identification, clarification and motivation of the research problem. This section attempts to use the ethnographic study to better identify the problems faced by HBHC in order to later, in the following section, motivate the need for the research and solution. Ideally the problem identification would also involve a component of literature review, although the literature relating to HBHC and the issues faced in HBHC were already touched on in the previous chapters (Chapters 2 and Chapter 3).

A number of problems and issues were identified in the ethnographic study which prevented the effective and efficient delivery of HBHC services in the examined communities of Kayamandi and Motherwell.

In order to identify the research problem it is important to keep in mind that when focusing on HBHC that by its nature HBHC attempts to overcome many of the environmental issues that prevent the delivery of more formal forms of healthcare. Thus when looking at HBHC particularly in rural and poor communities it is easy to mistake these environmental issues for formal healthcare as HBHC issues as well. For example poor road infrastructure and lack of private or public transport prevent patients from going to formal healthcare institutions, but HBHC delivers care at the homes of the patients overcoming this issue. Thus not all environmental issues are relevant and issues that prevent formal healthcare are not necessarily relevant to HBHC.

The problems that were identified using the ethnographic study that related to HBHC in general were grouped into two categories: macro-level encompassing governmental and institutional and micro-level encompassing institutional, caregivers and community.

The problems identified on the governmental and institutional (macro) level includes: a lack of sufficient direct government involvement and support in term of HBHC; lack of regulations and standards regarding the structure, definition and usage of data in HBHC; limited universal understanding of ground-level activities in HBHC which limits the success-rate of future projects and attempts to improve some aspect of HBHC and limited available support for the HBHC in terms of resources and funding.

Some of the problems which are experienced on the level of the HBHC initiatives (institution) and the caregivers, the micro-level, include: issues with collecting and using quality, reliable information; low penetration and usage of IT-based solutions; lack of sufficient infrastructure to support care activities or IT based solution; a lack of enough trained medical professionals (such as nurses) to act in a supervisory role; history of low success of external parties efforts to research/improve some aspect of the institution/caregiver/community; different meanings assigned to the same data-elements amongst caregivers and the various issues with the paper based forms have already been listed above.

Two related problems are addressed during the research, the solution problem is addressed using design and development based research which creates the research data in order to address the research problem.

When following a DSR approach when conducting research it is necessary that a problem is identified which can be addressed using a design-based approach. The reason for motivating the usage of DSR approach is that it provides additional validity to the research and the research findings.

The following section motivates the previously mentioned problems and the need for the research.

4.2.3 Problem Relevance and Motivation

When following a DSR approach in conduct research it is necessary that a problem is identified which can be addressed using a design-based approach. The reason for motivating the usage of DSR approach is that it provides additional validity to the research and the research findings.

The motivation of the problem is in part a literature review and part research, starting with an initial literature review followed by initial research to better ground and motivate the problem, supplemented with a literature where needed. This section will look at the motivation of the research problem and the solution problem using the ethnographic study as the literature components of the motivation has already been covered in Chapter 2.

All three of the initial conceptualized levels of HBHC (community, institution and governmental/management) ideally require comprehensive, timely and quality information to perform their roles effectively and efficiently. Although as seen from the ethnographic research there is no enforced standards from any of the governmental bodies to define what constitutes comprehensive, timely and quality data within the HBHC context. This means that at best the government responds to health situations based-on data of poor quality. But also important is that currently the government does not have access to individual patient data that could provide valuable input in the status of health at the community level. A better understanding of how data-elements are used in care activities outside the formal healthcare services can assist in defining an appropriate health record that could become the single electronic record for each person to be used by any healthcare facility.

Because of this lack of common understanding and standardisation that exists around the care data-elements it is difficult for HBHC initiatives to provide comprehensive information to the government in order to enable effective government action and intervention. This lack of standards also means that different healthcare institutes have difficulty sharing patient and practice information. So although it is fairly easy to refer a patient between institutions it is however to effectively share the patients data and information between these institutions because there is no common form or standard for which data needs to be shared and in what format.

Although it is possible for governmental bodies to create standards for data within HBHC, to do so effectively they would require a comprehensive, holistic understanding of data usage and definitions within HBHC. With the variety of differences that exists amongst HBHC initiatives this would be a daunting undertaking.

The development component of the research considers developing a simple tool for trying to capture and understand these HBHC data-elements and their relations. The literature noted issues around standardisation and comprehension of data-elements as well as pointing out some of the issues around standardising data and the benefit of being able to do so. The ethnographic research supported the literature findings by showing that there was currently no standard for HBHC data as well as differing usages and understandings of seemingly similar data.

A comprehensive system that can be used to understand the data-elements used in HBHC can potentially be of use and benefit to a number of primary and secondary parties. There are however several limitations namely: the system that was developed was only a first iteration of development, there exists clear issues in terms of software and research such as showing research rigour. Because of these limitations the software tool/solution itself could not constitute research, but by following a DSR approach the artefacts created during the design and development process could however serve as research tools.

During the literature it has become clear that HBHC particularly in rural areas has a low penetration of IT-based solutions and a wariness of new IT-based solutions amongst HBHC initiatives. The literature furthermore noted that rural HBHC faced numerous issues when trying to deliver care but also pointed to the various advantages which health informatics (HI) brings to the healthcare environment. The literature pointed to the possibility of HI being able to address or at least limit these issues faced by HI. But with the low penetration of IT and the low success rate of HI based solution in HBHC there was clearly a need from the literature to look at the design and development of solution in HBHC in order to increase the success-rate of HI solution in the rural HBHC context.

The initial ethnographic study also showed the concern where outside parties promised to solve some issues faced by individuals in HBHC and then failed to do so. The limited usage of IT and HI solutions within rural HBHC was also apparent from the ethnographic study but furthermore the ethnographic study showed some of the causal environmental factors, such as limited funding, limited supporting infrastructure and limited IT-based skills. The ethnographic study thus also supported the need for research into the design of an IT-based solution, which meets the requirements of the HBHC individuals and recognises the environmental factors in order to ensure the success of the intended IT-based solution.

4.3 Research Problem Revisited

This section will consider the problem being addressed by the research and the underlying development effort.

This research follows a DSR methodology, which as discussed in the Chapter 3 involves the creation of research artefacts with the intent that these artefacts can act as both a means for solving a given practical problem, but also can serve as tool for conducting the research.

The development effort, which produced the research artefacts, involved the creation of a semantic metadata repository with the intent of it being used as a simple tool for understanding the different usage of data amongst the various initiatives that attempt to provide HBHC services to communities in need. The development was undertaken as part of a project conducted at CPUT in Cape Town in the Western Cape of South Africa. The project's overall focus was to consider how ICT could be used to facilitate the addressing of stress in rural and poor communities with the specific focus being on HBHC within these poor and rural communities. By looking at the paper systems used by the different HBHC initiatives it became clear that there was no single standard HBHC paper system. By looking at the actual data fields on these forms it became apparent that the seemingly similar data-elements were being utilised differently and different measurements were used for similar things. There is therefore a need for a semantic metadata repository to capture the different uses of data-elements and to present the data-elements and their uses in a simple and easy to understand format. Such a repository can then be used by other ICT solutions to assist with the standardising of data-elements and their use in HBHC. There is a need for more ICT solutions in HBHC although the current lack of resources and more specifically funding remain a problem.

A number of reasons existed for the low IT penetration as detailed in the literature review: an overall lack of funds; a misconception of the advantages and uses of IT; low literacy rates and thus a fear of IT solutions amongst caregivers; limited infrastructure (notably in Motherwell where a stable supply of electricity is not always guaranteed); and previous issues and problems with similar solutions.

A large strain is placed on the HBHC initiatives; these initiatives work in an environment which does not always have the necessary resources in terms of money, equipment, facilities or personnel thus spending money on implementing new IT-based solutions seen as being important.

Kayamandi also has a number of stand-alone, none connected systems which perform similar functions but do not have the innate ability to transfer data between them. An additional weariness from the Stellenbosch Hospice also results from the fact that they perceive that any new IT-based solution would just add more work for them to do, as has been the case so far.

IS and IT artefacts and instantiations however do not function in a vacuum, the performance of an IT-based solution is related to the environment that it operates in. This is because the environment surrounding the solution defines how the solution will be used, by whom the solution will be used, where the solution will be used and the constraints under which the solution must function. Thus a clear understanding of the environment and the context of the intended solution is important for successful development of intended artefacts (March, Smith, 1995).

The context thus also provides the problem domain, or the opportunity domain depending on one's perspective, which gave rise to the need or desire to develop a solution in the first place. Furthermore beyond identifying the problem and the problem domain looking at the context provides some of the underlying justification and motivation for creating the solution.

Conceptualising the context not only provides important clarification around the solution artefact produce by the development but also provides important clarification for the intended research. As the research considers design considerations first, understanding the context in which a solution must function also helps to inform a researcher or developer about the issues and considerations that need to be taken into account such as the constraints or liberties. It would be unwise to claim that all HBHC initiatives within the entirety South Africa have the same characteristics and present exactly the same constraints and issues. It is still possible thou to generalise the findings from one or two contexts to such a degree that one could generate a list of possible issues or considerations that a researcher or developer would need to keep in mind while working within a given a South African HBHC environment. This, however, falls outside the scope of this study.

Chapter 5 Objectives

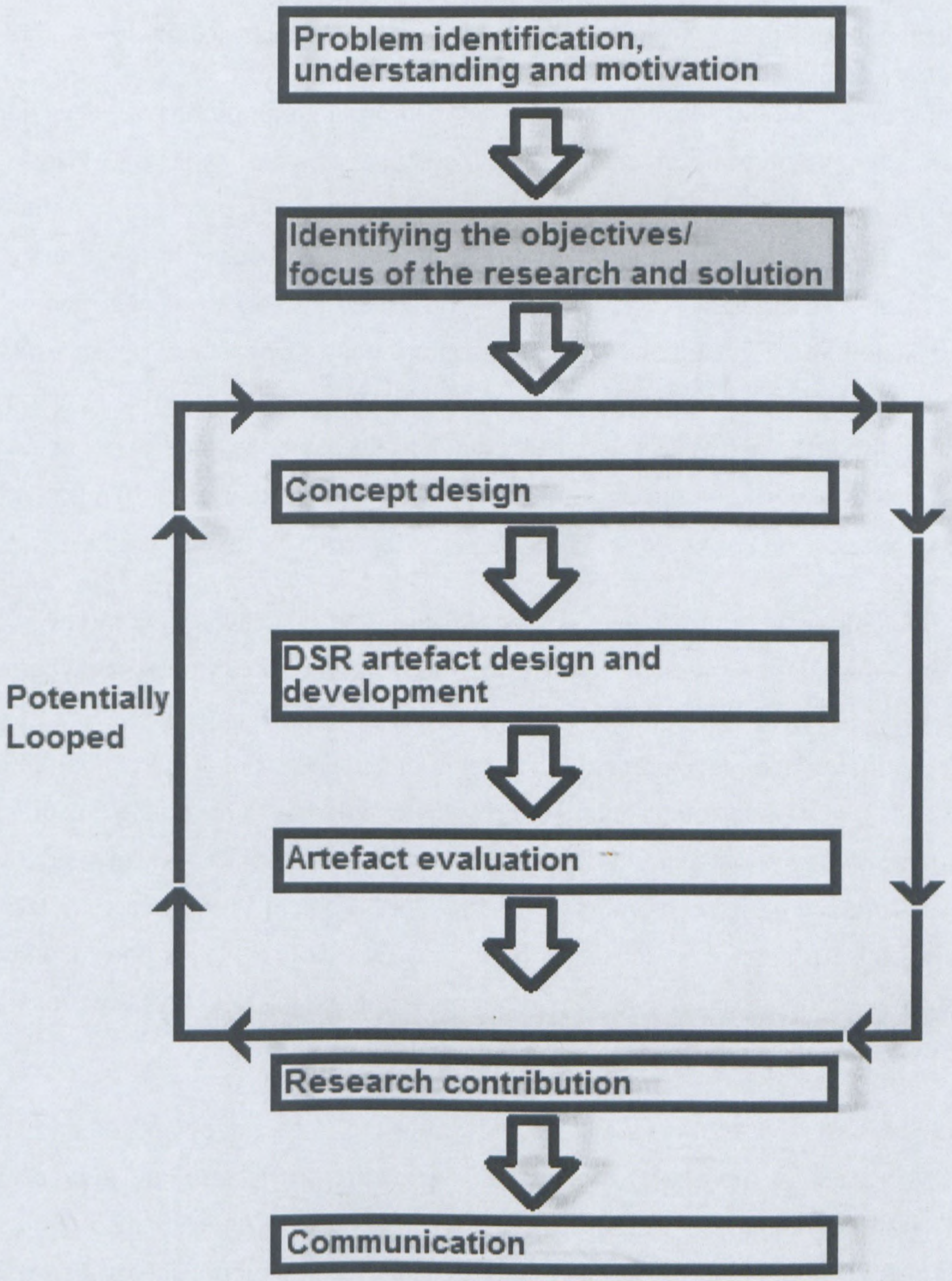


Figure 5.1, *Focus of chapter 5 in relation to the methodology.*

This chapter represents the second phase in the chosen research methodology and takes a closer look at the objectives of the research and the solution.

This chapter logically follows on from the previous chapter which looked at the identification and conceptualisation of the research problem relying primarily on ethnographic data. The ethnographic data provided the understanding of the research problem and the research context from which the objectives, which will be detailed in this chapter, originates.

Identifying the objectives of the solution artefact forms an important part of the overall design and development process involved in creating any particular type of solution. Detailed in chapter 3 under the heading of Artefacts in Design Science Research a number of artefacts are created during DSR to act as research tools to examine some aspect of the research area/problem. These DSR artefacts could potentially be created solely to act as research tools with no additional intent or purpose beyond research. It is however also possible that the DSR artefacts are created as part of the design and development of software intended to be used apart from the research. This is especially true if the research is undertaken as part of a pre-existing software development process, the produced artefacts would then be both DSR artefacts intended for research purposes as well as artefacts intended for practical use.

The objectives originate within the problem domain since these objectives define how and which specific issues within the problem domain need to be addressed. However research for the most part is a knowledge creation process with the primary objectives being the creation of knowledge or the development of insight. In contrast the development of a solution is a process that produces tools intended to achieve a number of possible outcomes beyond creating knowledge and insight. In DSR where solution artefacts are developed it is possible for these artefacts to address a different, if related, problem to the research itself. Thus it is possible for two sets of objectives to be in play, one set relating to the research and guiding the research undertaking, while the other set relates to the development process and guides the development of the solution.

The simplest form of DSR would thus ideally be one where the artefacts' objectives and the research objectives are near identical, for example the research considers how a problem can be solved and the artefacts are created with the purpose of solving them. Although conceiving that all DSR is limited to only this "simplistic" form with its identical solution and research objectives then it limits the possible range of DSR application. During conventional design and development efforts a number of related but different issues are touched on. Each of these issues potentially offers an opportunity for research but all these issues are collectively too complex to be addressed in a single research undertaking.

However this research attempts to pull these two objectives together as much as possible to simplify the overall research and the research process for fear that additional complexity would increase the likelihood of errors and inconsistencies from arising.

The first section within this chapter considers the objectives of the designed and developed solution; the second section focuses on the research objectives and attempts to unify these with the development objectives. The final section of this chapter discusses these two sets of objectives and the justification for the research.

5.1 Objectives of the design and development

This section will look at the objectives of the solution artefacts in order to provide a better understanding of the underlying design and development process which was used to generate the research artefacts which in turn were used to identify the research findings and contributions.

Within the process of software development the identification of the objectives forms an important part of the overall process of creating working software. Explicitly identifying the objectives does more than simply dictate what a solution should and implicitly what a solution should not be able to do. To meet these objectives it is necessary to define in detail the required functionality, the overall characteristics of the solution and to take into consideration the limitations and advantages of the intended context of use in which the solution must function. Specifying the objectives thus provides focus and guides additional analysis, design and development efforts which are undertaken in order to achieve the specified objectives.

It is the role of both the developer and the researcher (which in this instance is one and the same) to identify these objectives and to identify how these objectives can be met. While identifying the' how' it is necessary to keep in consideration the limitations and advantages of the context in which the solution will be used. In the case of this research the context is defined using the literature review and the ethnographic data.

These objectives can be co-created with the aid of the target users, with other stakeholders or by looking at the environment or the context. In the case of this research the objectives were created by combining the findings of the literature review and the ethnographic study, detailed in the previous chapter.

The design and development objectives for the underlying design and development originated in the identified problems specifically: the lack of standards and regulations relating to the definition; the usage and the structure of data within HBHC; issues surrounding the understanding of data-elements amongst caregivers; different usages of similar data and issues surrounding data quality on paper based systems.

The specific objectives of the solution artefact are:

- To be used as an informative tool, to help understand the data-elements being used in HBHC.

  Ideally this objective would be achieved by capturing the data and the metadata (structural data and semantics) and presenting it to users in a way that is easy to interpret (such as using diagrams or other visual means).

- To be able to capture and archive the different data-elements and metadata-elements used by the different HBHC initiatives. The system will need to keep record of which data-elements belong to which HBHC initiative.

- The solution should be able to store the source material of different types such as images, audio, video etc. from which the information has been derived, as an additional means of ensuring that the assumptions made during the process of identifying and capturing the source material into the solution.

- There is no initial need to make the solution usable by the caregivers themselves, and thus a technically minded individual is expected to be able to use the solution artefact. This is done to prevent major focus on the human-computer component of development, which does not form an important part of the research.

Most of these objectives are intended to be refined in additional iterations of the design and development process. The research involved only the first iteration of the design and development, meaning that during this early iteration most of these objectives have not yet fully been achieved or refined, but these higher-level objectives do provide clear requirements and guidance for the design and development process.

The following section considers the objectives of the research and attempts to align the research and development objectives.

5.2 Objectives of the Research

The objective of the research has previously been discussed in Chapter 3 Section 3.4. This section attempts to join the development and research objectives.

The development objectives are technologically focused, with the primary purposing relating to what the system is intended to accomplish, or what a person using the system is intended to be able to accomplish. Of course not all these objectives are purely technical, such as attempting to aid in comprehension. The research objectives however are more focused toward understanding and explaining, and thus aim to guide how this understanding and explanation can come about.

So where the development objectives attempt to guide the creation of the solution artefact, the research objectives attempt to understand what is required to create the solution artefact (which is in part heavily dependent on the practical development and objectives). In addition is it necessary to not just understand these requirements but in light of the research data relating to the context (the ethnographic data) to explain the why of these requirements of the solution artefact produced by the research.

By gaining this fundamental understanding of the "what" and "why" allows for a prescriptive knowledge to be developed. A further objective of the research is thus to provide some framework or guidelines by which further developments of similar solution artefacts (or artefacts in a similar context) can be undertaken with a potentially higher level of success.

The following section considers and discusses briefly the two separate sets of objectives at play during the research.

5.3 Separate solution and Research objectives

This section discusses the two sets of objectives identified previously. Although in the previous example the objectives are tightly coupled, and distinction between 'research' and 'development' objectives have little actual practical implication on the overall research, it is still of some interest to note and discuss.

At the start of the research only a single set of objectives was in place. As the research got underway it became clear that the research objectives which looked at the design consideration were not sufficient on their own. Further there existed clear objectives for the design and development (the 'carpentry') component of the research. It seemed clear that both practical and research objectives influences each other and both play an important role. Going back to the sourced articles it became clear that most of these articles argued for the utilisation of DSR on a fairly abstract intellectual level, but few gave practical examples of DSR being applied to more than simple research projects.

It was clear that although having two sets of objectives was not necessarily a characteristic of all design science-based research, it was necessary within this particular research to define and detail both set of potential objectives. Because of the nature of the research there were two distinct parts: the development and the research and although the development could simply have been treated like a research means or a method to create research tools, the development and the artefacts produced had goals beyond the research.

An additional reason for making a clear distinction between the design and development (solution) and the research and having both these aspects of the research having their own focus and objectives was because of the difficulty in proofing research rigour within the software solution. To prove the rigour and thus research contribution of the software solution it would most likely involve having to take the solution to industry and have it tested over a period of time in a practical setting, since the solution's relevance is difficult to show within a clinical 'laboratory' setting. Because of the scope of proving solution rigour and relevance it was decided rather to focus on a subset of the overall development, specifically looking at the design considerations.

Having two sets of objectives did thus prove useful. Alternatively if the solution's objectives were the same as that of the research, for example the research looked at the means to solving a research problem and the design and development looked at the tool that solved the research problem. This opens the possibility that the design and development or even the research might be influenced by the researcher's research of development biases. If DSR artefacts are created solely to generate research findings, then the researcher can create artefacts, knowingly or unknowingly, which are then inadequate in meeting the research objectives. These inadequate artefacts do not have any external criteria or objectives to guide their design and development and thus might not represent reality or address actual problems in an adequate manner. If the artefacts, the primary tool for producing research findings in DSR, do not represent reality then the findings produced by them might be questionable or completely misrepresentative of reality.

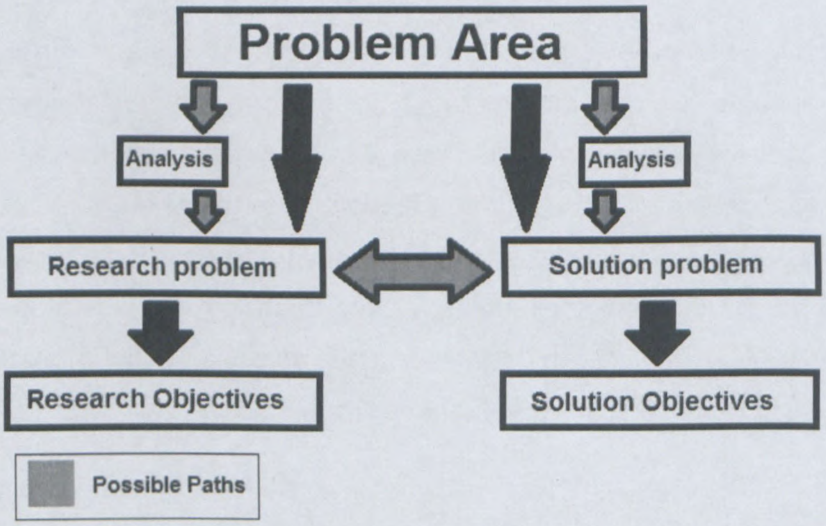


Figure 5.2: *Conceptualisation of the process from problem to objective.*

The Figure 5.2 shows a possible means for the development of objectives within DSR. The Objectives originate from the problem which in turn originates out of the problem area or the research context. It is possible for the problem to come directly out of the problem area, or as is the case within this research, that a certain amount of initial exploratory research (possibly involving a literature review) needs to be conducted to be able to identify the problem within the problem area. The Research problem and the Solution problem do not both have to spring directly from the problem area as one can create the other. For example if a software project is underway then research can spring from a given area of the software project and the opposite is also true, a given research undertaking can give rise to the creation of a software solution. In a case where one of the problems gives rise to the other problem it is still important to establish some link between the derived problem and the problem area, or risk the research falling prey to the issue identified previously where the design and development produced inadequate artefacts.

This chapter looked at the objectives of the solution artefact and very lightly touched on the topic of the research objectives since these have already been discussed in a previous chapter. An argument was also made for why two sets of objectives were necessary within this particular research. Although only the research objectives were of importance to the research, the solution objectives played an important part in ensuring the validity of the overall research.

Chapter 6 Design and Development

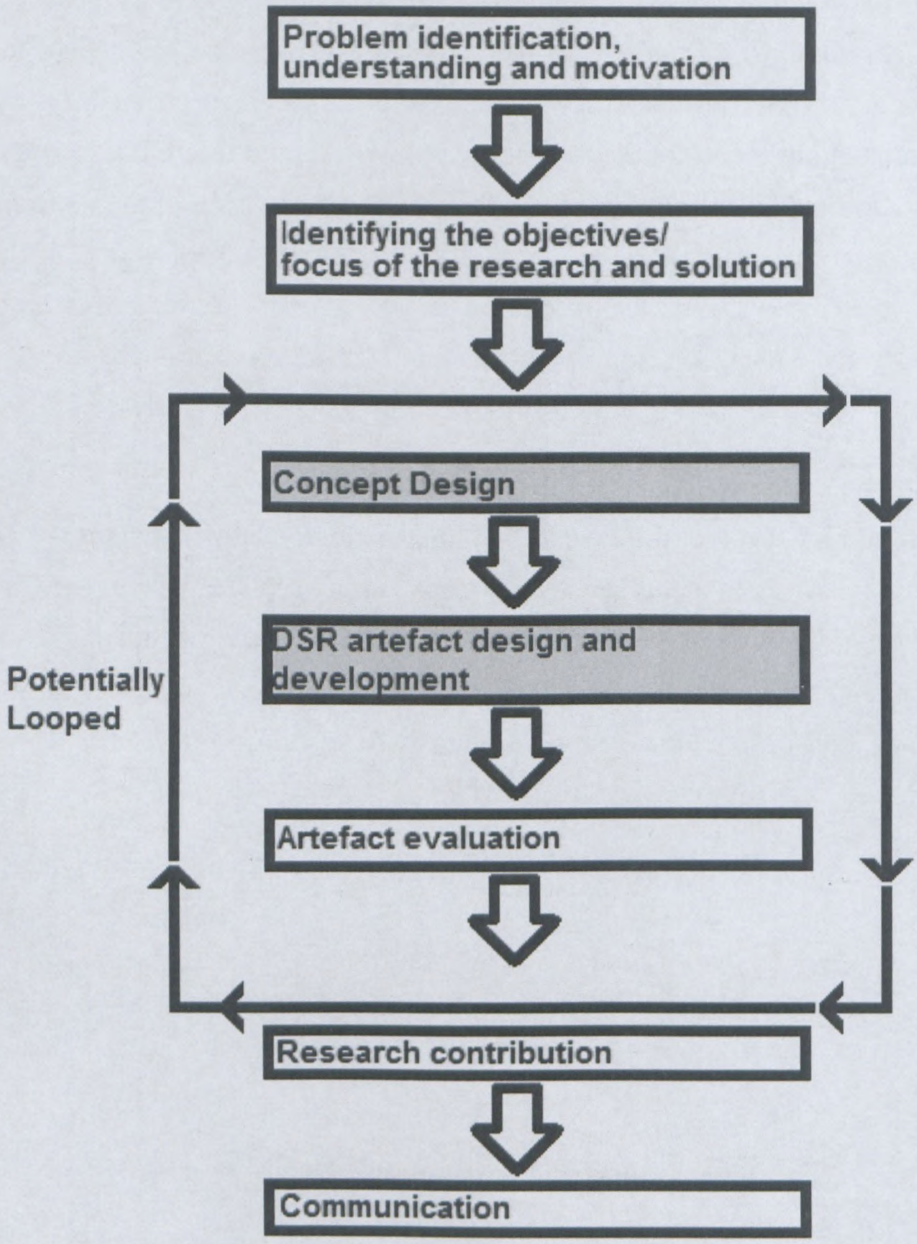


Figure 6.1: *Focus of Chapter 6 in relation to the methodology.*

Most of the literature sourced on the topic of DSR identifies a design and development component as being part of the overall research in either in either a single or spread across multiple phases of the research. DSR is based around the concepts of "build" and "evaluate" and frequently iterate between the two phases through the research process ((Kuechler & Vaishnavi, 2008). During the DSR process a number of research artefacts are created which act as tools for examining a given research problem. The artefacts are evaluated in order to generate the research data and findings providing some insight into a given research problem. Chapter 3 provides a more detailed discussion on the topic of design-based research.

This chapter focuses on the concept design as well as the solution design and development phases of the chosen DSR methodology, specifically the third and fourth step in the chosen research Methodology defined in Chapter 3. Within this chapter an overview of the design and development phases are provided as well as several of the design diagrams. Within this chapter some effort is also made to detail some of the design and development choices made as well as the factors that gave rise to these choices and the underlying logic behind the given chosen choices.

The primary focus of this chapter is to detail the design and development phases of the research. The design and development process produced the various research artefacts, and so this chapter simply considers the creation process and created artefacts.

The first section of this chapter focuses on the design of the solution concepts, providing the reasoning for the chosen design choices. The second section considers the design and creation of the solution architecture, the solution functionality and the creation of the data structures.

6.1 Concept Design

This section considers the design and creation of the solution concepts intended to solve the identified solution problem and achieve the identified solution objectives/aims. The first subsection considers the creation of the overall solution concept and details the process undertaken to decide on using a semantic metadata repository to attempt to solve the identified solution problem. The second subsection continues and considers the internal workings of the intended repository solution.

6.1.1 General Solution Concept

This section will discuss the process involved in designing and choosing the final solution concepts, that of a semantic metadata repository.

The motivation for undertaking the software development effort originated from overarching project with its primary focus relieving tension within rural communities in need, specifically looking at supporting HBHC services through the application of IT.

During the overarching project in which the research and development was conducted, the problem area was analysed and the solution and research problem were identified. The process of identifying the research and solution problem is discussed in more detail within Chapter 4.

The identified solution problem was based-on the ethnographic study (detailed in Chapter 4) which related to: the low penetration of IT-based solution within rural HBHC, the lack of standards for HBHC data-elements within rural South Africa and the resulting misinterpretation and misunderstanding of these data-elements.

Initially the research looked only at the lack of standards for HBHC data and the resulting problems and issues and also inversely the advantages of standardising these HBHC data-elements. One of the notable advantages of standardising the data-elements as identified in the literature, within Chapter 2 specifically Section 2.2.4 and touched on in Section 2.5 and Section 2.6, was the ability to enable interoperation. Once systems conformed to a specific standard, then interoperability would be possible and would be very advantages in South Africa and especially HBHC where little interoperability exists.

But as the initial exploratory research proceeded it became clear that the problems faced in HBHC ran significantly deeper than what was initially assumed. From further analysis the findings from the initial exploratory study, along with secondary findings from subsequent visits, it was clear that there was not just simply a lack of data standards. It was also made clear that the current IT solution was simple being used as a means of electronic backups i.e. capturing data into spread-sheets. In the case of Kayamandi and no computers at all in the case of Motherwell, other than as a means to printing the numerous papers forms.

A system that would have enabled some form of interoperability within rural HBHC would have been ideal. Such an idea could however not be achieved because the necessary level of technology was simply not there within the intended context of use. Even if the necessary technological infrastructure was available the literature review made clear mention of the fact that achieving interoperability was not a simple task and that literally decades of effort and work had gone into achieving interoperability with limited success. Thus focussing on interoperability within HBHC was not feasible, significant effort would first be needed to get the rural HBHC context's general level of technology to a level that system interoperability would be feasible.

Thus the focus of the development efforts shifted from looking at interoperability to looking even lower at the data-elements themselves. These data-elements had similar names, were used in similar context but no clear single definition existed for these data-elements and their relationships with each other was not clearly defined either. These differences existed not only between the HBHC initiatives but also within the same HBHC initiative.

Because of the above realisations the solution concept changed from the creation of a tool to enable interoperability, to the creation of a tool that could be used to in some way address this inconsistent treatment of HBHC data-elements, ideally developing a base for future efforts.

It was conceived that a new IT-based system could be created which would replace the current paper-based systems and would thus, if scoped, designed and developed correctly, capture explicitly the interpretations of the various data-elements. If these interpretations could be captures it could ease understanding of these differences potentially aiding in standardising the date elements and could even potentially enable the desired level of interoperability.

During the beginning of the CPUT/SAFIPA project such an attempt had been made to create a system that would replace the paper forms. This “replacement” system met with several problems however. Firstly all it did was digitise the paper forms literally copying the structure and field of the paper forms and putting them into a software system. In effect this intended solution offered no advantage over simply using the paper forms and served only as an example of implementing technology for the sake of implementing technology with no intent to innovate beyond digitising. A second issue was that the solution was specific only to a single HBHC initiative and would be irrelevant if implemented in another HBHC initiative. A third and final issue was that the replacement system failed to encapsulate all the HBHC initiatives requirements and more importantly the system failed to take into considerations that the forms would change. In fact the forms used by Stellenbosch Hospice had changed twice within the span of a year during the research period.

Simply creating a new system was thus not an option either; such a system would have a very low relevance in the general HBHC context. It is important to remember that useful software is developed to be used and be of use, thus for software to be useful it needs to provide some form of value in its use. It is thus important to understand the context of use, where the software will be used and for what purpose. Also important for the development of quality software is to not simply expect that all the contexts of use will be exactly the same.

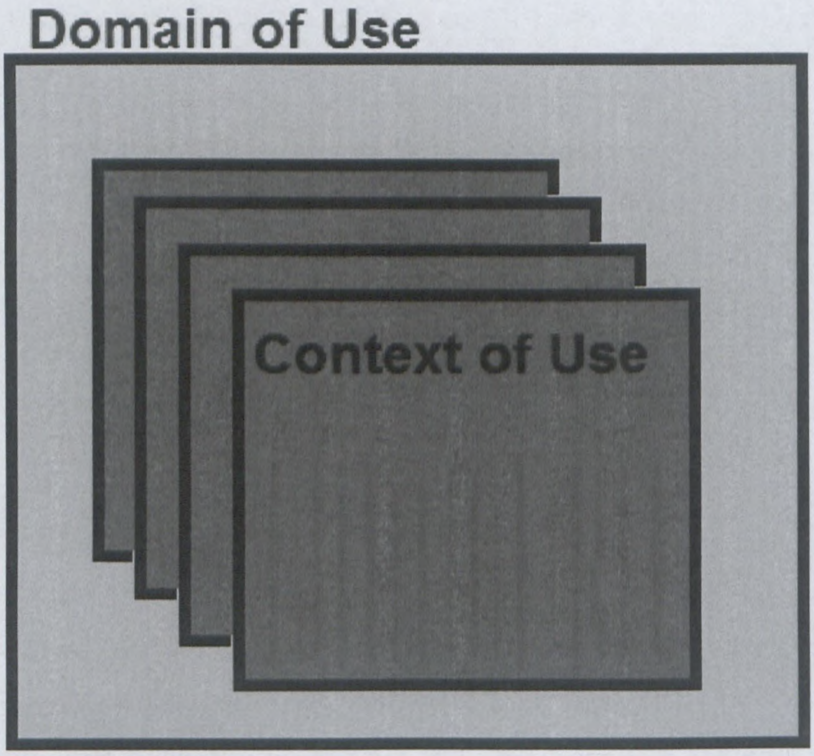


Figure 6.2, *Domain and Context of Use.*

Attempting create a system with a high level of relevance (thus making one system that could cater for multiple contexts of use or even the entire domain of use) would not only be a prolonged undertaking but would also possibly be unfeasible. Although the various contexts of use have several common characteristics (since they all belong to the same domain of use) they also potentially contain different characteristics which might vary greatly from other contexts of use and have a strong influence on any intended solution. These similarities and differences however can only be identified by analysis of the individual contexts of use.

In order to reach a high level of relevance within the solution artefact a detailed analysis of all the data-elements within all the HBHC initiatives would be required. Some form of agreement would have to be made amongst these HBHC initiatives as to the definition of any dissimilar data-elements and the system would also have to take into account the available infrastructure of all these HBHC initiatives. Not all initiatives have computer that could run such a system and some do not have computer available at all or for that matter a reliable source of electricity.

Thus the eventual solution concept was born, namely: that of differing interpretations and meanings placed on similar data-elements.

The solution would not try to solve the solution problem by forcing some sort of standard or simply digitising the current system but would rather be a tool which could be used to understand these different interpretations of the data-elements. By focusing on understanding rather than fixing the systems scope was made manageable (since the tool did not need to take into account every unique context of use within the HBHC domain of use) and offered long term advantages. If the tool was used successfully it would be an effective starting point for not only getting a ground level understanding of the situation within HBHC environment but also a good starting point for future attempts to fix these issues which surround data-elements.

In order to better understand the data-elements it was important to store data about these data-elements, thus creating and storing the metadata of these data-elements. As detailed in Chapter 2 Section 2.3, metadata has a number of uses including clarifying the primary data and enabling a number of additional operations, such as indexing and retrieval on the data. In the case of the research the metadata potentially offers a means of capturing the data-elements and their relations.

Thus the eventual solution concept was that of a repository that could store metadata that represented the data-elements used within the different HBHC initiatives. The repository was intended to archive the data-element's metadata and the source materials from which the metadata items were derived and allow these items to easily be modified. The system was intended to present the information in such a way that a clear understanding of the usage of the data-elements could be gained. The reason for storing the sources from which the content was derived was, to allow some traceability from source to product (the metadata-elements), thus allowing a means of managing the possible misinterpretation of source material.

The figure 6.3 shows the initial conceptualisation of the intended repository solution. The repository itself consisted out of two logical layers: firstly the data layer and secondly the metadata and semantics layer which corresponds to the data and further clarifying the data. The source of the data and the corresponding metadata and semantics would be primarily be the care documents used by the HBHC initiatives, but it was conceived that current systems and the people who work for the HBHC initiatives (the caregivers, nurses, administrative staff etc.) would also be able to contribute to the creation of the solution content (data, metadata, semantics).

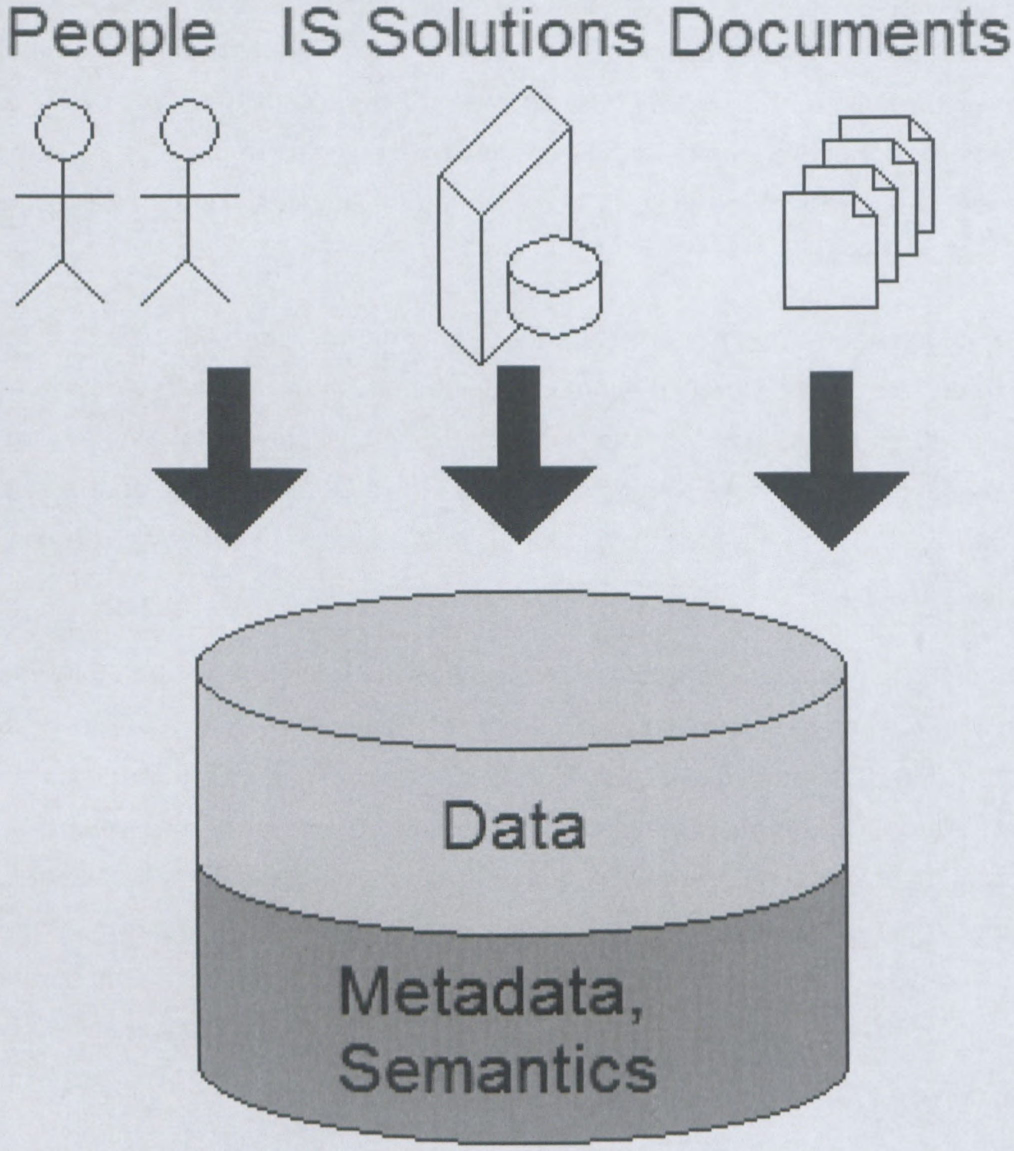


Figure 6.3: *Initial Conceptualisation of data flow.*

Initially the repository was not intended to capture the data itself, such as the patient's name, but to only capture the data-elements, the fact that a patient has a name. The focus of the research and the development was on the metadata/semantic layer in part because of the very real ethical issues that it would involve when dealing with real patient data. It was conceived that adding the data layer would add greatly to the future applications of the solution system.

As the development process continued it became apparent that simply storing the data-elements wasn't sufficient to develop an understanding of the data usage within HBHC. Two additional pieces of information needed to be stored: firstly, the semantics of the data-elements and secondly, the structure and relations of the data-elements.

It was clear from analysing the care documents that the same piece of data was known by several different synonyms, semantic differences existed within the same initiatives and between different initiatives. The semantics of the data-elements needed to be included in order to prevent accidental duplication and clarify the relations between data-elements. Without adding semantics it would also be difficult to effectively use data, someone looking at the data might not know if the piece of data he was looking at was in fact the correct piece of information or if there might be another more current or correct version available simply hidden behind a different name.

A single data-element on its own can rarely be of much use to anyone, simply knowing a patient name but knowing nothing else about the patient is not useful. Knowing the patient's weight on its own for example isn't useful either unless the patient's gender, age, height (and several other elements) are included in which case the patient can be assessed as being over- or underweight. Multiple data-elements would thus be required and how and when they related to each would also need to be captured in order to ensure that the repository and its content would be useful.

From the beginning it was clear enough that different data-elements had different structures and related to each other in unique ways. It eventually became clear that the structures and relations needed to be specified since they played an important role in collectively in performing specific tasks.

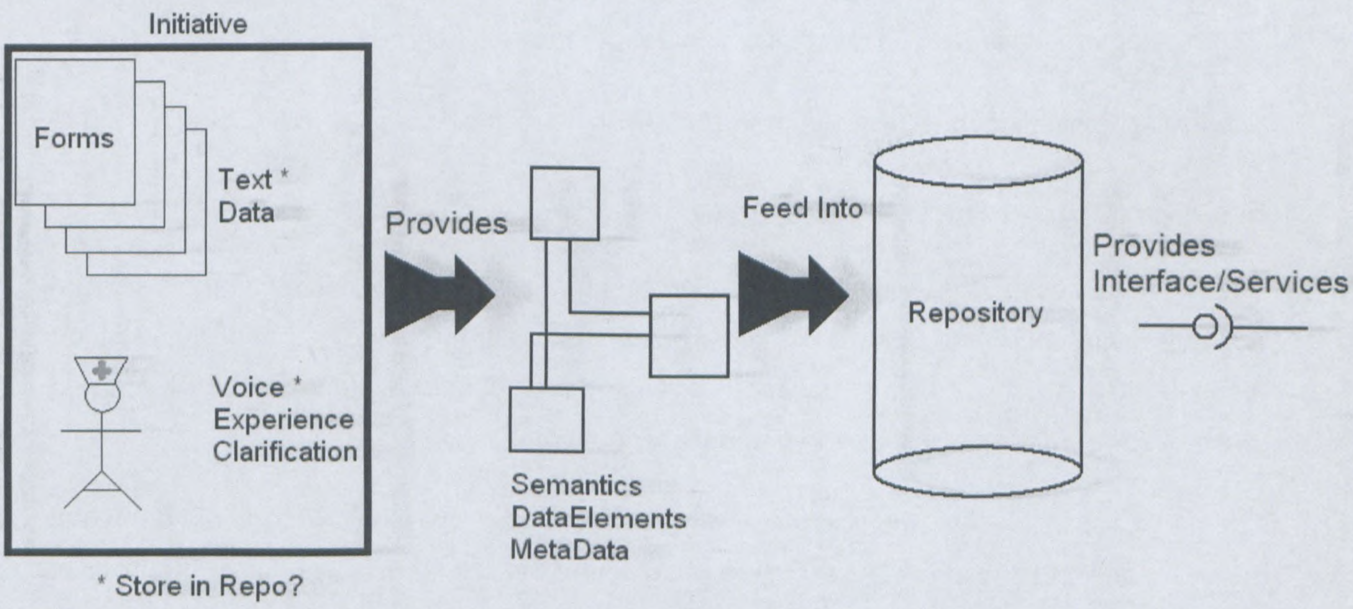


Figure 6.4: *refined conceptualisation of data flow.*

The Figure 6.4 is a further refinement of the initial conceptualisation shown in figure. The second refinement above placed more focus on the sources from which the data semantics and metadata were derived. In Figure 6.2 the initiative is identified as being the source of the varied documentation and the people who provided the data and the related semantics and structural-metadata which from the repository's content. This information originating from the initiative feeds into the solution repository which archives the information but also provides an interface in order to allow the archive data to be used. These interfaces between the solution repository and the external systems was, from beginning, identified as being a long term goal and beyond the scope of the immediate research and development efforts.

At this point the nature and storage of the source material (such as care documents, or caregiver interviews used to derive the metadata from) was brought into question specifically the types of source materials which the solution would need to cater for and how it would be stored. Initially it was assumed that the source material would be primarily narratives text-based such as interviews with the people involved in the HBHC initiative, or transcribed documentation. But if text-based interviews could be stored why not the audio files of the interviews? If transcribed documents then why not images of the documents? It became clear that numerous possible source materials were available and that the system would have to cater for them. It was thus decided not to limit the possible source materials and attempt to cater for as many possibilities as possible.

This section looked at the creation and logic involved in choosing a semantic metadata repository, based-on the ethnographic findings as well as the literature review. This section attempted to explain and justify the specific concept design decisions that were made as well as providing a necessary level of clarification about the solution concept.

The following section considers the internal solution concepts.

6.1.2 Solution Internal Concepts

The previous section focused on the creation of the overall solution concepts, focussing on not only the intended solution and its requirements but also the context of use. This section focuses only on the solution concept, that of the semantic metadata repository, and considers the internal workings and internal concepts of the solution.

The figure 6.5 focuses on the internal working of the repository. While designing the internal solution concepts it was necessary to explicitly define the relationships between the internal concepts identified in the previous section.

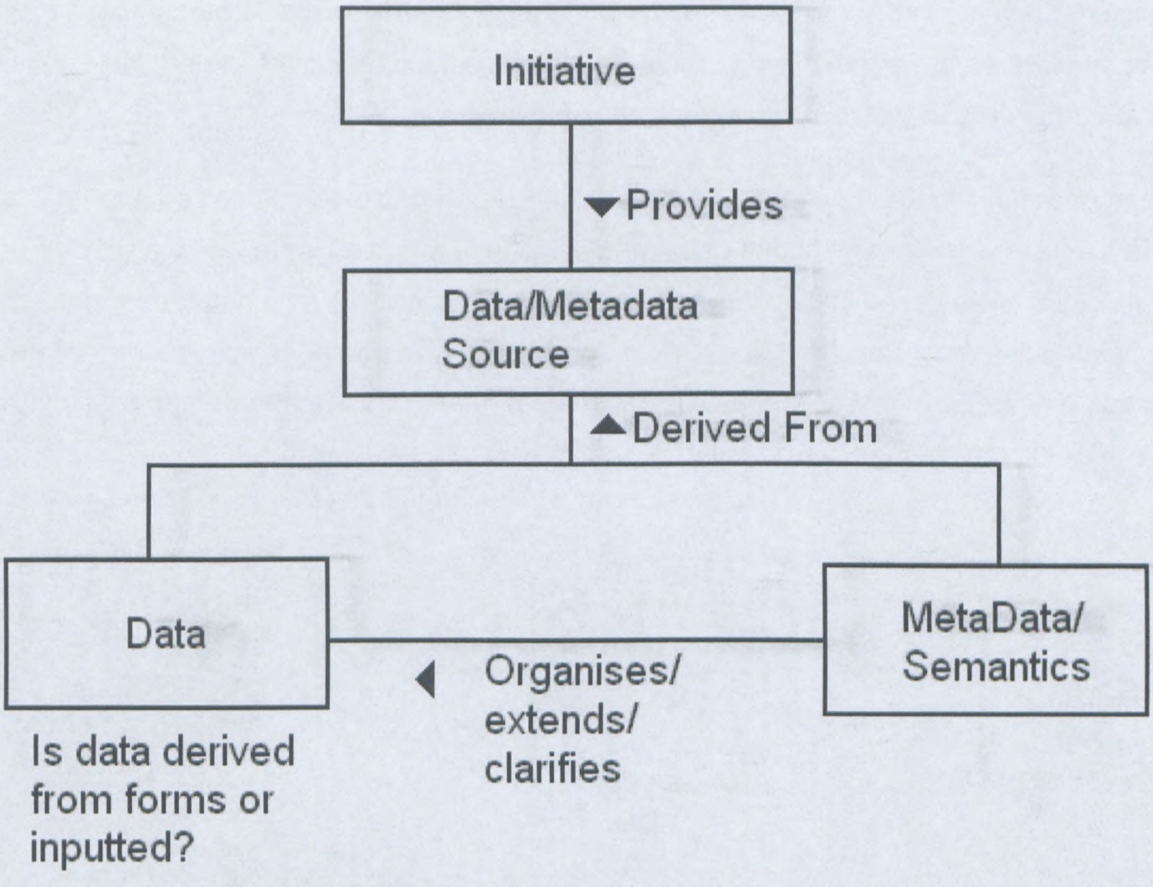


Figure 6.5: *Initial conceptualization of data relations within the repository.*

The HBHC initiative is the starting point for the creation of the data and metadata/semantic items, the initiative is responsible for providing the varied source materials. The source materials are used to derive the metadata and semantics as well as identify the data (although as mentioned before the data is not the primary focus of the research). Although initially assumed that the data could be derived from pre-existing sources it became obvious that there existed limitations to the idea that data could originate from pre-existing systems. Foremost of these limitations was that most of the pre-existing sources were paper forms, of questionable quality as discussed in Chapter 4 during the ethnographic study.

The most likely solution to gathering the data for the repository would be firstly to use the source materials, interviews and documentation being used by the initiative, to capture the data-elements and semantics (the metadata), and secondly capture as much historic data as possible, but capture most of the data in operation.

The purpose of the repository is to serve as an informative tool, to capture the data-elements and allow people to gain a deeper insight into the usage of the data-elements and relations between the data-elements specific to a given HBHC initiative. The data is not a major component of the solution system, although it seems clear that having actual data in the system as well can only help the primary purpose of the system, allowing the content to provide more meaningful insight. To that end the data would ideally be stored in the simplest format possible and a given piece of data would be assigned to a given metadata item which would further elaborate its structure, relations and semantics.

An important question arose of how the source materials could be interpreted and analysed to derive the system's content: the data, metadata and semantics. At this phase only a rough idea existed as to how this would be achieved. The possibility of automated data extraction and harvesting was considered but the literature on the subject (specifically metadata creation discussed further in Chapter 2 Section 2.3.6) showed that the means to effectively and efficiently do so does not yet exist.

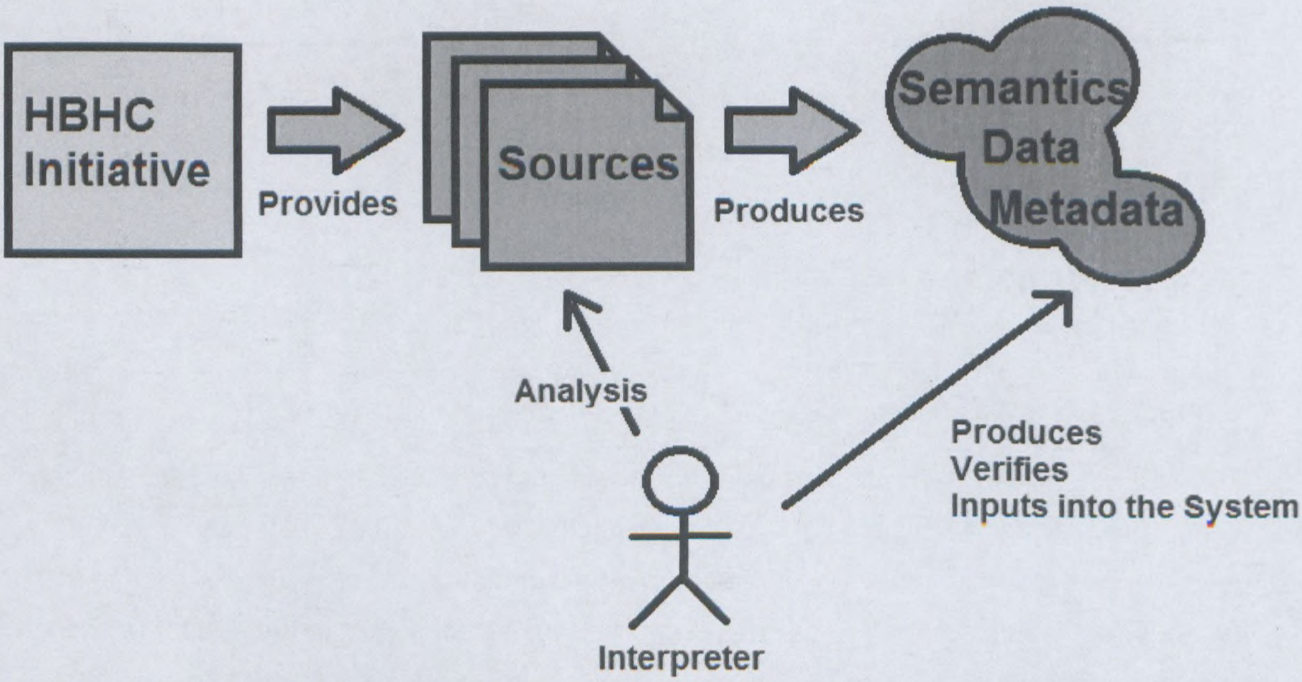


Figure 6.6: *the role of the human interpreter in content creation.*

Figure 6.6 shows the role of the human interpreter in the process of transforming source materials (documents, interviews) into repository content (metadata, data, semantics).

A human interpreter was thus envisioned to interpret and analyse the source material and derive the metadata, semantics and data. This interpreter would not be a member of the HBHC initiative but would be someone trained and chosen specifically for the role of interpreting the source material. A human interpreter would be advantages given the multitude of possible source materials of varying types which would be far easier for a human to analyse and interpret. The downside of course is that a human interpreter is susceptible to making mistakes and misinterpretation. Because the source material is available the origins of a given piece of information can be traced and thus correct and misinterpretation can be minimised by involving the people from the HBHC initiatives in the interpretation process. Furthermore the tool was never intended to capture some objectives truth or to define what the correct data-elements are, but rather to capture the interpretation and meanings of the data-elements as seen by the individual HBHC initiatives, whether these interpretation were correct or incorrect.

In order to limit the possibility of the human interpreter making mistakes it was decided to keep the data-elements and their relationships as simply as possible, as this would also aid in making maintenance of the repositories content simpler. A comprehensive solution could be created and would possibly provide better information to the intended users but the complexity would also increase the odds of the interpreter making mistakes during the initial input. Not only would a complex system create issues with date input but would potentially make management of the data-elements and their relationships more difficult.

In order to further elaborate on the relations that currently existed amongst the data, elements a blank form used by Stellenbosch Hospice was used and several data-elements were identified as well as their relationships. Only the most basic data-elements were chosen to clarify the relationships. This was a purposeful decision because the larger and more complex the chosen data-elements the more likely that possible issues might arise when attempting to clarify the relationships.

Table 6.1 looks simply at a patient data-element (sometimes called the client) and a handful of other data-elements that relate to the patient such as: the name, address, admission, service (care activity) and caregiver. The table provides an example of the identified elements (in bold) and the relationships (in italic).

Table 6.1: *Data-element and Relationship sample data.*

Patient/Client						
	Has A					
		Name				
			Composed of			
				First Name		
				Last Name		
		Address				
	Was					
		Admitted				
			Has A			

				Date		
				Institution		
	Is Provided					
		Service / care activity				
			Provided By			
				Carer / caregiver		
					Has A	
						Name

Figure 6.7 is based-on the identified elements and relations within Table 6.1. The figure was created in order to simplify the representation of the above relational data and to make it easier to examine and conceptualise.

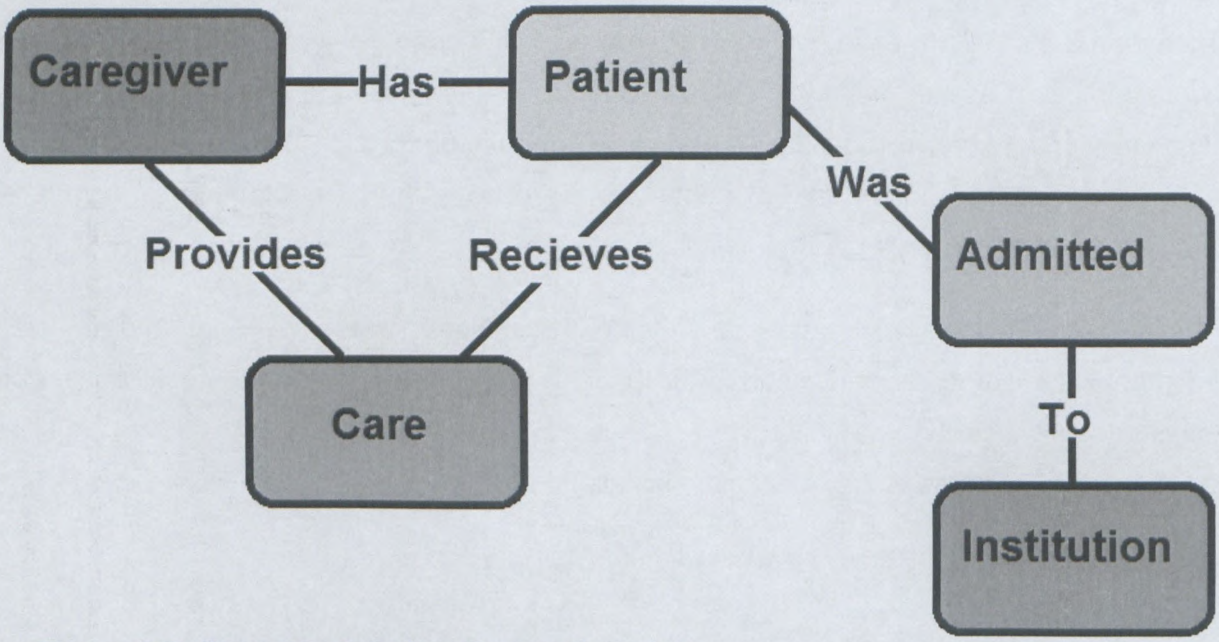


Figure 6.7: *Conceptual design of paper form data.*

Figure 6.7 did raise a number of questions such as: does a patient's Name constitute a separate data-element or is it simply an attribute? Is there a need to differentiate between attributes and core concepts such as a Patient and a Name? Would the name of a patient and a caregiver be the same data-elements or would there be a patient-name and caregiver-name? How would someone for example differentiate between a name and its components, first name and last name? Would admission be its own data-element, since it in turn can have separate attributes such as admission date, or would it be a relationship between a patient and an institution?

In order to attempt to answer the above question the Table 6.2 was created listing the identified relationships above.

Table 6.2: *Sample Relationships data .*

Relationship Name	Single/Multiple	Group/Not Group
Has A	Single	Not Group
Composed Of	Multiple	Group
Was	Multiple	Not Group
Provided	Multiple	Not Group

Table 6.2, was created by looking at the characteristics of the relationships identified and detailed in table 6.1.

As seen in Table 6.2 two additional attributes were added to the identified relationships namely: Single/Multiple and Grouped/Not Grouped.

Single/Multiple corresponds to the conventional Entity Relational Diagram (ERD) relationships 1:1 (single) and 1: Many (Multiple) although it is possible to extend the single/multiple to include additional cardinality (1 patient has 1 name or for example 1 caregiver has 10 or less patients). For the scope of the research and the initial iteration it was not deemed necessary to extend the cardinality information.

Grouped/Not Grouped was added to resolve the issue composite-style relationships, such as around first name and last name collectively forming a name. All the grouped elements that share a relationship collectively form the relating element. While none grouped relationships are individual, without direct relation to each other such as a patient, as a caregiver or a patient receiving care.

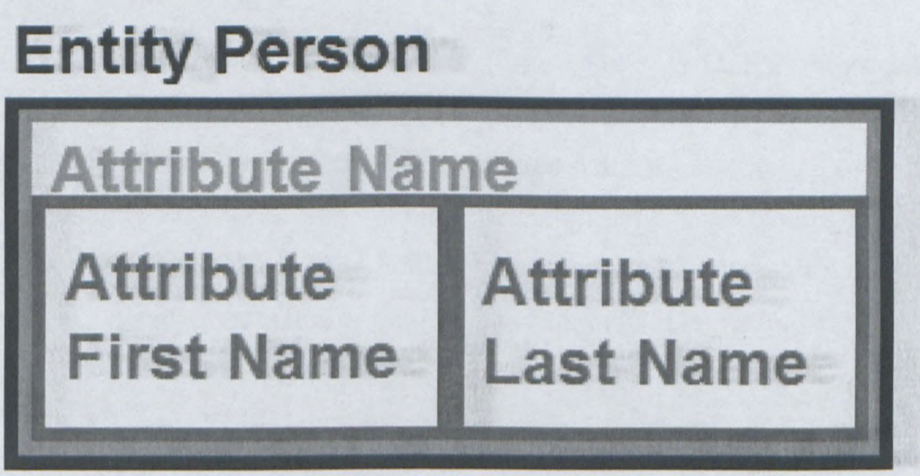


Figure 6.8: *Entity and Attributes as conceptualised in the development.*

A data entity is conceptualised as being composed of a number of data attributes, thus several options exists for handling the relationships between entities and attributes. It is possible to handle the entity and attributes distinctions within the actual data-elements themselves. A metadata-element (representing a data-element) can possible have some attribute which aids in differentiating whether it is an entity or attribute, but it is also possible to differentiate between entities and attributes based-on their relationship. The latter option was chosen because of the attempt to ensure that the metadata items were as simple as possible since it to ensure a minimalist-type of metadata was used (discussed in Chapter 2 Section 2.3.2).

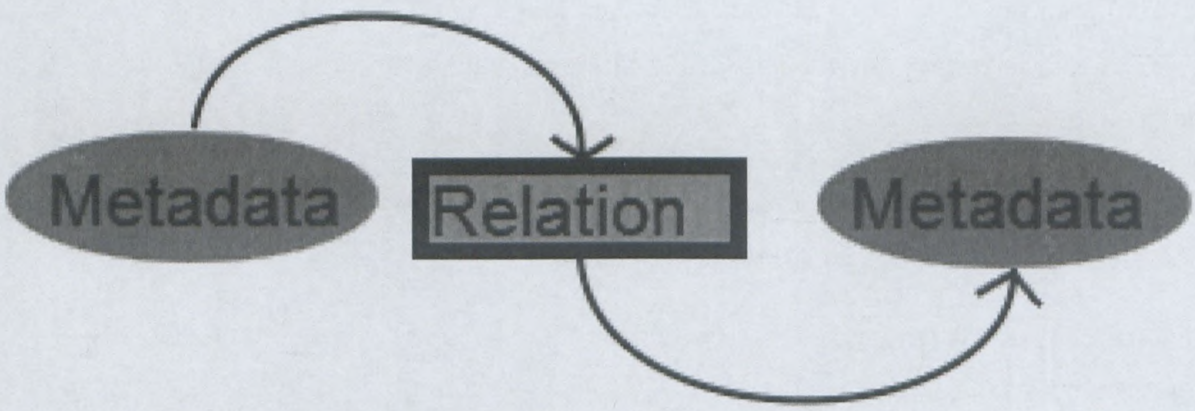


Figure 6.9: *metadata to metadata relationship conceptualisation.*

Thus the data-elements/metadata-elements and their relationships are conceptualised as shown in Figure 6.9. Two metadata items relate to each other through a relationship. This relationship is defined by the human interpreter along with the metadata which provides additional elaboration for the underlying data-elements. The relationship between two metadata-elements are one way, going from a 'parent'/'originating' element to a 'child'/'destination' element, but it should be possible to define multiple relationship between two elements.

The relationships between metadata-elements were conceived as being re-usable, as different metadata items can conceivable relate to each other in similar ways. Ideally these relationships are intended to be kept as simple as possible and the number of relationship to a minimum. Although it is possible for an uninformed human interpreter to conceivable define a unique relationship for every possible metadata combination, it was decided to keep the relationships as simple and open as possible to not prevent the expressiveness of these relationships.

The reasoning behind simple one way relations was that although an inverse relationship might exist and might exist in most other cases were the relationship is used but there will more than likely be exceptions. For example a Patient 'Has A' Name and a Name 'Identifies A' Patient but A Patient also 'Has A' Prescription and a Prescription 'Is Prescribed To' a Patient, thus in the first example the relationship is Has A and the inverse relationship is Identifies A, and in the section example the relationship is Has A and the inverse relationship is Prescribed To.

Table 6.3: *Sample test data.*

Data-element / Metadata	Relationship	Data-element/Metadata
Patient	Has A	Disease
Patient	Has A	Prescription
Prescription	Contains	Medicine
Medicine	Treats	Disease

Table 6.3 shows dummy data, some fairly simple data which isn't bases off of any real patient or medical data. This dummy data was used to test the purposed conceptual model shown in 6.9 and produced Figure 6.10.

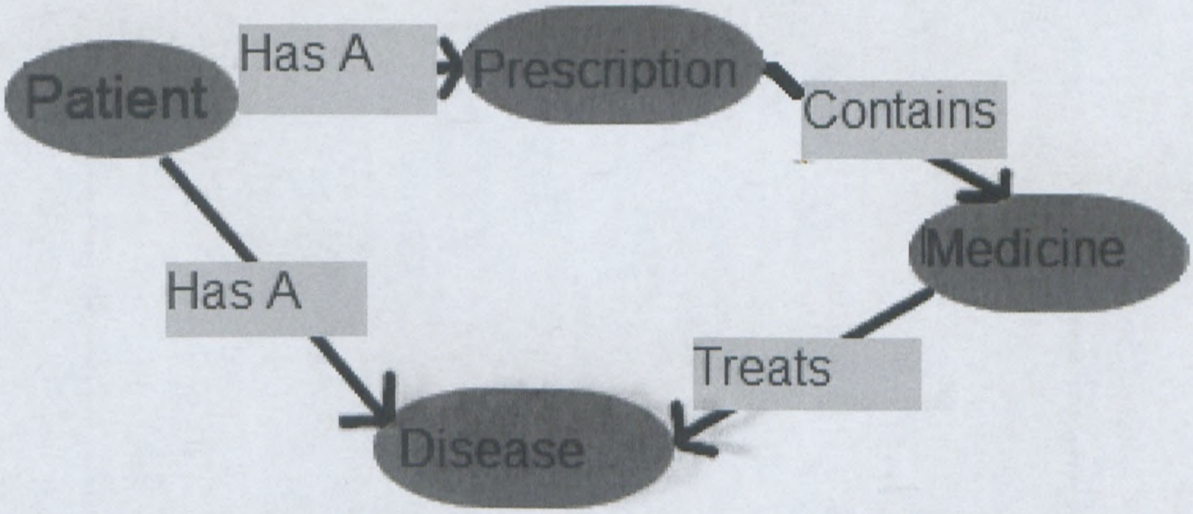


Figure 6.10: *Simple Dummy Data, Relationship conceptualisation.*

Figure 6.10 is initially detailed enough (without bringing in additional elements to define the 'attributes' of the various 'entity' elements and showing the inverse relationships) but this model doesn't successfully capture the semantics of the various elements. The simplest means of capturing the semantics for the various data-elements is to capture the synonyms, or simply a list of alternative names by which the data-elements are known. The ideal behind specifying the low-level semantics (the synonyms in this case) is to help with the usage of the data-elements. It is possible that if synonyms are not specified the same data-element might be known by different names and might be used and treated as actually being different data-elements. But simply specifying the synonyms seems of little use if it is simply a list of names that are only loosely connected to a given element.

To this end several examples of synonyms were identified: medicine when in storage is called stock, medicine when provided to a patient is known as a prescription and medicine given for a disease is a treatment (although treatment might include a number of other activities and items but for simplicity they are excluded from this example). Of course not all synonyms are associated with a relation, in some cases a Patient can be known as a Client.

From the above examples a synonym and a relationship are seen as being contextual in nature. Medicine in the context of a Patient is a Prescription; Medicine in the context of a Disease is a Treatment etc.

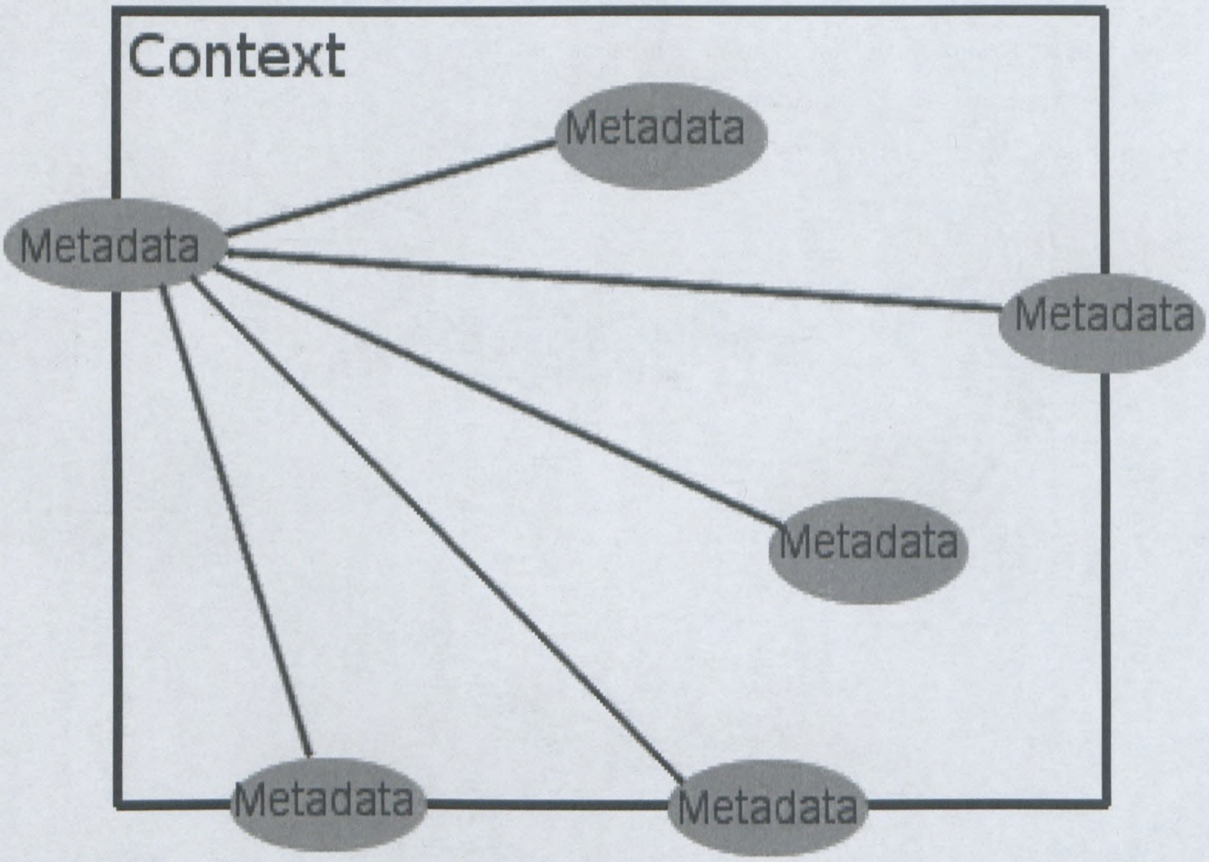


Figure 6.11: *Context in terms of metadata relationships.*

Figure 6.11 shows the context of the metadata/data-elements and their relationships. The elements themselves can be fully specific to a context; they are known only within a context and relate to other elements with the same context, or they can be partially inside the context, relating to other elements within the context but also relating to elements in other contexts. It is also conceptualised that the same data-elements/metadata-elements might relate to each other but in multiple contexts.

Thus the final conceptualisation of the data-element metadata is as shown in Figure 6.12, including the metadata elements, their relations to each other, their synonyms and the context which helps to define and clarify the relationships and synonyms.

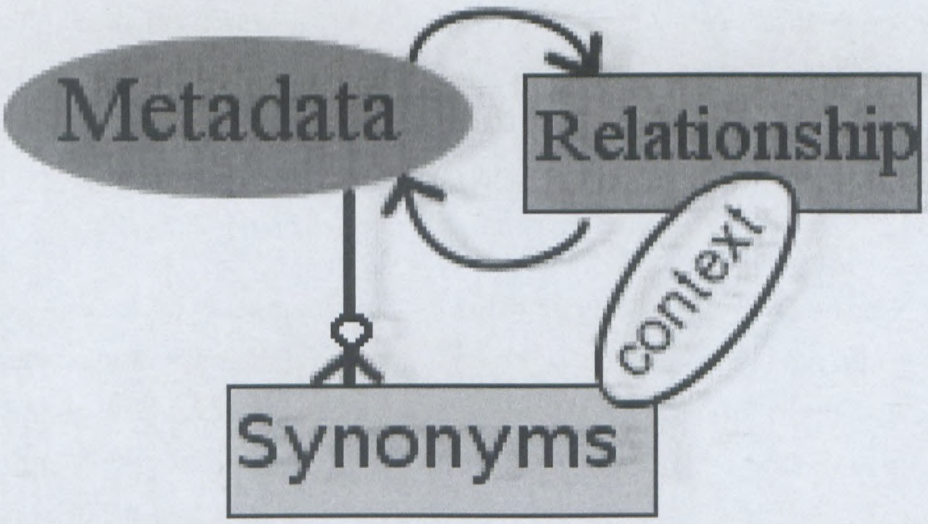


Figure 6.12: *the role of context in Metadata to Metadata relation.*

As shown in Figure 6.12. A given metadata-element (representing a data-element) relates to a number of other data-elements via a defined named relationship. These relationships also contains the context within the relations occurs. Each metadata item has Zero or more Synonyms which also define the context in which they occur, this provides additional information to the synonyms and helps to relate them to specific relationships. It should thus also not be possible to define the context of the synonyms and relationships. This implies that these synonyms are general (such as patient and client) and that the relationships are the same, for example Patient Has A Name, this should conceivably not change between different contexts. By specifying the context of a synonym, that synonym or synonyms for the metadata can be used when looking at the relationship with the same specified context.

This section detailed the conceptualisations of the internal works of the intended solution artefact. The following section takes the above conceptualisations and designs and constructs the numerous design artefacts.

6.2 Architecture construction, solution design and development

This section considers the actualisation of the previously identified solution conceptualisations. This section discusses the technical components of the design and development process undertaken as part of the research.

The repository was intended to be developed as part of an evolutionary prototype, where the prototype is refined as part of each iteration of the development process, where eventually the continually refined prototype becomes the finished product. Following an evolutionary prototype development process was ideal because there was no large upfront design involved in the development process. An initial ethnographic study was conducted (detailed in Chapter 4) which provided a deeper insight into the solution problem area, the context of use and offered numerous design ideas but the initial ethnographic study was not sufficient to answer all the design and development questions. Some design and development questions would not be known until they arose as part of the development process.

This is where DSR becomes useful as a tool for exploring some aspect of the problem area. The design and development of the artefacts takes known facts (in the case of this research the ethnographic study and the literature review) and attempts to develop a number of artefacts. During the development process the researcher/developer needs to make a number of design and development choices based-on what he known and might possibly need to re-examine either the context of use or the literature in order to address any problems that arise. As discussed in Chapter 3 the design and development components of DSR are thus creative knowledge generating processes.

The initial prototype on which the research is based was developed in other to serve as a proof of concept, to demonstrate that the assumption made during the design process was in fact correct. The design and development of the initial prototype also created a number of DSR artefacts that were further of use during the research.

Because of the personal skills and preferences of the researcher, it was decided that the initial prototype was to be developed using the Microsoft C#-programing language using the .Net Programming Framework 3.5 (which was later converted to the 4.0 Framework) and using a SQL Server 2008 relational database management system. Open Source software solution, such as Java was considered but as a proof of concept application that was not intended to be commercialised the advantages of using Java to create the solution repository, specifically the cost saving involved because of its Open Source nature, did not factor into the decisions and C# was settled on.

A further advantage of using Microsoft .Net based technology (C#) was the visual interface technology that was available. Visual Studio, the Integrated Development Environment (IDE) for Microsoft programming languages, offers an easy to use 'Drag and Drop' interface for creation UI as part of their Windows Forms technology which allowed for rapid development of UIs, although with limited stylisation and visual effects. Windows Forms was used to create the initial UI in order to rapidly create a functional system.

As discussed in Chapter 2 Section 2.4.1 under the heading of Semantic Repositories, a more visually rich UI offers a means of overcoming some of the usability issues inherent in semantic repositories. Here in too the Visual Studio IDE offered a useful technology, namely Windows Presentation Foundation (WPF). WPF is a newer technology used by Microsoft for the creation of UI, which offers more visual effects and presentations than Windows Forms. It is the intent in subsequent iterations to convert most of the repository UI to WPF, but for the iteration of development addressed within this research only the UI that would benefit most from a more visual display was converted to WPF, namely the Metadata display form (discussed in more detail within a subsequent subsection).

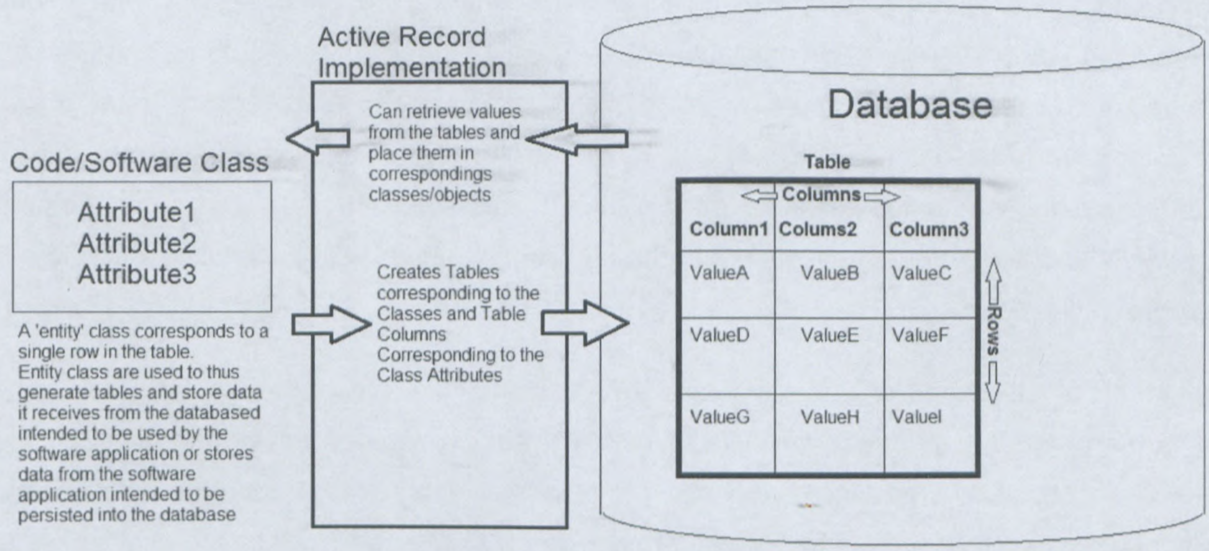


Figure 6.13: *simplistic representation of Active Record pattern usage.*

Castle ActiveRecord was used to persist and retrieve the data into and from database. Castle ActiveRecord is an implementation of the Active Record design-pattern for the .Net Framework and is similar to nHibernate (or more conventionally Hibernate for Java). Castle ActiveRecord enables databases to be created and modified 'on-the-fly' in response to changes made to the entity classes (programming code which represents data entities within the system). The active record pattern allows for more simple management of the database table and relations.

The reason for using an active record framework (Castle ActiveRecord specifically) was because it allowed for changes to be made to the database structure quickly and effectively. As attempted to be shown in the Figure 6.13 Castle ActiveRecord can map software/code classes to Database Tables, it can create and modify the database-tables based-on the software/code classes and connects the database-tables and code classes with each other to ease the process of reading data from the database and persisting data to the database.

Although the design and development had numerous objectives to meet (although it was never assumed that these objectives would be static) there was no big upfront design, a rough design was created and refined as the development continued. The reason for no large upfront design was because there was little design information available to go on (in fact identifying these design considerations is the primary goal of the research). Because of this hazy-idea of what needed to be done to ensure success it was obvious that evolutionary solution prototype would have be refined several times as new insights were gained as part of the design and development. Thus an active record pattern that allowed for these rapid changes was ideal.

Because Castle ActiveRecord is being used the underlying database isn't of any serious importance, and can easily be replaced by a number of other options, both proprietary and open-source. The active record pattern takes care of the conversion from coded programming data entities (such as programming classes) to relational database entities (such as database- tables). It is conceivable then that most relational database management systems would be able to be slotted into the repository solution with minimal changes to the programming code.

Because of the usage of an active record pattern, most of the classes design is loosely based around domain driven development (DDD). Although it was decided to leverage the advantages of Objected Orientation and not have the entity classes conform to Anaemic Entity Classes. Anaemic Entity Classes are simply entity classes in which the entity class is little more than a collection of accessors (getter and setter) with no functional code. Usually when Anaemic Entity Classes are used the functional code is placed in another class called the Data Access Object (DAO). Separating the DOA and the Data Entity offers significant simplicity but limits the ability of the Entity classes to benefit from Object Orientation concepts such as polymorphism, inheritance and code reuse.

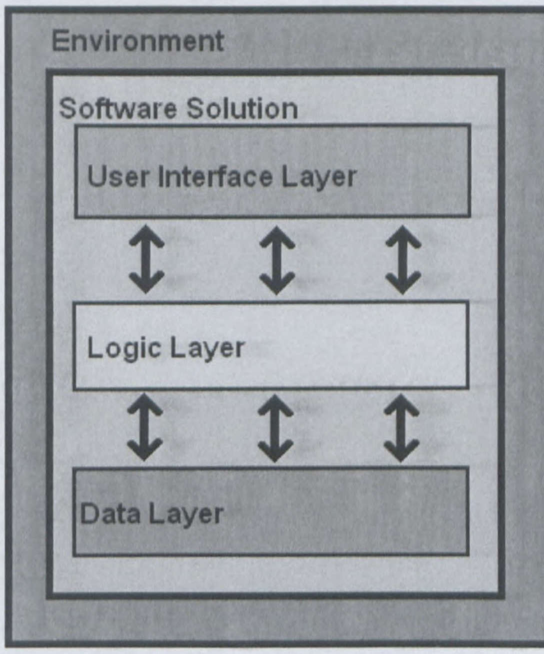


Figure 6.14: Basic envisioned components of a system.

The basic components of most software system fall into four categories: the User interface, Logic/Functionality, Data, environment. These components are not universal in all systems and their placement and interaction differs dependent on the overall architecture of the system. The repository solution created during the research was not intended to be multi-user, multi-platform or web-based, as a proof of concept prototype it was a single system intended to run on a single computer and be used by a single user, thus the categories shown in Figure 6.14 are all relevant to the intended repository solution.

The user interface (UI) is the component with which end users interact to meet a specific goal or perform a specific function. Not all systems have UIs because not all systems have human users some systems interact with other systems, but in the case of these systems that interact with other systems they still have interfaces by which the other systems can interact with them.

Through the UI the user is able to access the underlying logic/functionality of the software solution or initiate processes which access these functionalities. In some cases where the software is not particularly complicated the functional layer and the data layer might be one in the same, the advantage of course of separating the data layer from the functional layer is that it allows the data layer to be replaced without affecting any of the other layers.

Most software system manipulates data in some way; it receives it from the user or other external sources, formats the data, processes the data etc. However not all systems necessarily persist data for a long period of time, for example a simple calculator program only stores data for a short period and mostly only long enough as is necessary to perform any required calculation. Additionally not all systems use the same types of data or utilise the same types of storage technology (such as flat files or relational databases). Thus not all systems would necessarily have a data layer.

Finally, the environment does not fall within the bounds of the software system but it includes the underlying hardware, the users and the organisational context in which the software solution must function. The environment plays an important role in defining how, where, when and by whom the system will be used.

To this end the discussion of the architecture of the repository solution will touch on these topics specifically the functionality and the data layer. The environment of the solution is discussed as part of the ethnographic study within Chapter 4 and the user interface is discussed in the following chapter as part of the solution instantiation.

The next section will look at the functionality of the intended repository.

6.2.1 Functionality

This section considers the functional component of the designed and developed architecture.

The majority of systems are developed to meet specific objectives, these objectives are intended to solve problems, exploit opportunities, test propositions etc. In order to meet these objectives a system must perform specific task and provide specific functionality, after all a system that does nothing and does not allow an user to do anything is not very useful and has no reason to even be develop at all.

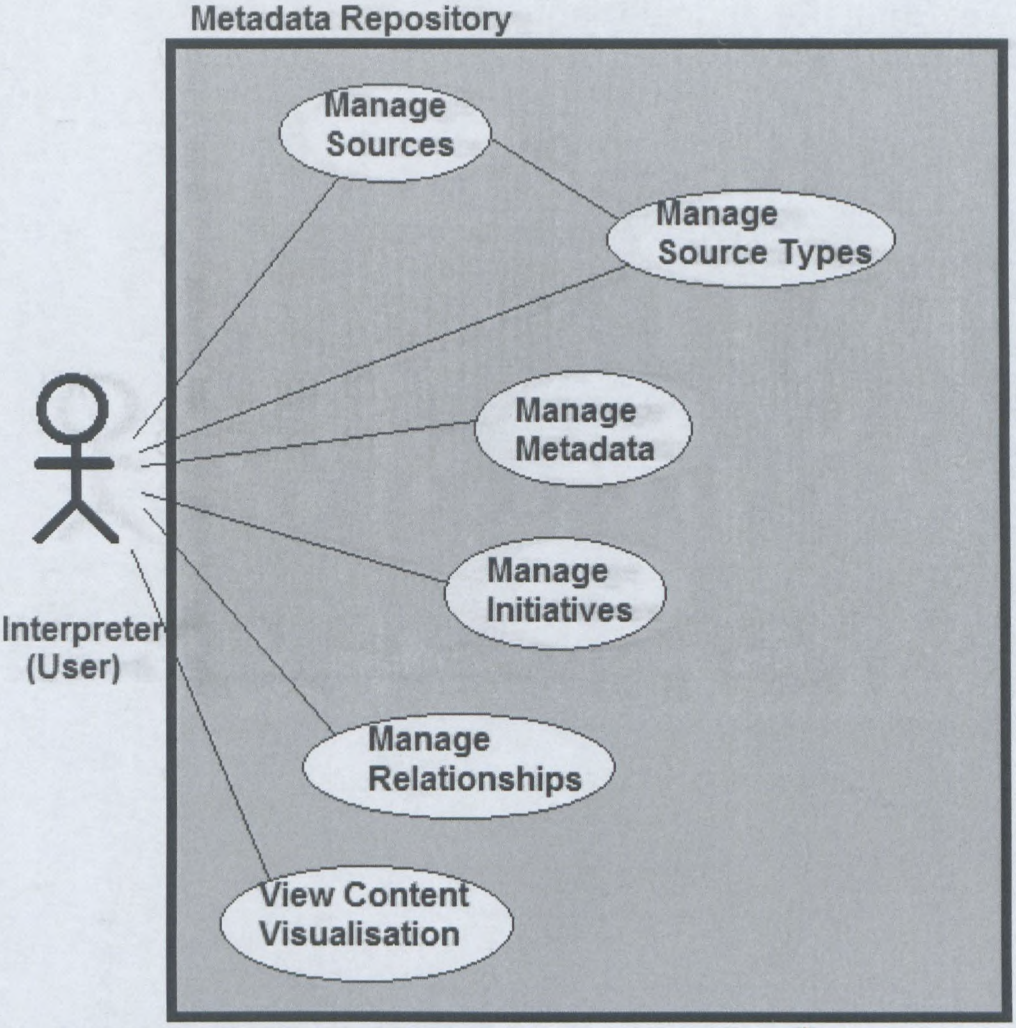


Figure 6.15: *Use-Case Diagram defining system functionality.*

The prototyped solutions intent it is be used as a tool for capturing the necessary semantic and structural metadata, it is essentially an input tool for capturing this data. Thus it must provide the necessary CRUD (create, read, update, delete)/management functionality for the different data components, allowing a user to create and manage the data.

But furthermore a solution system cannot be a 'black-hole' a system that allows you to enter data without getting anything out of it. Because the system is informative in nature, and thus prospective in its orientation, it also needs to be able to use and present the information in a beneficial way. In order to effectively allow for the management of the data-elements a mechanism for effectively displaying the semantics and the structure/relations is required, thus a visual interface which can visualise the data is required.

The following section considers the data layer of the application, specifically focussing on the database-table structures.

6.2.2 Database, Entity Relational Diagrams

This section focuses on the database structure created in order support the semantic metadata repository.

The ERD was developed based-on the conceptualisations listed in the first section (Section 6.1) of this chapter.

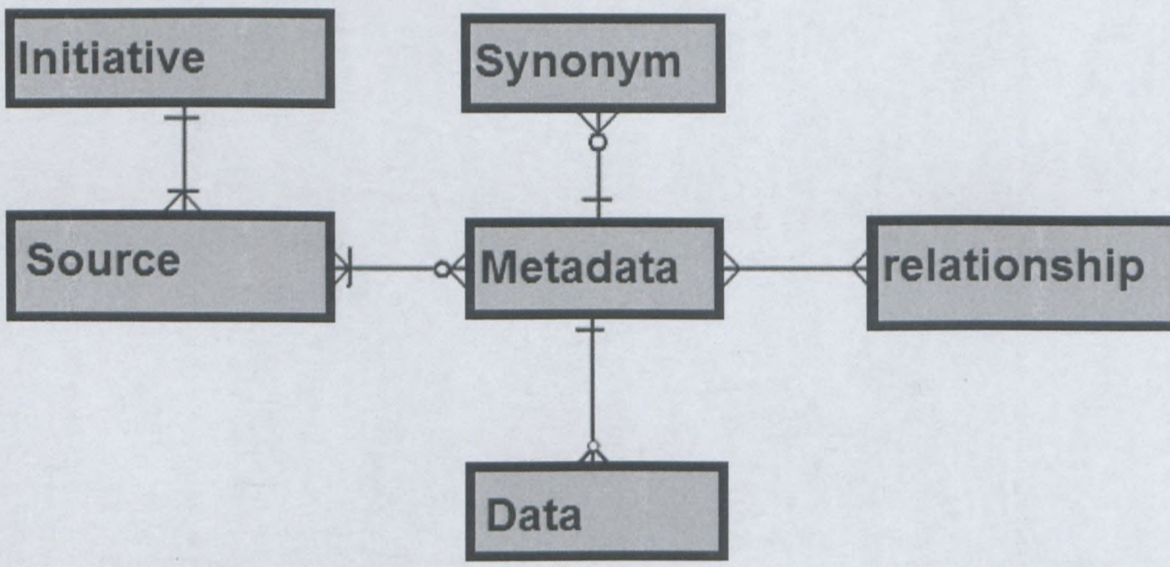


Figure 6.16: *Initial ERD.*

The initial requirement was for the repository to have the ability to store the data-elements from different HBHC initiatives and store the various source materials from these initiatives which the data-elements are derived from. Since both the source materials and the data-elements (called metadata in the ERD shown in Figure 6.16) are unique to a given HBHC initiative thus it was decided to relate the metadata to the source material and the source material to the initiative. The importance of also storing the source material from which the metadata was derived was discussed earlier but essentially the source material is stored in order to ensure that metadata can be traced back to its source material. The hope is that it will aid in the management of metadata items (in case of misinterpretation or if doubt exists around a given metadata item then a user can trace the metadata back to its origin and either ensure that the item is correct or make any needed changes).

The relations between HBHC initiatives, Source Materials and derived Metadata-elements does itself present a new issue, if a identified data-elements does not have a source it cannot be related to a given initiative. However enforcing the requirement that data-elements need to relate to a source material is acceptable since if a data-element could relate directly to an initiative it could lead to a situation where the source of an identified data-element is uncertain, and the quality of the data-element might be in question without there being an established means of ensuring its authenticity or correcting any mistakes.

Relationships were deemed to be generic enough, if they were kept simplistic enough, to be usable across different multiple initiatives and thus relationship data is not linked to a specific initiative or piece of metadata. A specific relationship is only related (indirectly) back to an initiative when it is used to relate two metadata items (which are directly linked to an initiative) to each other.

The initial ERD only caters for very low level semantics, essentially only providing a simple list of string values denoting the different synonyms of the metadata. It was considered to possibly extend the initial model to include the references to external reference sources such as SNOMED but this was beyond the scope of the initial development. A reference code could however be inserted into the synonym database-table to enable some form of connection to an external reference source but the system was not intended to cater for these external reference vocabularies.

Because the intent of the metadata is to make the underlying data more usable, there needs to be a relationship between the data and the metadata. The data relates to other data-elements via the metadata and the metadata relationships and the data are purposefully stored in as simple a format at possible as decided in the initial conceptualisation. Because the Metadata-elements are meant to be fairly generic (such as symptoms) and the Data is intended to be more specific (headache, fever, stomach pain etc.), a single Metadata item can relate to a number of Data items. Data thus represent a specific data value and the Metadata represents a specific type of data value.

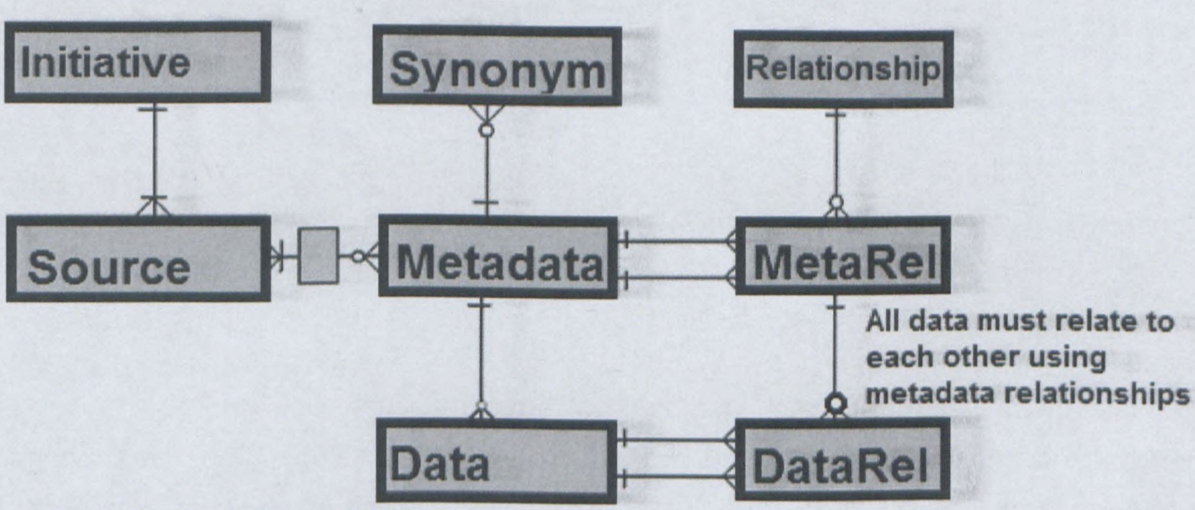


Figure 6.17: *Refined ERD.*

Because of the limitations of most relational databases systems, the initial ERD shown in Figure 6.16 was not yet practically implementable since most database systems have difficulty handling Many to Many relationships (such as between Metadata and Source and between Metadata and Relationship). Additionally the initial ERD was still not yet detailed and refined enough with several issues. The initial ERD was capable of defining a fair amount of the metadata structural information but simply linking the Data with its corresponding Metadata was not sufficient in transferring the metadata's structural specification to the data-elements itself.

The DataRel was added to the ERD as shown in Figure 6.17 in order to allow specific instances of the data-elements to relate to one another. The DataRel uses a user-defined relationships between metadata specified in MetaRel to ensure that the data instances cannot relate to each other in anyway other than those specified by the metadata. The DataRel thus defines how two Data items relate to each other but a DataRel must be based-on an existing MetaRel item, which in turn defines how two Metadata items relate to each other.

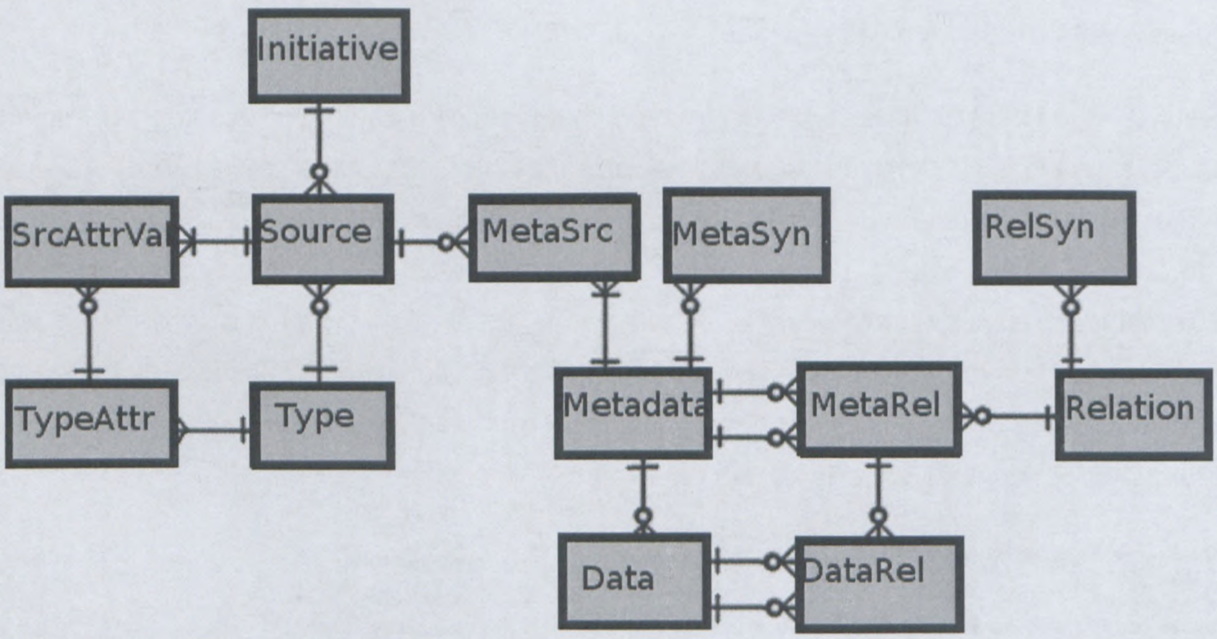
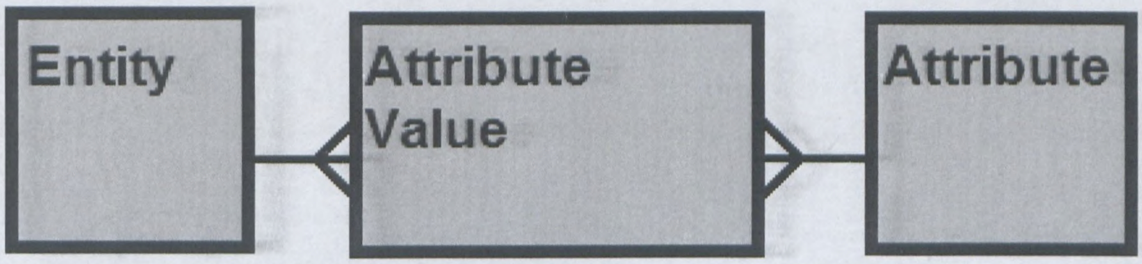


Figure 6.18: *initial implemented ERD.*

Figure 6.18 shows the ERD in more detail, synonyms were also added for relationships and the source model was further refined. This ERD was used for the first implementation of the prototype repository.

It was not the goal of the research and the development project to look at what constituted a standard for HBHC source material thus an Entity-Attribute-Value (EAV) approach was followed to define the attributes for the different types of source materials. A more explicit form of the EAV model was used for the Source, Type and attributes and was considered for the Metadata component as well but had to be refined, as seen in the ERD, before it could be used.



Person	Gender: Male Height: 1.7 m Name: John Smith	Gender Height Name
Building	Height: 3 floors Colour: White	Height Colour

Figure 6.19: *Simplistic example of Entity-Attribute-Value (EAV) model.*

Within the conceptualisation it was required that system caters for as many types of source materials as possible. To cater for all source materials would have meant that the Source Table (or the Type Table) would have had to have an extremely large number of fields (since although source materials have many similar attributes they also have a number of differing attributes). An EAV model allows the attributes a given entity (in this case the Source Type) to be defined as needed. By using an EAV inspired pattern any type of Source Material can be cater for since any required fields can be added for the specific type of Source material.

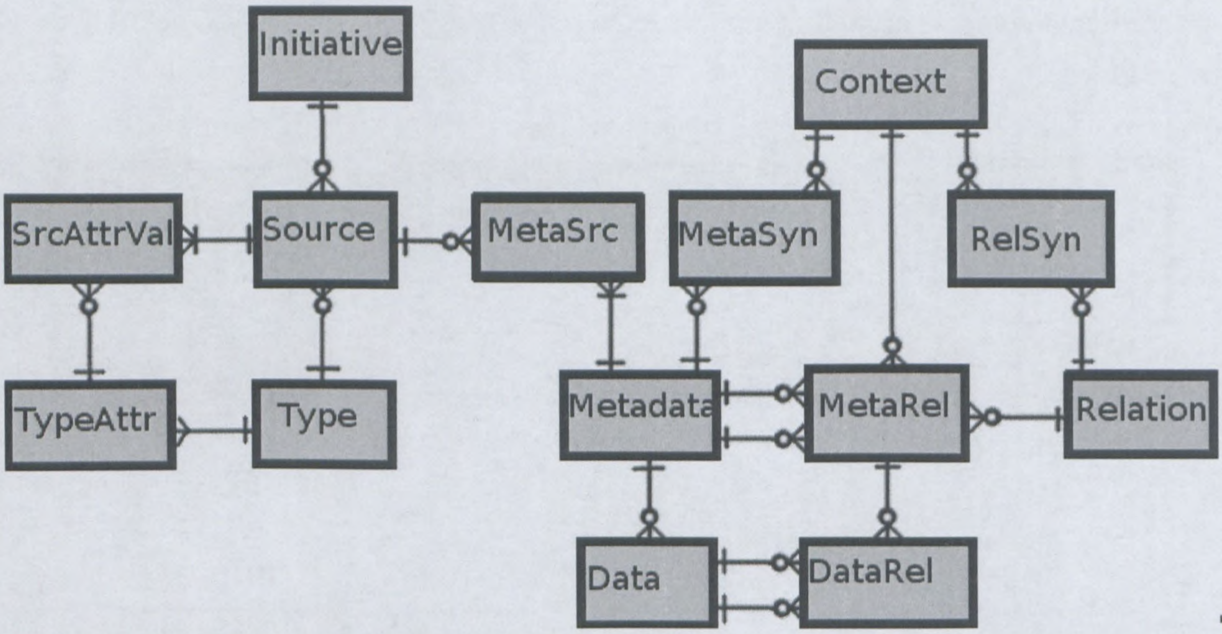


Figure 6.20: *Refined ERD, second implementation.*

Although the first implementation of the previous ERD was successful, once the system was being examined and tested it became clear that the system did not provide a detailed level of information necessary to gain the required insight into how the HBHC initiative used their data.

After the Implementation of the initial ERD it became clear that the context needed to be included. The conceptualisation of the context is discussed in more detail within this Chapter 4 under Section 4.1.2 and was seen as playing an important role in defining and understanding the data-elements. The context was added to the ERD and linked to the metadata synonyms and the relationship synonyms as well as to the metadata relationships. The context was added in the second iteration of development.

This section discussed the database structure used in the semantic metadata repository. The following section briefly considers the environment of the solution artefact.

6.2.3 Environment

This section very briefly touches on the environment of the repository solution. Most of the HBHC environment was discussed in Chapter 4, so within this chapter only the most relevant environment characteristics are discussed.

The initial prototype application was developed for a standard Windows based desktop PC (personal computer) with no need for internet access or networking capabilities to be in place to use the intended repository system.

The user of the system (previously referred to as the interpreter within this chapter) is not necessarily a member of the HBHC initiative, but would ideally work with members of the HBHC initiative. The role of the interpreter is to examine the capture the source materials (documentation, interviews with the caregivers etc.) into the repository and then to interpret these source materials and create the necessary database content (metadata, relationships, data etc.). The interpreter would also play an important role in the maintenance of the repository content and would along with knowledgeable individuals form the HBHC initiative help to ensure the quality and correctness of the repository content. In order for the interpreter to perform these functions, the interpreter is intended to be technologically literature, trained, individuals with some understanding of database design and training in the usage of the repository.

The advantage of defining that trained, IT-literature users are intended is that the initial prototype and the overarching research did not wish to heavily focus on the Human-Computer components of the system. Human Computer Interaction and designing for specific types of users is a fairly in-depth topic which could potentially constitute its own research, and thus was placed outside the scope of the research.

The intended users of the repository are not the HBHC initiatives themselves, since they will have little practical use for a tool which captures their own understanding and use of data. The intended users are third parties, who are interested in understanding the usage of data by HBHC initiatives. The repository allows for data to be stored for multiple HBHC initiatives thus allowing these third parties to gain an understanding of multiple initiatives.

These third parties could potentially be governmental bodies, researchers or software developers who are intent of developing software solutions for these initiatives. As discussed in Chapter 02 in Section 2.8, ontologies (which define the concepts and the relationships between these concepts) are useful in aiding software development. The repository does something very similar in that it defines the data-elements (the concepts) and their relationships to each other. The repository thus helps to create semi-ontologies for the HBHC initiatives that it captures data for.

Demonstration and Evaluation

Chapter 7 Instantiation

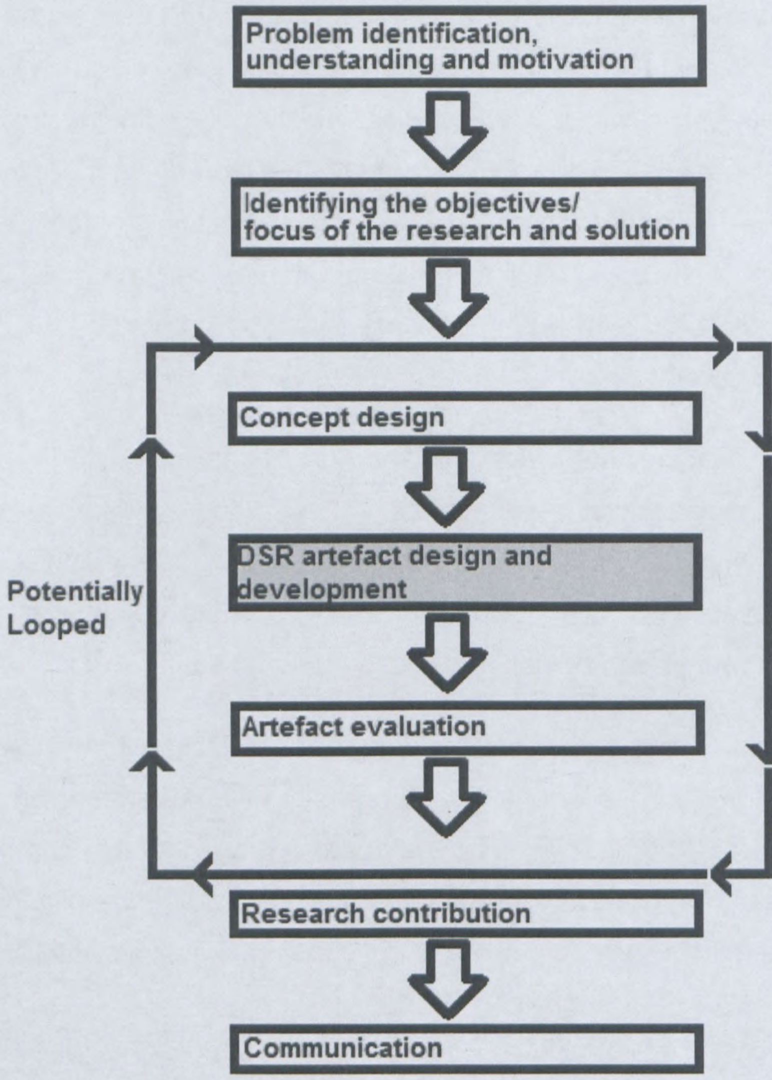


Figure 7.1: *Focus of Chapter 7 in relation to the methodology.*

This chapter considers the actualisation and instantiation of the DSR solution using the concepts and designs discussed throughout the previous chapters.

Following a DSR-based methodology the underlying development processes formed an important part of the research namely: the analysis and requirements/objective gathering processes form an important part in initial data gathering, while the design and development contributed in data gathering and creation as well as providing proof of concept.

Because the primary focus is usually placed on the analysis, design, development and evaluation it is thus conceivable that in some cases the presentation of the solution instantiations might not be included in the research or that it would most likely not be thoroughly covered. However within this research it was however deemed important to present the instantiation since it could constitute some form of contribution and the presentation of the instantiation aids in clarifying the research and the research findings.

This chapter follows logically from the previous chapters. Within Chapter 4 using an ethnographic study the problem area and the solution context was discussed and analysed and a deeper insight has gained into the research and solution problems. Within Chapter 5 using the results from the ethnographic study the objectives of both the research and the solution artefacts were identified. Within Chapter 6 the objectives were used to guide the design of the solution concepts and these solution concepts form the basis for the design and development.

The following section considers the user interface of the initial-prototype solution created as part of the research.

7.1 User Interface

The UI is the most visible portion of any IS system. Users hardly ever see or interact with the database or the hardware (such as the central processing unit (CPU) or hard drive disks (HDD) that support most IT-based system. The users usually do not need to know about things such as the different application layers or network topologies or programming languages in order to use a piece of software or interact with a system. This is because of the UI (if designed and developed correctly) allows to the user to effectively and inefficiently interact with all these components without knowing they are. The UI is the component with which the end users interacts and in most cases it is it is also the primary means by which the user can receive output from the systems.

The UI of a system and the functionality of a system are however two different but related concepts. Although most users, because they only interact with the UI, might think that the UI is the system but the UI is only a component of the overall system. A good system has all the required functionality and beyond that is very usable (the UI has a high level of usability). Although it is possible that a system might be able to do everything it is intended to do, but it might not be usable, i.e. the user might have difficulty using the system but the system performs and does as expected. The opposite is also true, a system might be lacking functionality or even have broken functionality, but the UI can be intuitive and very usable. Thus it stands to reason that both, the UI and the functional components are equally important in achieving a good system, but they are still fairly distinct (although both would be influences by environment and the underlying objectives and requirements of the system).

The research conducted was focused on the technical aspects of the design involved in creating the solution artefact. The research does acknowledge the importance of the user's ability to interact with the system, and even lightly touches upon it, but the core focus is on the technical design and development of the intended solution. Thus the UI developed during the initial prototyping was intended to be used by technical and trained users, limiting at least in part the need to create UI that caters to the user characteristics.

The UI thus has limited rigour and relevance and the researcher does not claim that it is perfect of correct, but the presentation of the UI is still important in presenting the overall developed prototyped solution.

The initial UI was intended to be developed quickly and was kept as simple as possible to ensure rapid development. The idea behind a simple UI was to test the ability of the underlying data structures to meet the intended objectives. The UI was created using Windows Form technology, which is fairly easy to develop rapidly because of the ability to drag-and-drop UI controls.

Ideally the UI would have been developed from the start using the intended WPF technology instead of Windows Forms. Because of the uncertainty of the underlying data structures and design concepts to meet the intended objectives a disposable prototype UI was developed, which would then later be replaced after any required changes have been made.

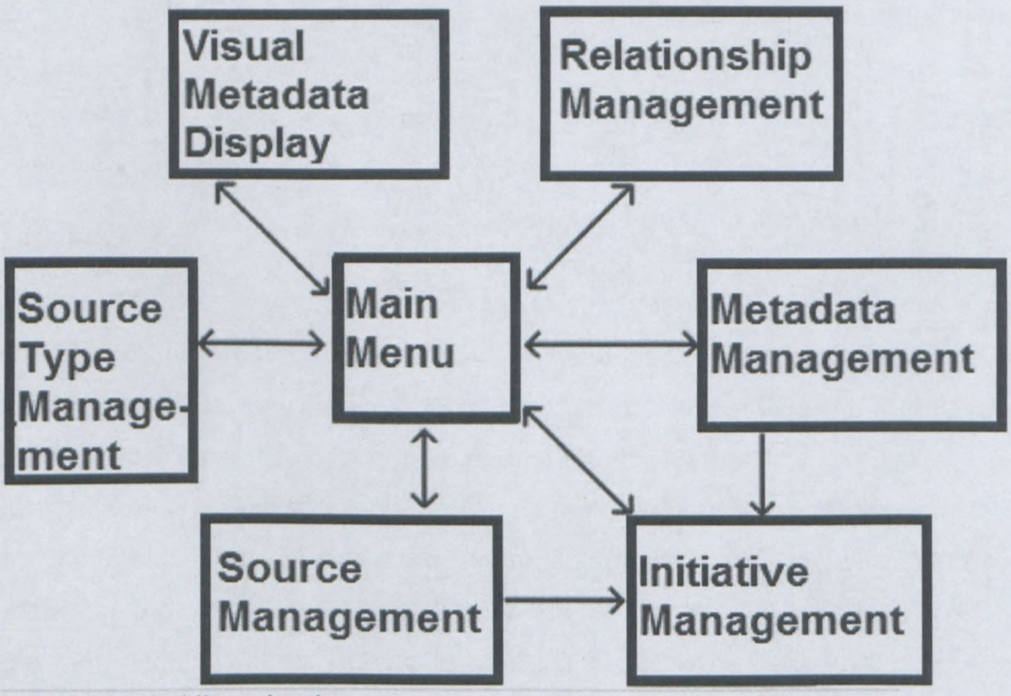


Figure 7.2: *Basic UI navigation.*

Figure 7.2 shows the navigation of the initial prototype application. The main menu is the first UI presented to the user and from it the user can navigate to any of the four main sub-UIs (institution, sources, metadata and relationship) each responsible for managing an important aspect of the structural metadata. Sources and Metadata during the conceptualisation was identified as being specific to a HBHC initiative. Thus an Initiative would firstly needs to be captured before the related metadata-elements and source materials; although because of the assumed generic nature of relationships they can be captured even without any initiatives present.

The Initial prototype application consisted simply of the Main Form and a number of sub-forms. These sub-forms are primarily intended for CRUD (Create, Read, Update and Delete) functionality with each form covering a specific data aspect. A single WPF form was created to visually display the data-elements and their relationships. The WPF form formed part of the intended goal of the repository not only to capture metadata-elements but also to display these elements in an informative way.

Most of the sub-forms follow a simply navigation structure, every Windows Form based sub-form has two buttons, Back and Exit. The Back button closes the sub-form and refocuses on the parent form (in most cases returning control back the Main Form); the Exit button closes the entire application. Obviously unless any changes have been saved these two buttons would cause the changes to be lost.

In the following sections the various the sub-forms that made up the prototype interface are briefly shown and discussed.

7.1.1 Main Menu Form

This section considers the UI of the main menu, the initial starting page for the application used to access the various other sub-forms.

The main form is the default starting form for the application and is thus the first screen seen when access the application. From here all the sub-forms can be accessed. It has little functionality other than providing a means for the user to access the other forms and their corresponding functionality.

Figure 7.3 presents a screenshot taken of the Main Form of the prototype applications.

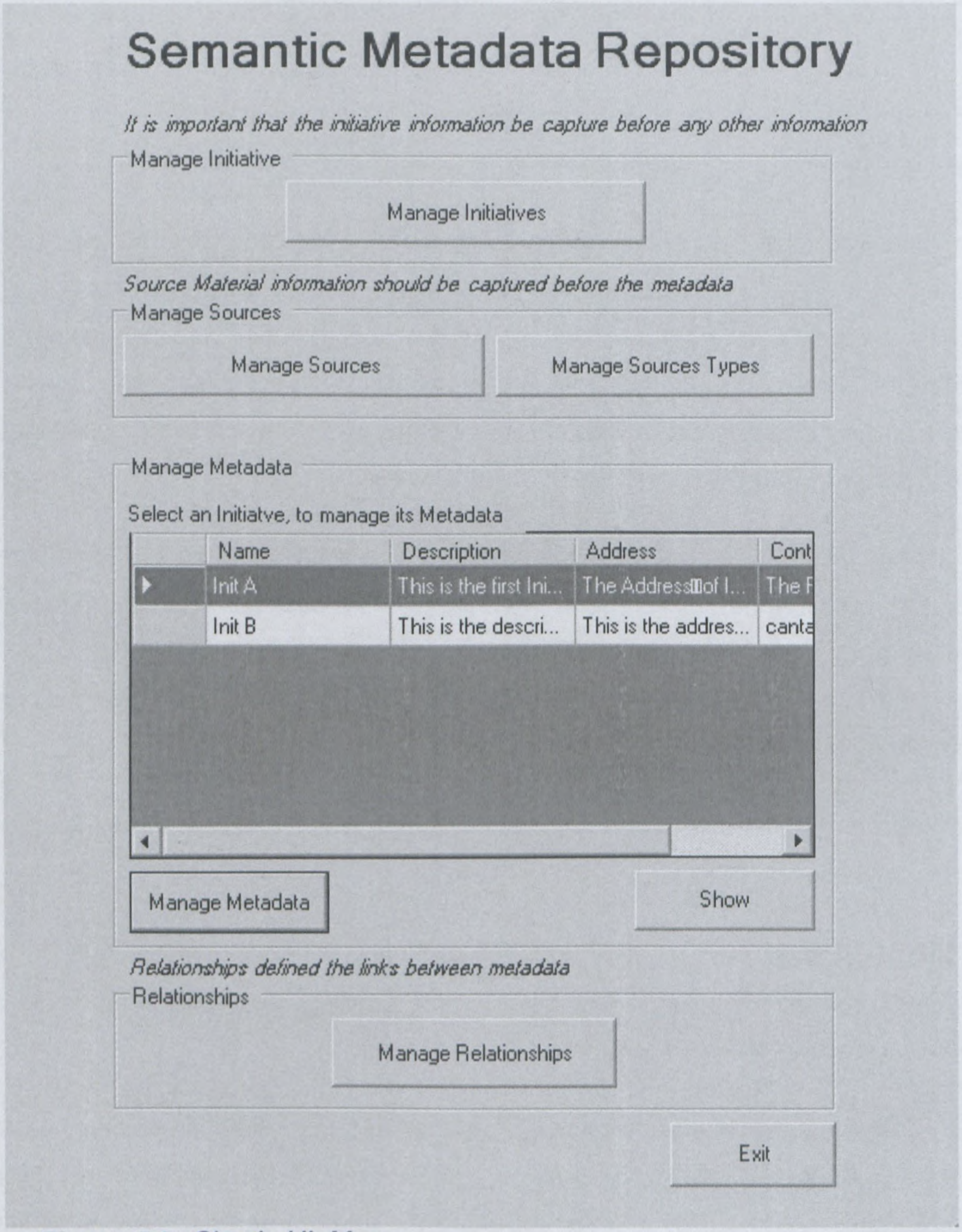


Figure 7.3: *Simple UI, Menu.*

The buttons that allow access to the various sub-forms are grouped together as: 'Manage Initiatives', 'Manage Sources', 'Manage Metadata' and 'Manage Relationships'. Each of these groups has a sort description accompanying them that is intended to aid in the usage of the sub-forms, although the application was intended to be used by trained users so these descriptions was done to help ensure effectively initial usage of the application.

The button groups are ordered from top to bottom in the order necessary to capture the needed data. Firstly the information for a given initiative must be captured, secondly the sources material originating from the initiatives needs to be captured, if necessary data about the type of the source needs to be captured before the source material themselves, third after the sources have been captured the metadata can be captured. Relationships information can be captured after or before the metadata information although conceptually the metadata would be captured first and then the relationships details and the metadata would be updated to include these relationships. Although one could capture the relationships before the metadata, but it was more important to first know which data-elements exists and then to define the relationships between the data-elements rather than create a number of relationships beforehand (which might be too generic or unnecessary). Although there is no limit placed on which forms can be accessed first (with the metadata forms being the exception because it is necessary to first select an Initiative from the list in the middle of the form).

The following sections displays and discusses the various sub-forms; the next section considers the initiative.

7.1.2 Initiative Form

This section considers the initiative sub-form. The form is intended to capture information related to an HBHC provider/initiative.

An Initiative or provider is conceptualised simply as an organisation or group which has the primary focus of providing HBHC services, such as the Stellenbosch Hospice or the Motherwell Health Clinic.

The intended goal of the repository solution was to provide a system for capturing the metadata and semantics related to the data-elements used by the HBHC initiatives in order to provide a tool for understanding how these data-elements are defined and used. Thus capturing comprehensive details for the initiatives was not deemed necessary as the initiative did not form part of the core focus of the application but rather aided in categorising the repository content.

Initiatives would usually be added to the repository but rarely ever modified or removed. Further it was not expected that the limited number of details captured for an initiative would frequently change. Removing the initiative from the repository would most likely also be accompanied by removing the initiative related and derived data (such as the source materials and the metadata items) which would remove content and thus purpose from the repository.

It is thus expected that usually the initiative sub-form would be used for adding new Initiatives to the repository or making small, infrequent changes.

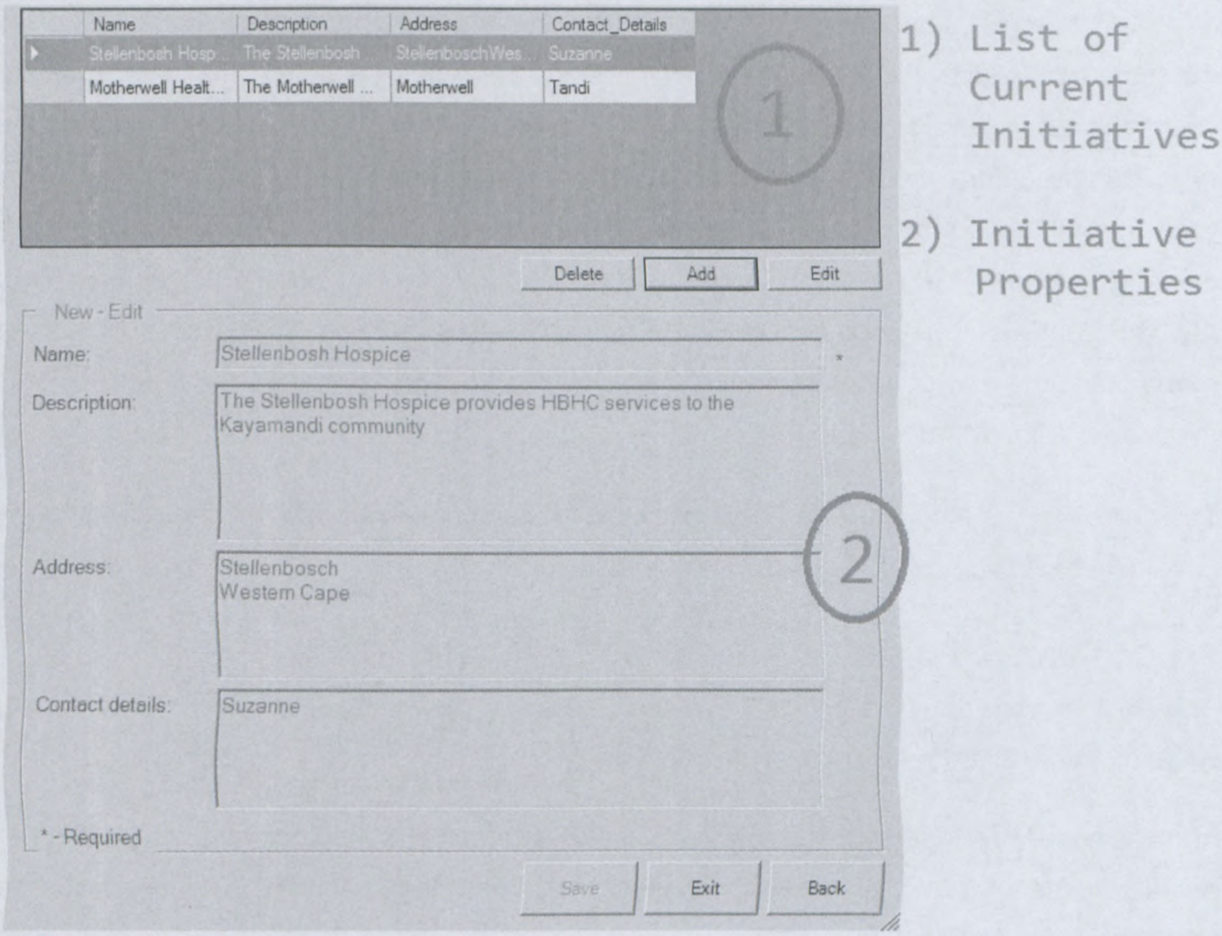


Figure 7.4: *Simple UI, Initiative management.*

Figure 7.4 shows the simple Initiative Management form, a sub-form within the prototyped application.

The Initiative form offers simple CRUD functionality, all the initiatives are listed and can be selected deleted, edited or a new initiative can be added to the list. The attributes chosen for a given initiative (name, description, address, contact details) were all chosen arbitrarily. The goal of the application wasn't to store and manage initiative information and as such only the most basic information was capture as such the only required field was the name of a given initiative to ensure that the initiative could at least be identified.

However in different contexts and within different requirements the chosen fields would obviously differ as well as what is considered required and what is not. It might also be possible that in a context the initiatives might have to be categorised or that information related to the types of services the initiatives provided might be required. But the more complex these data requirements and their relations to each other becomes, the more difficult it would be for the user to use and maintain.

The content of this form only contains the most basic information necessary, the name of the initiative in order to allow the users to identify the initiative, a description of the initiative in order to further aid in identification, the physical address of the initiative and contact information.

The following section will focus on the Source Type form.

7.1.3 Source Type Form

This section considers the Source-Type form used to manage information about the different types of source materials which can originate from the HBHC initiatives and which are used to derive the metadata-elements.

Within the conceptualisation it was decided that the repository was intended to cater for as many types of source materials as possible to not unnecessarily limit the number of source materials that could be captured and used to derive the metadata-elements. Only a handful of source materials were used (primarily document based) to create the repository. It however seemed clear that numerous other types could conceivable be used and that it would be difficult to cater for all the different types of details at would need to be stored for each type of source material.

In order to ensure the correct level of detail without forcing any possibly incorrect assumptions, it was decided to use an EAV-derived approach to allow the user to define the attributes of the different types of source materials. The EAV allows attributes to be dynamically created as discussed in Section 6.2.2 and shown in Figure 6.19. By allowing attributes to be dynamically defined it prevented having to make unnecessary assumptions about the source materials, such as the types or characteristics.

It is not possible to guess all the possible types of source materials. It is however assumed that eventually when the repository has been in use long enough that most of the different types of source materials would have been defined and the source type forms would only be used in the rare situation where a new type is discovered or changes need to be made to an existing type.

Figure 7.5 shows a screenshot of the Source Type Form used to create new types of source materials, from which the repository content can be derived.

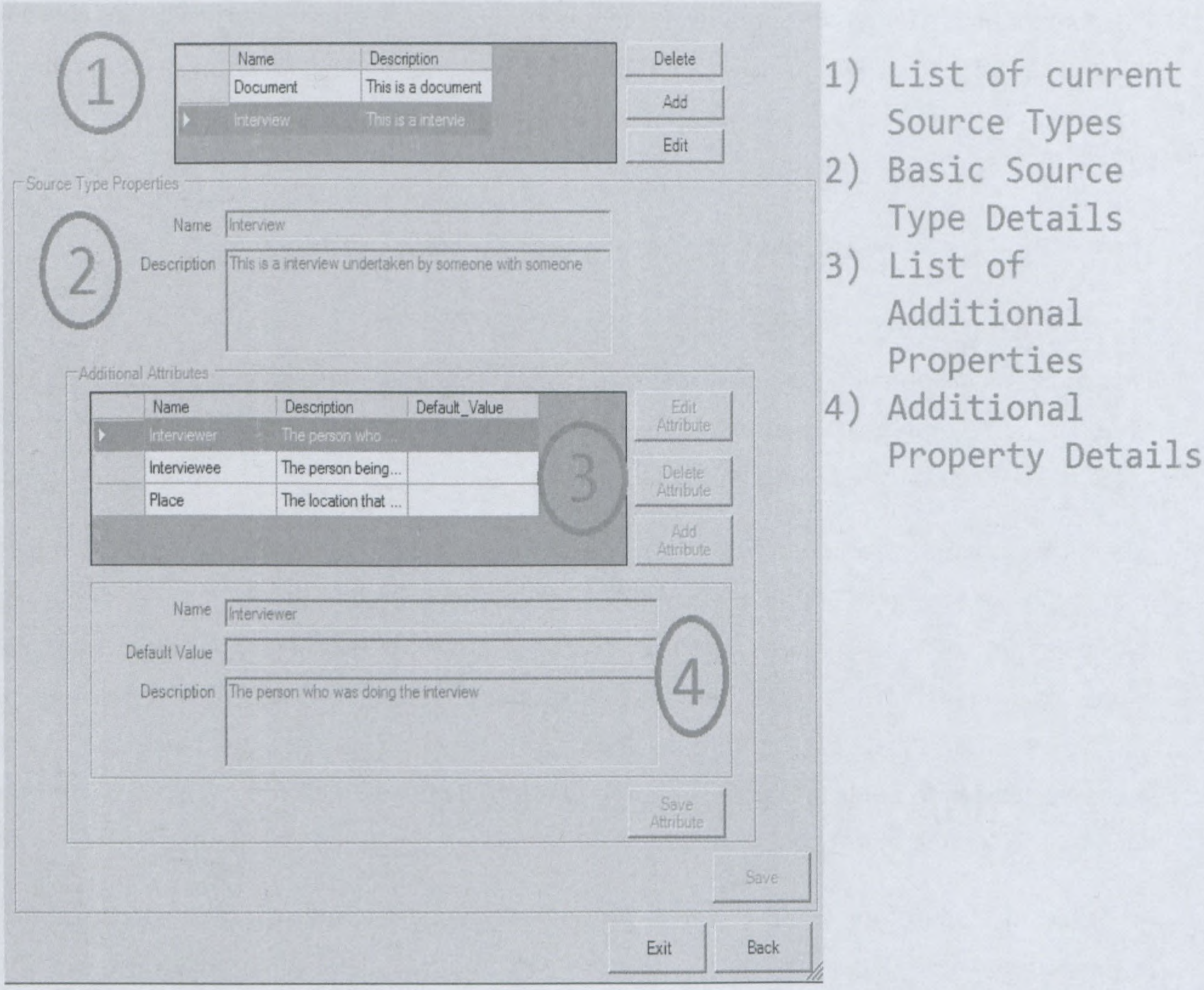


Figure 7.5: *Simple UI, Source Type Management.*

The source type form caters for the CRUD activities of the different types of source materials in addition it also allows for the definition and management of specific Source Type attributes.

The basic-properties of the source type consist of only a name and a description. The reason for this is the fear that defining too many basic properties might led to the inclusion of unnecessary properties that might not fit all the different source material types. For example a page number is relevant to document-type source materials but would not be relevant to pictures. An interviewee is relevant when an Interview (either written or audio) is used as a source, but interviewee is not applicable to reports.

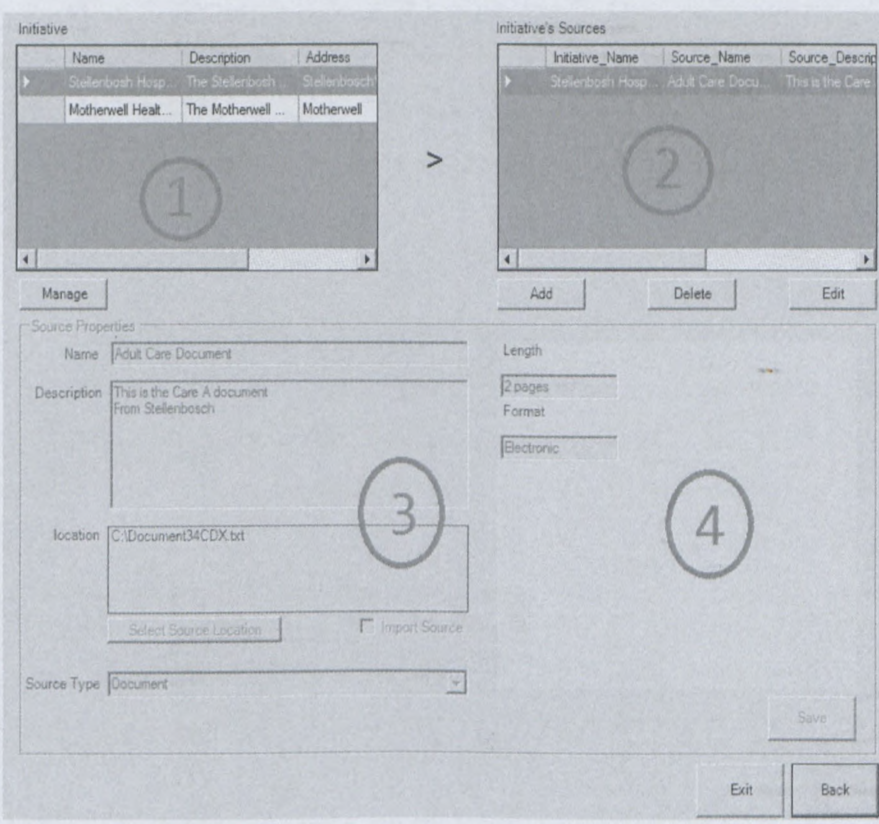
The Source Type data is defined into basic properties and additional attributes. The additional attributes can be defined specifically for a given source type, each additional attribute contains three basic pieces of information, name, default value and description. These are fairly self-explanatory, the name identifies the attribute, the description describes the attribute and if a user does not provide a value for it (when creating a source material) the default value will be used instead.

By providing the ability to define additional attributes which can be specific to a type of source material, it allows for the most relevant data to be captured while providing for unknown, future types of source materials.

The following section considers the Source Material form.

7.1.4 Source Form

This section considers the Source Form which caters for the management of source material information. The source materials originate from a given initiative and the source materials are in turned used to derive the metadata-elements.



1) List of HBHC Initiatives
2) List of Selected Initiative's Source Materials
3) Basic Source Material Details
4) Additional Details

Figure 7.6: *Simple, Source-Material Management Form.*

The Source Form allows for source materials for a specific initiative to be added and managed (provides the CRUD functionality). From the Source Form the Initiative Form can be accessed directly accessed without having to go back to the Main Form first. Selecting a given initiative from the list on the left, will automatically loads its source materials (if any is available) into the list on the right and allow for them to be managed by the application user.

A given source has two groups of data that can be entered for it, its basic attributes and its type attributes. Its basic attributes include name, description, location and type while the type attributes are user-defined within the Source Type Form and depend on the selected source type. The location value denotes the actual file that constitutes a given source material, that is the audio, image or text-file received from a specific initiative that will be used to identify and create the data-elements. A file must be selected and the location cannot be typed in, this is done to ensure that there is in fact a file associated with a source material record.

When a type is selected for the source material its type attributes will automatically load, the type attributes are specific to a given type (described a defined within the Source Type Form). These attributes all accept text values, although if a text value is not provided the default value specified for the specific type attribute is used.

As already explained previously for the Source Type Form the need for specific attributes for specific types of source material is because of the variety of different types of materials which can possible constitute a source material (ranging from voice interviews, photographs, text documents etc.) and the difficulty in defining common attributes for all the potential types of source material that can be used.

The following section considers the Metadata form.

7.1.5 Metadata Form

The metadata form is responsible for providing the CRUD functionality for managing the metadata for a given HBHC initiative. The metadata form also provides for the low-level semantics of the metadata-elements.

The metadata forms the primary focus of the repository solution, as each metadata item relates to a specific data-element within the HBHC initiatives. Most of the other UI forms provide additional information intended to ensure that the metadata-elements are useful and detailed.

It is thus envisioned that this form would be used most often for creation, updating and deleting metadata. As was seen during the ethnographic study, the data within the HBHC initiatives are constantly changing, thus the metadata will need to be changed to correspond to these changes as well.

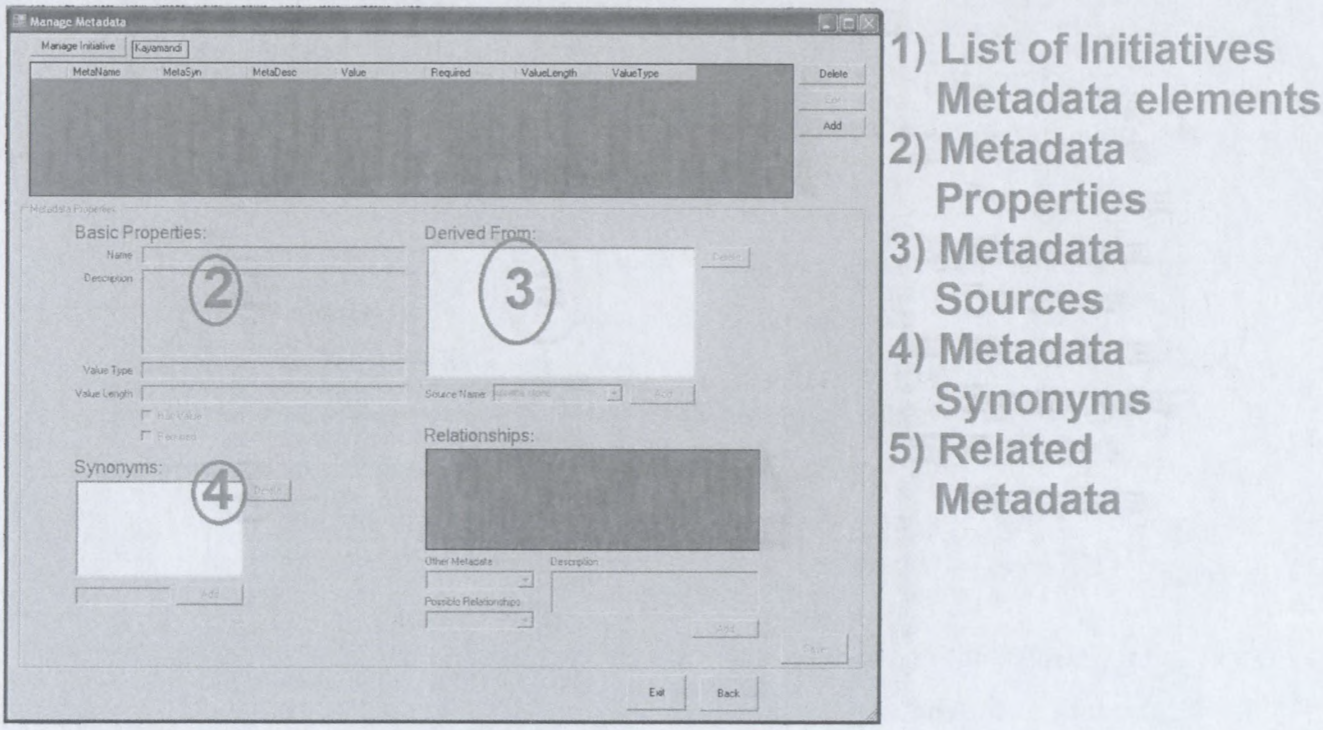


Figure 7.7: *Simple UI, Metadata-element Management.*

Figure 7.7 shows the Metadata Management form, a sub-form within the prototyped application.

The metadata form provides CRUD functionality for the metadata information. The form lists all the metadata that belongs to a given initiative, allowing for the metadata to be easily selected in order to allow it to be deleted or edited. The metadata form also offers access to the Initiative Form, discussed previously.

There are four groups of data that can be entered for a piece of metadata (as denoted by the labels) specifically: basic properties, derived from, synonyms and relationships.

Basic properties are the attributes which are common to all conceptualised data-elements and include: Name, Description, Value Type, Value Length, Has Value and Required. Name and description are fairly standards; the value type defines the type of value (string, integer etc.).

The basic properties aid in defining the type of data-element it is. If it has no data then it could simply be a category that other items belong it or some form of entity with relating items being the attributes or something equivalent. Although the intent was to keep it as open as possible thus providing properties that did not constrain the possible implementation negatively.

Synonyms (within the synonym portion of the UI) are simply a list of alternative names by which the metadata can be known. These are simply string values and can be added and removed for a given metadata item as needed.

Relationships denote the relationships between the metadata currently being added or edited and another pre-existing (within the application) metadata item. The two metadata items are related via a relationship. These relationships are one way, from the current metadata item to the specified pre-existing metadata item. Discussed more within the conceptualisation of the solution in Section 6.1.2 and demonstrated in Figure 6.9, the one-way relationships were purposefully chosen since numerous example were identified where the same relationship existed one-way but then different relationships in the opposite direction.

Derived from (within its respective portion of the Metadata UI) denotes the source of a given metadata item, as a means of ensuring the quality of a data-element the source must be specified that led to the creation and identification of the metadata item. The source material originates from a specific initiative and thus helps to link a metadata-element to an initiative.

The following section details the Relationship Form.

7.1.6 Relationship Form

This section considers the relationship form.

The relationships constitute meaningful relationships between two metadata-elements. It was conceptualised as necessary for metadata-elements to explicitly define the relationships between these elements as these relationships can potentially define how the data within a HBHC initiative is being utilised and what data is necessary to perform specific functions.

It is assumed that these relationships would be kept as simple as possible and thus would not be specific to a given HBHC initiative, only the way it relates the metadata would be unique to a given initiative.

Much like the Source Types it is envisioned that this form would be utilised mostly in the initial phases when identifying the data-elements and that eventually the repository would contain a robust enough collection that the Relationship Form would primarily be used for managing existing relationships, and adding the rare newly uncovered relationship type.

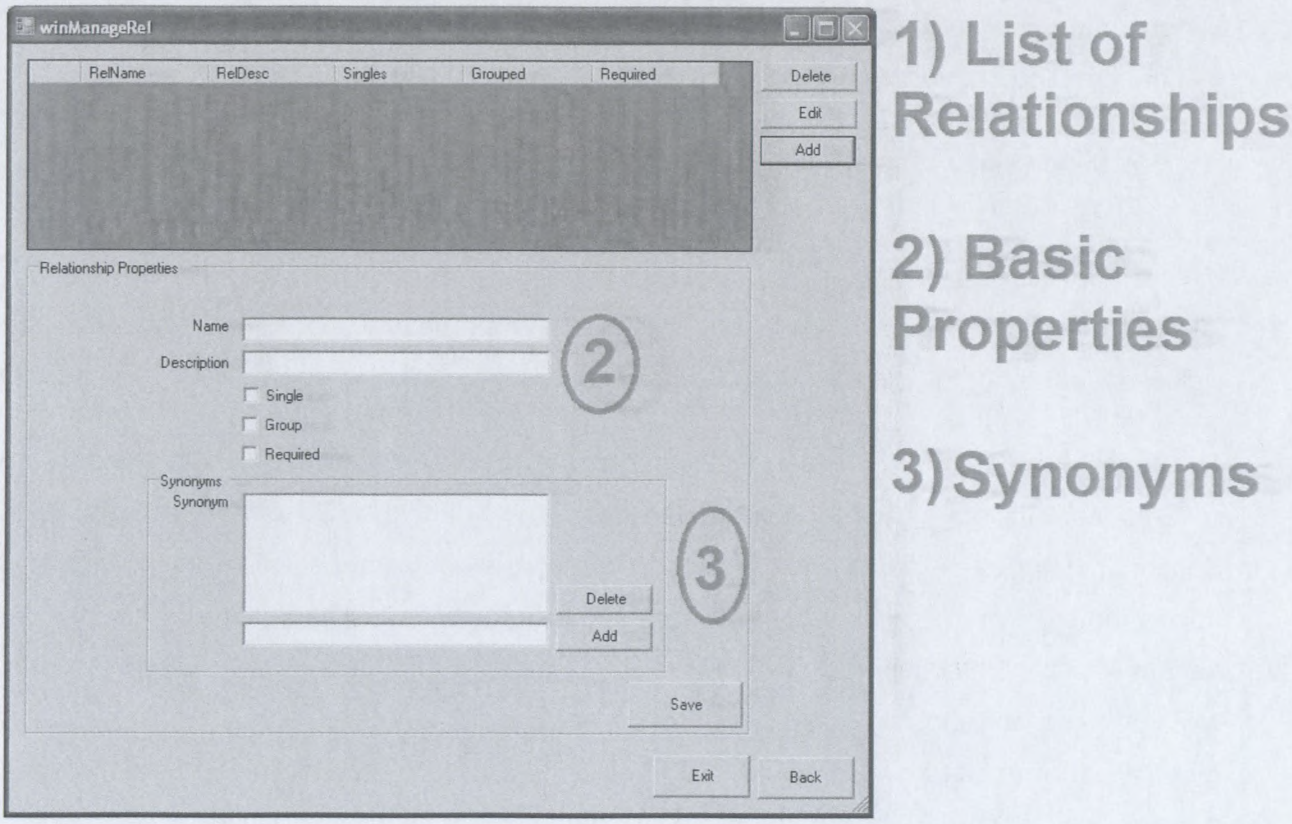


Figure 7.8: Simple UI, Relationships management.

The Relationship Form shown in the Figure 7.8 caters for the relationship information; it provides the CRUD functionality necessary for managing the metadata relationships. Like most of the other UI forms update and delete is accomplished by using the list, which is intended to show all the relationships defined within the repository application.

Relationships are one way, one element is related to another but it is not automatically inversely related. Although this does not automatically denote the reverse relationship (MetaA relates to MetaB via RelC but MetaB does not relate to MetaA via RelC), this is done because relationships are seen as being generic enough to not be related to a given initiative and exceptions exist when explicitly implying an inverse relationship.

The relationship information is placed into two groups, basic information and synonyms. The basic information consists of Name, Description, single, group and required. The Name and Description are simple string based textfield which provide identification and a description of the relationship. As discussed within the Chapter 5 within the conceptualisation, the Single property denotes that the relationship between the two metadata-elements only involves the two elements while the Group property denotes that the relating elements collectively relate to the metadata item (i.e. a person name consists of first and last name combined). The Required attribute denotes that the relating elements must both be present (if the one appears than so must the other).

Synonyms much like the synonyms of metadata-elements are simply a list of string values that denote different names by which the relationship can be known.

The following section discusses the metadata visualisation UI.

7.1.7 Visual Metadata Form

This section considers and describes the Visual Metadata Form which visualised the repository content.

The Visual Metadata Form was initially intended to provide visualizations of the repository content. As detailed in the literature review under the heading of Semantic Repositories in Section 2.4.1 an issue arises with repositories such as the intended solution, as more data is added the system becomes more complex, not only because of the volume of content but because of the potentially complex relationships between the repository content. As data was gathered and inputted into the repository it became clear that simply looking a list of metadata (not matter how complex or simplistic) did more to hamper the understanding of the repository content. Visualisation offers of means of showing potentially complex data in a way that it is easy for a human user to interpret.

The Visual Form is also thus a key component in achieving the goal of the metadata repository in that it will aid in understanding the usage of data-elements within HBHC initiatives by presenting the complex data-elements and their relationships in such a way that it is easy to gain an insight into the usage of data-elements and their relationships.

The visual metadata form unlike the previous forms is not based-on the Windows Form technology on Extensible Application Markup Language (XAML), specifically it was created using Windows Presentation Foundation (WPF) technology. WPF is used for creating graphical interfaces and has the advantage of providing in-built support for developing 2.5 dimensional graphics (two dimensional graphics within a three dimensional space), allowing for a more visual presentation of the UI content. As seen from the previous sections Windows Forms is primarily based around using and extending existing controls (textboxes, labels, button etc.), it does contain the ability to draw two dimensional graphics but from previous experience WPF is a far simpler solution.

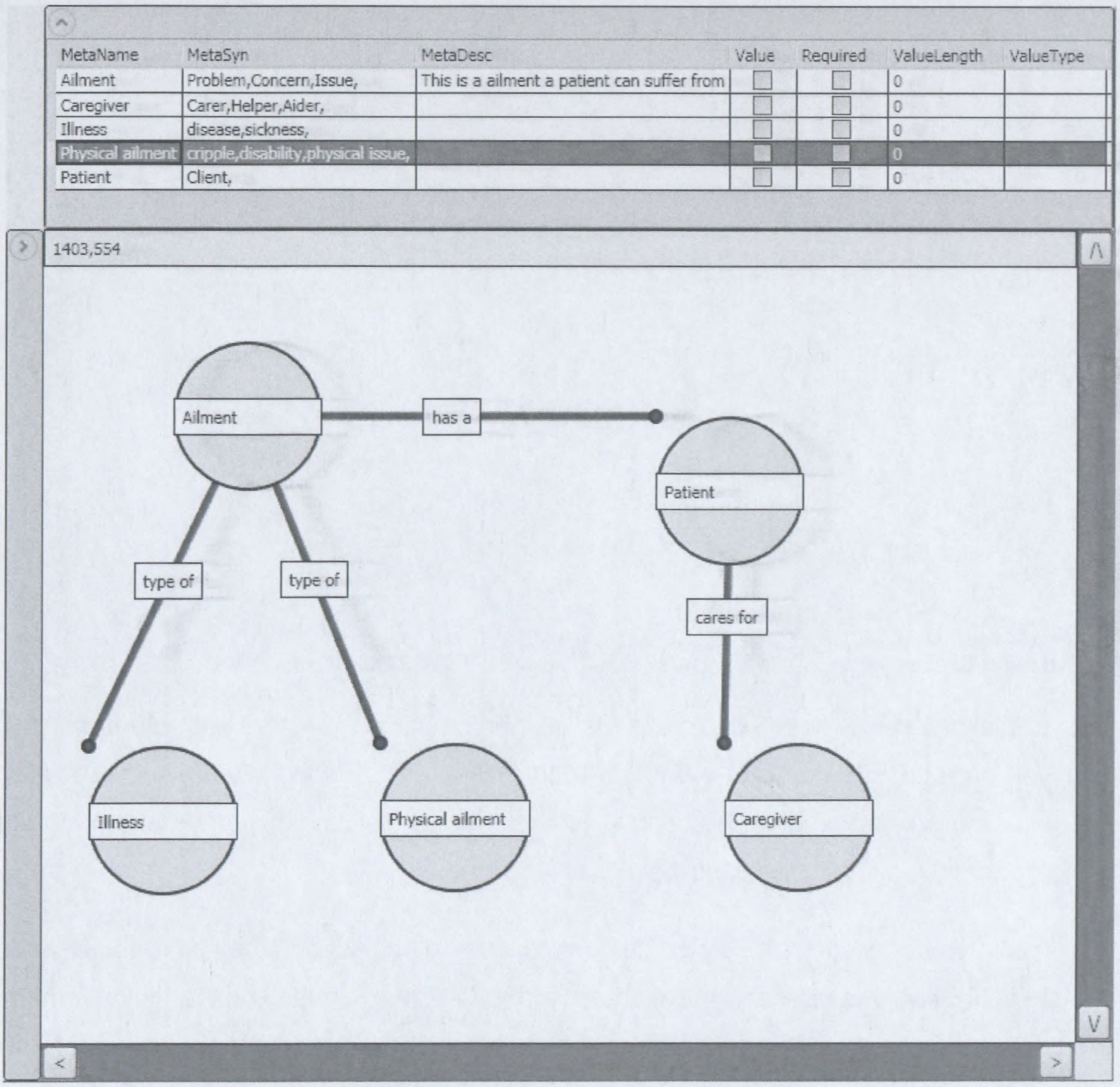


Figure 7.9: *Visual metadata UI.*

The initial development was done using windows forms as a proof of concept, once the underlying concepts were proven to be adequate, the development is to continue but replacing the prototype forms with WPF forms. Although the possibility did exist that rather than replacing the existing forms with exactly the same forms, only in WPF, that the functionality of the other forms could be incorporated in the Visual Metadata Form, although that is only intended to take place in future iterations.

The Visual Metadata Form lists all the metadata for a specific initiative (which was chosen on the Main Menu Form before opening the Visual Metadata Form). These metadata-elements can then be loaded into the 'view' area, as seen in Figure 7.9 the yellow circles denote a given metadata item. When multiple items have been loaded and a relationship exists among them, a blue line will be visible on which the name of the relationship appears.

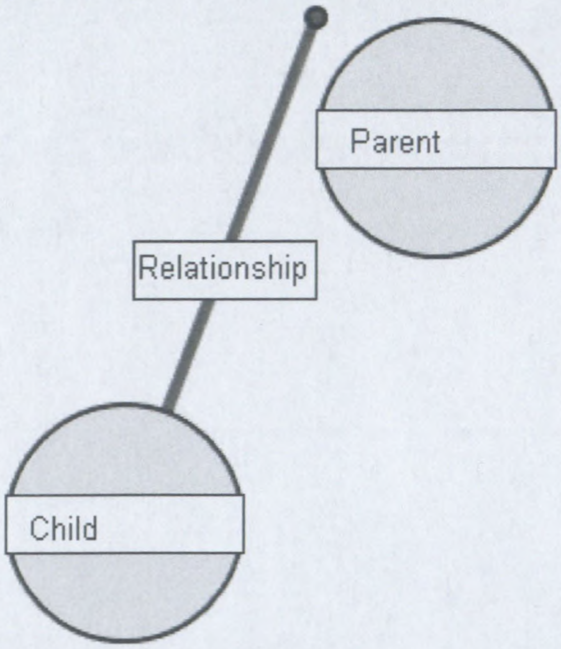


Figure 7.10: *Example of a Metadata-Relationship-Metadata visualisation.*

Because relationships between the metadata-elements (which represent data-elements at a meta-level) are one way, concept of a parent (or relater) and a child (related) exist between two relating metadata-elements. As seen in Figure 7.10 the relationship starts at the top left of the parent and leads into the centre of the child element.

Being a prototype application however the context was not included in the initial visualisation form but ideally would appear much like shown in the Chapter 5 in the concept development. Currently the metadata-elements can be arranged around the display area by simply dragging them around with the computer mouse. Ideally of course some form of automatic arrangement or placement would be used, although this would be a daunting task primarily consisting of trial and error and would require the more active involvement of end users to ensure that whatever layout is selected is most adequate.

7.2 Code Class

This section considers the implementation of the DAO (Data Access Objects) responsible for managing retrieval of data from the database to the Entities and persistence of Entity data into the database and the Entity classes that temporarily store the data for the application. This section will also lightly touch on the Display Classes, the classes that correspond to the entity classes but provide formatting for the Entity class data in order for the data to be presented on the UI.

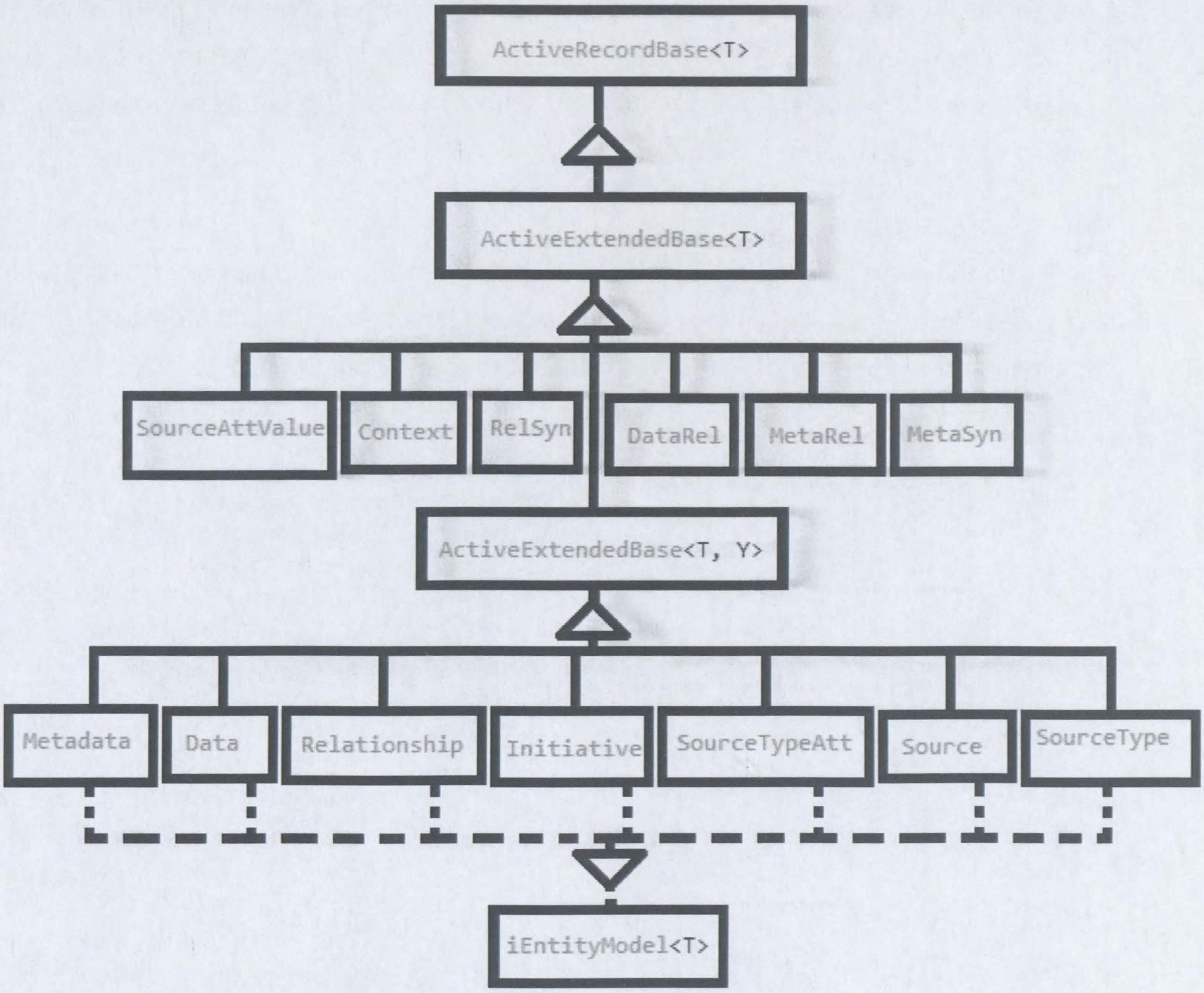


Figure 7.11: *Class Hierarchy.*

The Castle ActiveRecord Framework, which is an implementation of the Active Record pattern for the Microsoft Dot Net Framework, was used to provide for data access and persistence to the underlying SQL Server database. Castle ActiveRecord provides many of the basic DOA functions (such as deleting data in the database, persisting/saving to the database and reading data from the database). The ActiveRecordBase class from the Castle ActiveRecord framework provided this generic functionality but this class was extended to provide functionality more specific to the current application (for example the application and database uses Global Unique Identifiers (Guids) to identify unique data entities, the ActiveRecordBase class was extended to allow retrieval and deletion of data using Guids).

To prevent an Anaemic Domain Model the data layer did not separate the DAO from the Entity classes. No separate controlling class (DAO) was created for each entity; rather two generic DAO classes were created one for entity classes that had a corresponding display class and one for entity classes that did not have a corresponding display class. The Entities themselves inherited and extends these DAO classes for methods specific to the Entity (for example functionality to find all the derived metadata for a given source or to find the source of a given metadata item). The Entities that have a corresponding display class also implements the iEntityModel interface which defines a method for converting between the Entity and the Display Class.

Not only did combining the Entity and DAO classes aid in increase the speed of development (usually a separate DAO class would have been created for every Entity class) but it also aided in overall maintenance (since without inheritance it would have been difficult and time-consuming to update the different DAO classes). Combining the DAO and Entities aided in simplifying the overall structure of the application classes.

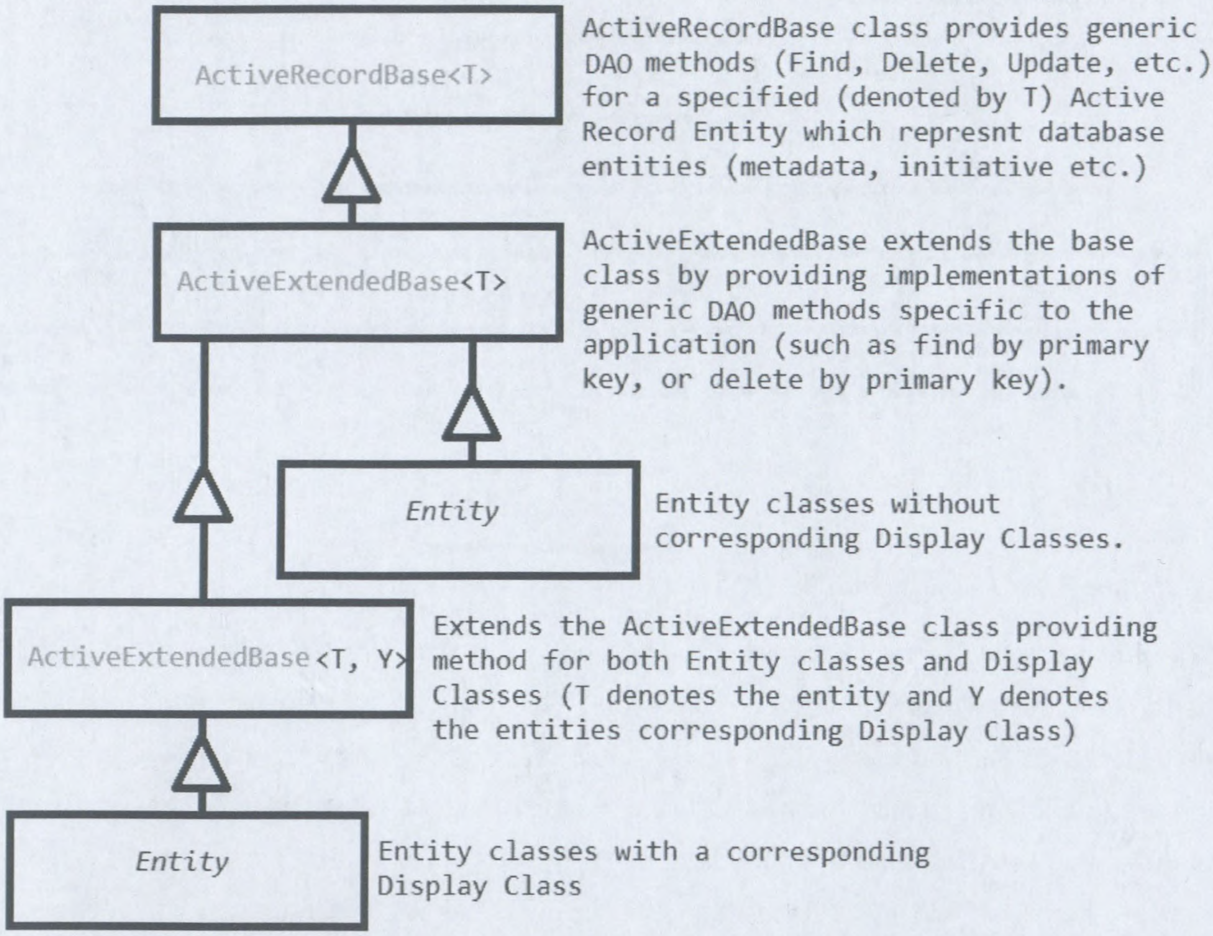


Figure 7.12: *Clarification of class inheritance.*

Figure 7.12 provides a quick overview of the non-Anaemic Domain Model used in the repository application. In this model the entities have all the necessary code and functionality to manage themselves (and do not have to rely on external classes to provide these functionalities). The Entities thus are not Anaemic (which can be the case when implementing a Domain Model where the Entities are little more than a collection of property Accessors, the getters and setters).

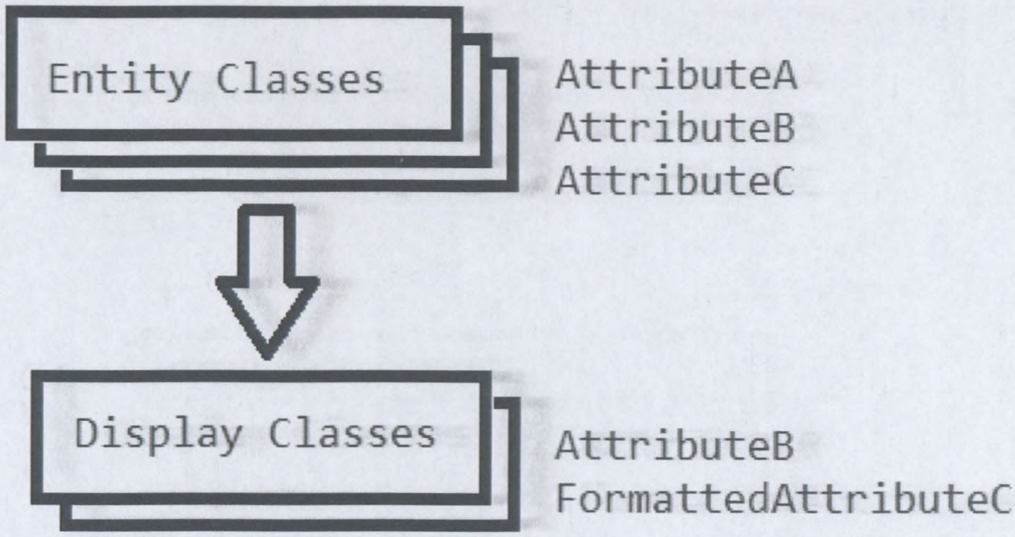


Figure 7.13: *Entity / Display Class examples.*

Some Entity Classes have corresponding Display Classes. These Display Classes are intended to be used to represent information retrieved using the Entity/DAO classes to the user in the UI. Some information stored in the database and thus within the Entity classes are not important for the user to know in order to use the application, information such as database primary keys (which uniquely identify database records) or database foreign keys (which identify the relations between database records within). Some information is also more useful and meaningful when it has been formatted, such as Date and Time values. The Display Classes thus hide unnecessary information, such as Database Primary Keys and Foreign Keys, and formats less presentable information, for example formatting DateTime attributes to make it more presentable.

Many of the UI controls that come standard with the Dot Net Framework allow its content to be automatically generated based-on what data is given to it (such as Datagrids and Combo-boxes). Using Display Classes the data from the database (via the Entity classes) can be formatted and provided to the standard Dot Net controls to aid in populating the UI.

Chapter 8 Evaluation

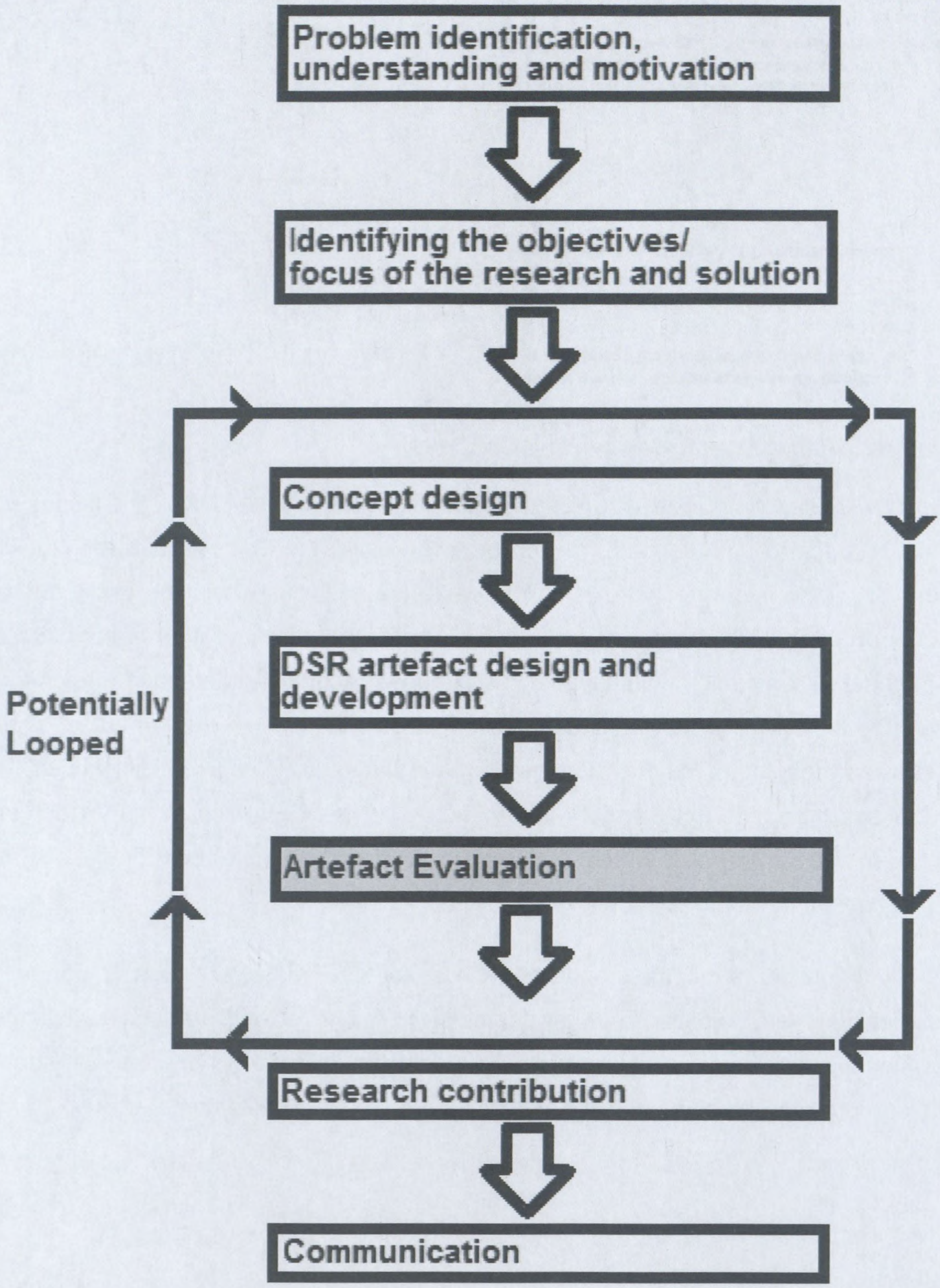


Figure 8.1: *Focus of chapter 8 in relation to the methodology.*

As has previously been discussed within Chapter 3 in Section 3.7 and touched on within Chapter 2 in Section 2.7, a number of DSR artefacts are developed as part of the DRS research process. The process of developing the DSR artefacts is detailed in Chapter 4 through Chapter 6 which covers the initial exploratory research and research problem, the concept design and the solution design and development and the final instantiation of the artefact is detailed in Chapter 7.

Within design-based research the DSR artefacts are developed as research tools and the evaluation stage of these DSR artefacts in design-based research is equivalent to the data analysis stage in other forms of research, because these DSR artefacts are viewed as being knowledge containing. DSR artefacts are seen as knowledge containing because: they are constructed with a specific goal in mind; they contain knowledge and insights of the designers and creators as well as a number of assumptions spanning a variety of aspects such as the intended usage, context of use and process of construction. The evaluation aids in assessing if the knowledge was relevant and correctly applied and if the assumptions were correct or not.

This chapter covers the evaluation stage of the research. The evaluation will specifically focus on: the overall solution artefact (the repository) and will touch on several of the sub-components. In order to manage the bias involved in the evaluation process, where possible both the positive and negative components of any specific evaluation criteria is presented and discussed.

The following section will focus primarily on the evaluation of the overall solution artefact.

8.1 Solution Evaluation

The solution is currently still in its prototyped phase and thus certain aspects still need more refinement and development (specifically the UI and interfaces to external systems). Some effort was made to ensure that the core ideas and concepts were incorporated within the prototyped application as the prototype was intended to serve as proof of concept. The conducted research was also structured in order not to focus on these incomplete components. These components were not the primary focus not only because they required more work, but also because these components themselves could potentially constitute their own research, for example the process of designing, prototyping and refining the UI and developing or implementing standards for interfaces with external systems.

The evaluation of the solution considers the underlying functionality and technical aspects. As has been previously been discussed, the research assumes that within a given application the UI components and the underlying functionality both play equal roles in ensuring that a software solution is effective and efficient, but that these two components are separate and can, to a certain extent, be separately evaluated.

The metadata model is evaluated with the most relevant of the NISO requirements for good metadata specifically: the metadata must be appropriate for intended us; include a clear statement of the intended condition and terms of use; must support long term management; metadata records are objects and must conform to the requirements of quality objects: authority, authenticity, achievability, persistence, and unique identification (NISO Press, 2004; Park, 2009).

The repository itself was measured along Thibodeau's (2006) axes for evaluating a digital repository: orientation, coverage, collaboration. Although Thibodeau (2006) defined three broad categories of evaluating digital repositories, no specific or detailed criteria for measuring the repository was provided meaning a more quantitative evaluation is not possible, but the three axes can serve as a topic for qualitative discussion and evaluation.

The following section considers the evaluation of the metadata component of the repository solution.

8.1.1 Metadata

This section considers the evaluation of the metadata model used in the repository solution which was developed as part of the research. The repository was created to explicitly capture structural semantic metadata and thus the metadata forms a core component of the repository solution.

The literature review that was undertaken as part of the research, detailed in Chapter 2, provided a significant and varied amount of information on the topic of metadata and also clearly showed that there was no single definition, type of usage for metadata. Because of the variety of metadata types and usages found during the literature review, the literature could not provide any universal criteria for evaluating metadata.

The NISO provided a number of requirements for good metadata, although not all of these requirements were relevant to the type of metadata used in the research, thus only the most relevant of the NISO requirements were used to evaluate the solution's metadata component.

The following section considers these relevant requirements for good Metadata provided by NISO.

8.1.1.1 Appropriateness

This section evaluates the metadata according to its appropriateness, the ability of the metadata to meet the intended requirements or enable the necessary functionality. Appropriateness is one of the requirements for good metadata defined by NISO.

Table 8.1: *Metadata Evaluation, appropriateness.*

Appropriateness	Issues faced
Offers a means of creating pseudo-ontologies and thus representing the data-elements, their relations and contexts.	Although appropriate there might be a better solution to capturing, storing and representing the data-elements, relations and contexts. A technological due-diligence was conducted and no alternatives were found but it is still possible.
Represents the data on a meta-level and thus is not affected by changes in data values but only by changes in data usage and data definition.	As data is actively used its definition, structure and relations might change, no feedback currently exists from the data level to the metadata level.
Metadata model provides a generic model in which to capture the metadata information, being generic increases the potential data which can be captured.	Being generic it might be inappropriate for specialised types of data, for example procedural information or best practices.
The Metadata model can easily be visualised to aid in representing and understanding the data-elements.	Visualisation is not inherent part of the metadata model and how the data can be represented is not standardised.
Source material information is also included in order to ensure that the lineage of the derived metadata can be traced, aiding in management and understanding.	Source material analysis is still a human activity and even by including the sources and lineage errors can still occur. Although automatic extraction is potentially even more problematic.

One of the problems that gave rise to the research was the lack of standards for data-elements, specifically within HBHC. Initially the research had attempted to look at a way of enabling interoperability, potentially through the usage of metadata which could define the ontologies and overcome the issues surrounding semantic and ontological inconsistencies which inhibited successful interoperability. Although it soon became apparent that the issues faced by rural South African HBHC ran deeper than simply enabling interoperability, namely: a lack of basic IT based solution and the necessary infrastructure to enable IT-based solutions as well as the prevalence of paper-based systems and the lack of standardisation of HBHC data. These listed issues first need to be addressed before any research and effort could be made to enable interoperability at an IT and IS level.

Not only did the lack of standards give rise to the research but also the differences that existed in the definition and usage of data-elements between different HBHC initiatives and the daunting task of attempting to standardise all these different definitions and usages. It was decided not to attempt to gain an insight into all the data usages across all the rural HBHC initiatives and neither to attempt and purpose some standard for HBHC data which may or may not fit all the possible HBHC contexts. It was rather decided to rather create a mechanism which would enable the effective capture of these data usages and definitions. By providing a means of capturing these definitions other parties could potentially undertake the process of standardization or by having a means of understanding the data-elements other parties could potentially implement IT-based solutions.

From the literature review it was found that ontologies provided a means of representing the data and the data relationships within a given domain and exist explicitly or implicitly in most systems whether paper-based of IT-based. From the literature findings metadata was seen as a way of creating a pseudo-ontology, the metadata could be used to define the HBHC data-elements as well as their relationships. Using metadata was seen as the most appropriate solution to understanding the data-elements within HBHC and a structural metadata model was created specifically to solve this problem.

The metadata connects the identified data-elements to specific source material originating from a given HBHC initiative in order to ensure that derived data-elements can be traced back to a given source. This connection of data-elements and sources helps to ensure that a given source was interpreted correctly to create the data-element i.e. a user is able to go back to the source material(s) and ensure the source material was correctly interpreted and that the data-elements were derived correctly. This is important because automatic data harvesting was not deemed the most effective solution, because of the issues identified within the literature detailed within Chapter 2 in Section 2.3.6 under the heading of Metadata Creation.

The purposed metadata model however does have certain shortcomings.

It was intended to be fairly generic in order to be used to explicitly define a generic ontology used by a given HBHC initiative and was further created using paper forms from real HBHC initiatives. The generic nature might be a potential shortcoming since the model might be too generic to capture specialised data and the paper forms used in creating the metadata model might not have been sufficient to provide a global view of all data-elements in HBHC. However, this is debatable of course since it is fairly hypothetical. The metadata model was able to successfully capture the data-elements presented by the paper forms used by these HBHC initiatives and not enough evidence was present to bring into question the effectiveness of the metadata model. However it is difficult to say with absolute certainty that the purposed metadata model is perfect for all possible situations and might have to be changed, or even made more specialised to cater for certain, as yet unknown, types of data.

There is however certain types of information that cannot be captured easily using the metadata model such as: procedural information and best practices. Although the metadata model could capture this type of information to a certain degree, a lot of potentially important information might not be explicitly captured. It was however not the focus of the research or the development process to cater for this type of information.

From the literature metadata was deemed the most effective for the intended purpose of capturing the types and relations of the data-elements used in rural South African HBHC. The metadata model is somewhat limited in certain aspects such as its ability to effectively capture certain types of information. However the use of metadata to solve the intended solution problem was deemed most effective, barring any unknown alternative but none was found in the literature.

Table 8.1 lists the factors which make the usage of metadata and the metadata model appropriate in achieving the intended goals of the Metadata repository, but Table 8.1 also lists some of the issues. By providing both the pros and the cons a more balanced evaluation was conducted, but most of these issues were subjective and in the realm of unknowns such as that there might be a better model to represent the data-elements (although none were found), or that there might be types of data-elements the model cannot cater for (although none were found).

The following section details the terms of use for the metadata.

8.1.1.2 Statement of terms of use

One of the requirements for good metadata, according to the NISO requirements used to evaluate the metadata, is a clear and explicit statement of the intended terms of use.

The terms and usage of the metadata is fairly clear, if somewhat technical. The metadata is used to capture the structural and semantic data for a data-element which are derived from source materials originating from rural South African HBHC initiatives. The metadata is used to assign additional meaning to the core data (the data-element) such as: how it relates to other data-elements and in what context these relations occur and by what name a given data-element within a given context is known (i.e. medicine can be a treatment when given to a patient, a prescription when being prescribed to a patient or it can be stock when being stored).

The intent of the metadata is to enable the capture and storage of these structural metadata for a given data-element which can then be used along with other data-elements and related metadata to gain a deeper understanding of how they are being used. This in turn could be used to develop some sort of information management solution or develop standards for the data-elements being used in rural HBHC.

The statement of use for the metadata is perceived to be clear enough, but it does have a fairly technical component to it which might cause the intended terms of use for the metadata model not to be immediately obvious to none-technically minded individuals. As such the users of the metadata model and the overall solution artefact (the repository) was intended to be technically minded or trained individuals.

Most likely the only reason why a clear statement of purpose is needed is because it allows the individuals involved in creating the metadata to give specific and explicit thought into the purpose and goals of the metadata, indirectly helping to guide the development process. However in order to conduct the research it was necessary to explicitly state the terms of use, thus as an implicit part of the research this requirement was already met.

The following section evaluates the metadata used in the solution artefact according to its long term manageability.

8.1.1.3 Long Term Management

Another of the NISO requirements for good metadata is the metadata's long term management. Long term management is of varying importance in different types of systems that actively use metadata, within a system that stores metadata that regularly changes. Long term management is of a lesser importance, but within the repository solution created as part of this research long term management plays an important role in ensuring the quality of the metadata.

The Table 8.2 demonstrated the evaluation of the metadata model in terms of its long term management. Specifically the following criteria are looked at: inclusion of source material information, metadata visualisation and simplified metadata information and relations.

Table 8.2: *Management Characteristics of the Metadata.*

Long Term Management Characteristics	Addressed Metadata Management Issues
Inclusion of source material information	Updating metadata-elements based-on changes to the definition and usage of the HBHC's data (reflected in changes to source materials). Tracing the lineage/origin of a given derived metadata-element, and thus ensuring that the derived metadata is correct.
Metadata Visualisation	Working with a large volume of complex data. Providing the necessary level of understanding of the metadata-elements in order to manage the metadata.
Simplified metadata information and relations	Potential complexity in managing a large volume of complex metadata-elements.

The metadata model created in this research was intended for long term use and thus had an implicit requirement to enable long term management. It is envisioned that although the data-elements and their associated metadata will change overtime, these changes would be limited to only a few times a year and that these changes would not affect a significant portion of the metadata-elements at any given time. So although the initial entry of the data-elements' metadata might take a while, the management that occurs afterwards would be infrequent and potentially quick. The management however is still important none the less because it supports an important goal of the metadata model, specifically that the metadata model represents the current state of the data within a given HBHC initiative. As the data within the HBHC initiative changes the metadata model must be changed to reflect these changes.

The metadata would obviously be most useful if a large enough number of metadata-elements are present within the repository solution; in order to get the most detailed insight into the current state of the data-elements and their relations. Some issues might arise with the long term management in this regard because as more metadata-elements are added to the repository the overall model becomes more complex and more difficult to manage. To ease some of this complexity the metadata items were specifically segregated together based-on the HBHC initiatives from which they originate. But even if the metadata is being separated based-on their originating HBHC initiatives, it is still possible for a single HBHC initiative to have a large number of data-elements with complex relationships and usages.

Another way that the metadata attempts to ease the burden on the user who will be responsible for managing the metadata, is that the metadata is kept minimalistic. The metadata-elements are kept as simplistic as possible; the fear of course is that if the content is too complex the user will be more likely to make mistakes or it would be too difficult to maintain. However context was not included in the metadata model to help ensure that the metadata would be kept as simple as possible, context was eventually added in order to ensure that the metadata was more meaningful. Simplicity in the metadata thus needs to be balanced, since if it is too simplistic it might not be meaningful and if it is too complex it might be too difficult to manage.

Overall however it is presumed that since HBHC is a fairly specialised field catering for only a specific form of care and offering only a limited number of services, the number of data-elements which could be identified and the corresponding amount of metadata that could be generated would be of a manageable volume. The purposed metadata model would however suffer from this complexity if it was implemented in an already complex environment such as a hospital with a fairly large number of specialised departments and services. So however it is not expected to be an immediate priority, the metadata still caters for the issues surrounding the management of large volumes of information.

Source materials information was added to the metadata model in order for the lineage of the metadata-elements to be traced back to source materials originating from the rural HBHC initiatives. By adding source material information a user can trace the identified metadata-elements back to their sources and ensure that the source materials were correctly interpreted and that the derived metadata does correctly represent actual data within the HBHC initiative. By adding source material information it offers the ability of a human user to track more easily and make changes to the repository content. If these sources were to change and thus the corresponding data definition as well the source material information would ease the process of identifying the previously derived metadata-elements making the necessary changes to allow them to reflect the changes in the source materials.

The overall solution artefact (the repository) was designed and developed with the goal to ease the creation and management of the metadata items. One way to do so is to conduct a number of usability studies and ensure that the UI is optimal for the intended use and users, although this falls outside the scope of the research as the research does not focus on UI design.

Beyond tailoring the UI for a specific group of users the literature analysis provided another means of managing large volumes of data, detailed in Chapter 2 under the heading of Semantic Repositories. By visualising the repository's content (the metadata and the data-element relations) it allows a user to visually see and interpret the repository content and eases in understanding the potentially complex information and in part easing the content management process. The repository implements a fairly simplistic content visualisation in order to aid in the management of the repository content.

It is fairly difficult to evaluate the long term manageability of the metadata on its own, since the primary means by which the metadata is created and managed is via the repository solution. The metadata however does take management in regard including some additional information that could aid in the management, specifically the source material information. The repository itself also caters for the management of the metadata by providing visualisation components.

The following section will look at the metadata as a quality object.

8.1.1.4 Metadata as a Quality Object

According to NISO because metadata are objects themselves they need to have the qualities that define good objects such as: authority, authenticity, achievability, persistence and unique identification (NISO Press, 2004).

The merit in using this quality object criteria for metadata and its value in relation to this research is debatable since NISO was referring to metadata in the context of digital libraries, where digital library material is stored and the metadata is secondary enabling functionality related to the core content such as retrieval, categorisation and management. Within the developed repository-artefact however the metadata is the core content. Because the metadata is intended to be stored within a digital repository the criteria of a quality digital object was deemed to be relevant however.

Because the purposed metadata model is built on top of a relational-database many of the requirements for quality objects is provided by the DBMS itself. The chosen relational database, SQL Server 2008 Express and most relational-database technologies support ACID (atomicity, consistency, isolation and durability). The DBMS ensures that the metadata items are being stored (or persisted) conforming to one of the qualities of a good object. The DBMS also provides unique identification for the metadata items using Primary Keys common to relational database models, because of the DBMS the metadata items are being archived and the authority and authenticity is also being catered for.

Although authority is not such an important requirement of the metadata model itself but more of the repository in which it occurs and since the repository is still in its prototyped phase. Thus no in-depth consideration was given and no implementations were made for different potential types of possible users and the accessibility to specific part of the metadata model by these different types of users. Since no ethical information was being stored, in the prototyped repository at least, limiting and selective access to repository content was not deemed to be of immediate importance.

8.1.2 Metadata Repository

The previous section focused on the evaluation of the metadata, focusing on a fairly micro-level, this section focuses on the evaluation of the overall repository solution created as part of the research taking into account the various subcomponents. The evaluating of the repository is done according to the three axes specified by Thibodeau's (2006) for digital repositories specifically: orientation, coverage, collaboration.

As has previously been discussed within Chapter 3 under the heading of Data Analysis in Section 3.11 and touched on in Chapter 2 under the heading of Repositories in Section 2.4, the criteria purposed by Thibodeau (2006) are only broad criteria for the evaluation of digital repositories, as more specific criteria are unique to the type of repository, the type of repository and the repository content and the intended use and users. The three axes of Thibodeau (2006) none-the-less provide good headings for a qualitative evaluation.

The repository itself represents the combination of the various designs detailed in Chapter 6 and is thus also the instantiation of the conceptualisations created in Chapter 5. The repository acts as a proof of concept and is, according to the literature on the subject of DSR, a form of instantiation-artefact, one of the types of artefacts produced by DSR.

The final instantiation artefact does serve as proof of concept, but there are limits to what can be extrapolated or derived from the success or failure of the instantiation-artefact, for example: if the instantiation is proven to be adequate or is proven to be a success it is difficult to state that its sub-components are successful. All one can state however is that the subcomponents collectively are correct, that their synergy produces the desired outcomes or phenomena. The inverse of course is also true; the failure of the instantiation-artefact does not mean that all the subcomponents are faulty either. It is thus necessary to evaluate the artefacts created within DSR on multiple levels to gain a holistic view.

The evaluation of the repository in conjunction with the previous evaluation of the metadata, and the metadata model created in solution are both important in gaining insights into the research area.

The following section evaluates at the orientation of the repository.

8.1.2.1 Orientation

This section evaluates the repository according to its orientation, which was described by Thibodeau (2006) as falling somewhere in or between the categories of retrospective (focused on storage) or prospective (focused on providing services to the users).

The repository was intended to be prospective, to be used regularly and meet user needs such as access to the repository content and management of repository content. Long term storage of static content or information which is rarely updated or accessed is thus not a goal of the repository and did not form one of the criteria by which the repository was evaluated. However the repository is intended to store content for a long period of time with regular user access but that the repository content is not intended to be static.

Table 8.3: *Repository Orientation Evaluation.*

Orientation Criteria	Positive	Negative
Access to Repository Content	Repository provides CRUD UI, based around lists and textboxes to display data. Repository also provides visualisation UI that represents basic content information	The CRUD UI is primarily focused on management and creation of repository content The visualisation UI only shows basic information and is more a proof of concept showing the value of visualisation.
Ability to Manage Repository Content	A number of CRUD interfaces are provides with all the necessary, validation and checks to manage the repository content.	The visualisation interface has no CRUD functionality and limited connections to the CRUD UIs.
Regular access to repository	Content and the UI is stored and structured in such a way that regular user access is catered for.	Repository because of its prototype nature is focused on desktop, single user access.
User ability to comprehend Repository Content	Visualisation UI specifically designed to aid in the representation of repository content in a meaningful way. CRUD interfaces offers some, if limited, ability to comprehend the repository content.	Visualisation is a prototype and shows limited/basic information. CRUD is specialised for management and not for content presentation and comprehension.

The repository would be fairly useless if there was no mechanism for inputting content. The repository offers a number of UIs that allow for CRUD functionality (detailed in Chapter 7 such as the Initiative Management, Relationship Management etc.) allowing a user to both create and update content and view the current content of the repository. The CRUD UIs however were ill suited to fulfil one of the requirements of the repository namely, to allow users to gain a better understanding of the different data-elements used by a specific HBHC initiative. The CRUD UIs did contain and represent all the necessary information and content. The CRUD UI however was limited in its ability to successfully represent the content, because the UIs were based around a multitude of lists and simple text fields, and it became fairly apparent that users would not easily be able to gleam any useful insight from looking at raw data.

Because of the limited ability of the CRUD UIs to informatively present the repository content to human users a more visual interface was added to visualise these relationships and data-elements. Visually representing the repository's content made understanding the data-elements and the relationships amongst the data-elements easier. The visual interface thus aids in the intended orientation-goal of the repository, specifically allowing users to comprehend the repository content.

However some issues did arise when visually representing the repository content, firstly it was difficult to effectively display every last piece of information captured in the CRUD interfaces, thus for the initial prototype only the minimal amount of information was shown focusing on the bare minimum. Attempts to show any significant amount of information left the UI cluttered with information, leading to a state of information overload where a user's could not pick up on the important information amongst the large amount of information being displayed. The visual component much like the rest of the repository was only a prototype, a proof of concept showing the value in visually representing the repository content.

Furthermore providing CRUD functionality on the visualisation was problematic, and was thus left out of the initial visualisation UI. Currently the CRUD (management) functionality for the different components is housed on several different UI forms, attempted to fit all of these controls and UI elements onto a single form, prove too complex for the initial iterations of the prototype and it was thus decided to keep the initial CRUD forms and use the visualisation UI for display purposes. The visual UI does however link back to the CRUD UI forms, meaning that the two are not totally disconnected.

The following section will look at the coverage of the repository.

8.1.2.2 Coverage

This section discusses evaluation of the repository in relation to its coverage. Thibodeau (2006) defined coverage as: "how well it covers the universe of assets it should or might hold". The coverage thus relates to the content of the repositories, its ability to acquire, store and use the content.

Table 8.4: *Evaluation of Repository Coverage.*

Coverage Characteristics	Positive	Negative
Content Acquisition (Human interpreter who analyse source materials to generate content).	Human user (interpreter) can analyse potentially obscure information and can communicate directly with knowledgeable member from the HBHC initiative. Metadata model takes into account management and error correction (using source material information).	Human user (interpreter) can potentially make errors under the best of circumstances. It can be time-consuming to analyse, interpret, clarify and input the derived metadata content into the repository.
Content Storage (Metadata model).	Refer to Section 8.1 where the metadata and metadata model is evaluated and discussed.	
Content Usage	The scope of functionality was purposefully limited in the early conceptualisation and design phases, in order to ensure that all important usages are included. Necessary CRUD functionality is included to help create and manage the repository content. Visualisation of repository content to aid in simplifying comprehension is included.	Interfaces with other systems is not included, but for obvious reasons (few or no such systems present in rural HBHC). Visualisation component is only a proof of concept phase and does not visualise all repository content.
Content Type	Successfully stores data-elements and the relationships between data-elements	Procedural and best practices cannot easily be quantified and stored

The repository was intended to store semantic structural metadata, although some thought was also given to storing the underlying 'core' data as well, for example: the metadata defines that there is a patient and a patient has a name, but the data would be the patient's name. The metadata model created in this research does have the ability to store the underlying data and relate it to a given metadata item but for the initial prototypes created in the research focused primarily on the metadata itself, in part because of the issues involved in retrieving the data and the ethic consideration that would needed if the actual data was to be stored.

The semantics component is however somewhat limited in the initial prototyped artefacts, offering only limited synonym management and some contextual information to help clarify the synonyms. More work can however be done on the semantics possible including the semiotics, but this was seen as being a fairly large and complex endeavour and beyond the scope of the research.

The repository content was intended to be entered by a human user, who would serve to analyse and interpret the source materials, most likely working closely with knowledgeable members from the HBHC initiative from which the source materials originate. As discussed in Chapter 2 under the heading of Metadata Creation there is two ways of automatically creating metadata, harvesting and extraction. Since no current explicit metadata exists in the HBHC care forms (which were used to create the repository) or within the ethnographic data collected as part of the research (which also aided in creating the repository) harvesting was not an option. Furthermore somehow finding a way to automatically extract the metadata from the many varied types of source materials (spanning varied paper forms to audio interviews) was not feasible given the complexity and time required to find such a solution.

The management of the repository content would also be a human centred activity. A human user would make the necessary changes in order to reflect changes to the data and the data usage that had occurred within the specific HBHC initiative.

As has been detailed in Chapter 6 under the heading of Database, Entity Relational Diagrams the content of the repository is stored within the created metadata model. The metadata model is built on top of a SQL-based relational DBMS and using an active record pattern for persisting content to the database means that the solution can work with any of a number of SQL-based relational DBMSs.

The initial concept of the repository was to provide it with interfaces to external systems and allow these systems to both pull and push information into and out of the repository. Ideally if the repository could be linked to working systems then the underlying data component could regularly be updated and possibly the metadata components could also be automatically updated. However the lack of compatible IT-based systems within rural HBHC and the scope of the endeavour to provide this type of interface meant that it was not included the prototype.

The primary goal of the repository was thus to allow human users to input content, manage the content and present the content through visualisation in a meaningful way that would allow the users to gain a better comprehension of the repository content.

The following section discusses the evaluation of the repository according to its collaboration.

8.1.2.3 Collaboration

The collaboration of a repository according to Thibodeau (2006) falls somewhere on or between isolated and collaborative and is defined as: "whether it can achieve its purpose operating in isolation or whether it collaborates with other organizations in order to achieve success".

The initial intend of the repository was to have a limited amount of collaboration offering interfaces to other systems that would allow these external systems to pull content and push content into and from the repository, ideally minimising the need for a human user to actively maintain the repository content. Because of the lack of IT-based solutions within the rural HBHC domain the initial prototype repository does not include interfaces in external systems or any data export/import functionality.

Within the initial iterations of prototype development it was decided not to focus on the collaborative functionality, both in terms of multi-user and in terms of external interfaces to other systems. The focus was rather placed on the concepts and components that were in question, such as the metadata model being used and the visualisation, in order to allow more effort to be placed in refining these components which were not yet proven to work effectively then focus effort on developing functionality which did not serve as proof of concept.

Focus was not placed on the collaborative functionality because it would be more of a technical challenge than a research challenge. The initial prototype was a desktop-based single user application. Changing the application from desktop-based to web-based and making the necessary changes to how the database functions, how data is cached and locked would have enabled multiple users. This would however have added unnecessary complexity to a proof of concept application with little research value.

Because the repository system is intended to work in isolation there are limited evaluation metrics that can be applied to the developed repository instantiation artefact in terms of collaboration. If however the repository was intended to be collaborative then several more metrics that could be applied to evaluating the repository such as the degree of collaboration, the importance of collaboration etc.

The most relevant evaluation metrics of course would be if isolation is or is not the most effective approach. However since the repository was designed to work in isolation and to work in an environment with limited active IT-based solution, the answer would be yes, isolation is an effective approach given the environment, goal and intended uses of the repository.

8.1.3 Use of the Repository

Within the previous sections the repository developed as part of the research was evaluated based-on criteria found within the academic literature on the topic of digital repositories, specifically the criteria specified by Thibodeau (2006).This criteria provided broad categories by which the repository can be evaluated. However to gather even more detailed evaluation results not only must the repository itself be evaluated but also the actual use of the repository and the results of this usage (the "product") needs to be evaluated.

This section specifically focuses on the evaluation of the usage of the repository by looking at the result of using the repository; specifically this section will look at how the source materials are captured and how the resulting information is stored within the repository. This section will also focus on how the data is represented in the Visualisation UI.

The underlying reason for evaluating the "product" of the repository is in measuring how effectively the repository is able to capture and represent a "reality". The repository attempts to represent a "reality" of how data is defined and used by a given HBHC initiative. The repository does not seek to capture "the truth" or "the reality" but a truth that is true for a given HBHC initiative. Thus it is important to gage how accurately the repository can capture the data-elements, their relations and context for a given initiative as this directly relates back to the usefulness of the repository.

The following section will focus on the example document that will be used to test the usage of the repository.

8.1.3.1 Example Document

This section will focus briefly on the example document that will be used to show the usage of the repository, ideally by providing a quick overview of the document it will better place the document's intent and purpose thus helping to clarify the resulting "product".

OLIVE LEAF FOUNDATION

CONFIDENTIAL

Adult Client Form

Reference number (province/site/number)	

1. Client Consent

I ______________________ agree to the recording of my details.

Signature ______________________ Date ______________________

2. Support Group member/HBC client Member Particulars

Surname				
First Name				
Known as				
Date of birth (dd/mm/yy)				
Age (in completed years)				
Place of birth				
Sex				
Nationality				
ID Number				
Home language				
Religion				
Employment status	Employed	Un-employed	Self-employed	Other
Marital Status	Married (customary)	Single	Divorced	Separated
Residential address				
Contact numbers	Tel (h)			
	Tel (w)			
	Cellphone			
Contact person in case of emergency (different from own)	Name:			
	Telephone:			

Figure 8.2, *Paper Form from the Olive Leaf Foundation, Adult patient details.*

Figure 8.2 shows the adult patient details form used by the OLF in Motherwell, Eastern Cape of South Africa. Motherwell and the OLF are discussed in more detail in Chapter 4 as it formed part of the initial ethnographic study. The form is intended to capture the personal details (called particulars) of a HBHC client (or patient as it is referred to on other OLF paper forms) or of a member of the support group.

This particular paper form was chosen because of its fairly limited number and variety of fields, with only mild interconnectivity between the data-elements. The form also does not make reference to complex medical data or data located on other forms, thus all the necessary data is located on one form in one place. The primary context of the form is also very apparent as the details on the form all relate to an adult patient.

The following section will look at the identification of the various data-elements.

8.1.3.2 Identification of data-elements

This section will focus on the process of identifying the data-elements on the example document from the OLF that will be used to show the usage of the repository.

The repository is a tool intended to be used and be useful for a human interpreter. The role of the human interpreter is someone with the needed insight and experience in the usage of the repository and with the ability to work with knowledgeable individuals from a specific HBHC initiative and analyse or interpret various source materials (paper document, interviews, images etc.) that originate from the HBHC initiative. The goal of the interpreters actions is to gain a understanding of the data-elements, their relationships, context etc. as seen by the specific HBHC initiative and then to use the repository to capture this information with the intent of making this information available to others who can benefit or use this information to help or support the HBHC initiative.

The identification of the data-elements (the content that will be inputted into the repository) is a human process. Automatic metadata harvesting or extraction was ruled out as a possibility since no effective solutions currently exist (or more accurately a detailed literature review did not reveal any such solution). Thus an automated solution would have needed to be created from 'scratch' and such an undertaking would have taken a completely different route to the intended research with limited value or insight being added to the research.

Looking at the example care document in the previous section the data-elements are fairly simple to identify. Each field is a data-element (the name, the surname, the nationality etc.) while the corresponding values (which are not included for obvious ethical reasons) would constitute the data. Collectively this data relates to an adult patient or an adult support group member, presumably of course by explicitly identifying the patient as an adult there would be other types of patients as well but only working with this single form means that these other types are unknown, but could easily still be managed.

It is possible by looking at the example form to capture the data in simple entity/attribute style, where an entity (such as an adult patient) is constructed by its various attributes (name, surname, nationality etc.). There are potentially a number of other entities as well such as the contact person, which could potentially have just as many attributes as the adult patient, but according to how the OLF views a contact person, a contact person only has the bare minimum number of fields.

Certain entity relations are more difficult to define, for example telephone and cellular phone numbers. A patient is seen as having three contact numbers, a work number, a home number and a cell phone number, a patient's contact person also has a telephone number. Conceivable cellular phone and telephone numbers would be very much a similar in their format; within the South African context both have a similar ten digit format (999 999 9999), but they are treated as related but different. Thus there is a single telephone data-element with two specialisations, work telephone and home telephone, and there is a single cellular phone element, all collectively forming a patient's contact details.

From the form alone it is also difficult to know which values are important or could be left out; the form does not explicitly make these differentiations. Looking at rural HBHC it is hard to imagine that all patients would have telephones or cellular phones. In the case of OLF and Motherwell there is barely any cellular phone coverage in the area at all thus it is hard to imagine that all patients would have working cellular phones. One can however assume which information is and is not important or needed; names, national identification numbers obviously seem to be important. However assumptions about what is and is not important might be incorrect and negatively affect the quality of the repository content, thus once again putting emphasis on the human interpreter's role in working with the HBHC and solving these types of inconsistences.

Most of the attributes values are left empty for fairly obvious reasons it is difficult to imagine all the possible names and surnames that potential patient could have. Other fields such as employment status and marital status have a finite number of potential values. It is obviously important that these empty and finite values are included and reflected in the repository itself.

Another important consideration is which data needs to be captured in the repository, for example on the example form a patients' needs to provide his consent to have his details captured and to sign the paper form. The need for the patient to give his consent is an ethical, legal and procedural requirement and the signature is simply proof of consent, but overall it has little practical use in terms of care delivery. As has been shown in the previous evaluations, a limit of the repository is that it is not specialised to capture procedural information, although consent can be included with limited hassles.

The following section will look at the creation of the repository content.

8.1.3.3 Repository Content Creation

The identification of the data-elements on the example form was discussed in the previous section; this section will look at the information generated by using the example form adult client form.

Because only a single document was used there were limits to what could be derived for example: the example form only had a single context that is the data capture for an adult patient and the relationships between seemingly similar data-elements could not be gleamed.

Because only one context and only a single form was used, the semantics and the various synonyms could not fully be identified and captured. Overall this does not heavily or negatively impact on the overall quality of the derived metadata in this particular case because the goal of capturing semantic information was to have it aid in clarifying any semantic inconsistencies. The semantic information thus doesn't form the core data but merely aids in ensure the quality and comprehensiveness of the core data.

Table 8.5: *Example Source Material.*

Name	Description	Confidential	Document Type
Adult Client Form for Support Group/HBHC Client Members	A care document used to capture the particulars (details) of an adult client (patient) of the Olive Leaf Foundation. The content of the form is confidential and it is required that client provides his consent to have his details recorded.	Yes	Adult Client Form

Table 8.5 shows the Source Material's information that was created in order to show the usage of the repository. The source information in Table 8.5 represents the example care document. Two additional attributes were defined 'Confidentiality' and 'Document Type' and although this information is also captured in the description-field, capturing the information explicitly was considered to be the best solution because these additional attributes were important in describing the document.

Capturing detailed source material information does not directly lead to higher quality metadata information, since the reason for capturing source material information is to provide a lineage which over time can be used to aid in ensuring the quality of the derived metadata items.

Table 8.6: *Example Metadata.*

Metadata-element Name	Metadata Description	Has a Value	Is Required	Value Type
Place of Birth	Location where the person was born.	TRUE	FALSE	Text
Religion	Religious information	TRUE	FALSE	Text
Name	Name by which a person is known.	TRUE	TRUE	Text
ID Number	Identification (ID) number of a person.	TRUE	TRUE	Text
Home Language	Home language spoken by a person	FALSE	TRUE	Text
Age	Age in completed years	TRUE	TRUE	Numeric
Consent	Whether the relevant parties provided consent for the information to be captured and stored.	TRUE	TRUE	Entity
Cellphone	Cellphone number by which a person can be contacted	TRUE	FALSE	9999999999
Residential Address	Address where the person lives, the person's residence.	TRUE	TRUE	Text
Employment Status	Current state of employment.	TRUE	FALSE	Employed / Un-employed /Self-employed / other
Surname	A person's surname	TRUE	TRUE	Text
Contact Numbers	Telephone and cellphone numbers by which the person can be contacted.	FALSE	TRUE	Entity
Signature	A person's signature	TRUE	TRUE	Image
Contact Person	Contact Person in case of emergency (different from own).	FALSE	TRUE	Entity
Client	A generic type of client, created to manage the different possible types of patients that could exist	FALSE	FALSE	Entity
Nationality	Nationality of the person.	TRUE	TRUE	Text
Telephone Number	A telephone number	TRUE	TRUE	(999) 999 9999
First Name	A person's first name	TRUE	TRUE	Text
Adult Client	An adult client of the Olive Leaf Foundation, receives any one or combination of services from the Love Leaf Foundation	FALSE	FALSE	Entity
Date of Birth	Date when the person was born.	TRUE	TRUE	dd/mm/yyyy
Marital Status	Marital status of the person.	TRUE	TRUE	Married (customary) / Single/ Divorced/ Separated

Table 8.6: Example Metadata (continued).

Metadata-element Name	Metadata Description	Has a Value	Is Required	Value Type
Telephone Work	Telephone number by which the person can be contacted with when at work.	TRUE	FALSE	(999) 999 9999
Date	A date in terms of day/month/year	TRUE	TRUE	dd/mm/yyyy
Known As	Name by which a person is colloquially known, other than his first name or surname.	TRUE	TRUE	Text
Telephone Home	Telephone number which the person can be contacted from when at home.	TRUE	FALSE	(999) 999 9999
Sex	Sex (male or female) of the person	TRUE	TRUE	Male/ Female

Table 8.6 shows the derived metadata-elements, based-on the Adult Client example document.

As discussed in the previously in Section 8.1.3.2 under the heading 'Identification of the Data-elements' all of the elements corresponded to fields on the table. Certain fields were grouped together and obviously they were similarly grouped together when being entered into the repository, for example home telephone, work telephone number and cellular phone number collectively form the Contact Numbers.

A fair number of elements were derived and by simply looking at the list it is difficult to be able to gain any meaningful understanding of how these data-elements relate to each other or how they function. This yet again emphasises the need for data visualisation when it comes to the content of the repository.

Figure 8.3 shows the visualisation of the data-elements and relationships generated from the metadata derived from the Adult Client example document. As has previously been discussed as part of the evaluation of the repository detailed in Section 8.1.2 and its sub-sections, the visualisation component does not visualise all the captured information (being that like most of the repository the visualisation is simply a proof of concept). However the visualisation does help to show more clearly how the various elements relate to each other (although most of the relationships could be further refined using additional source materials).

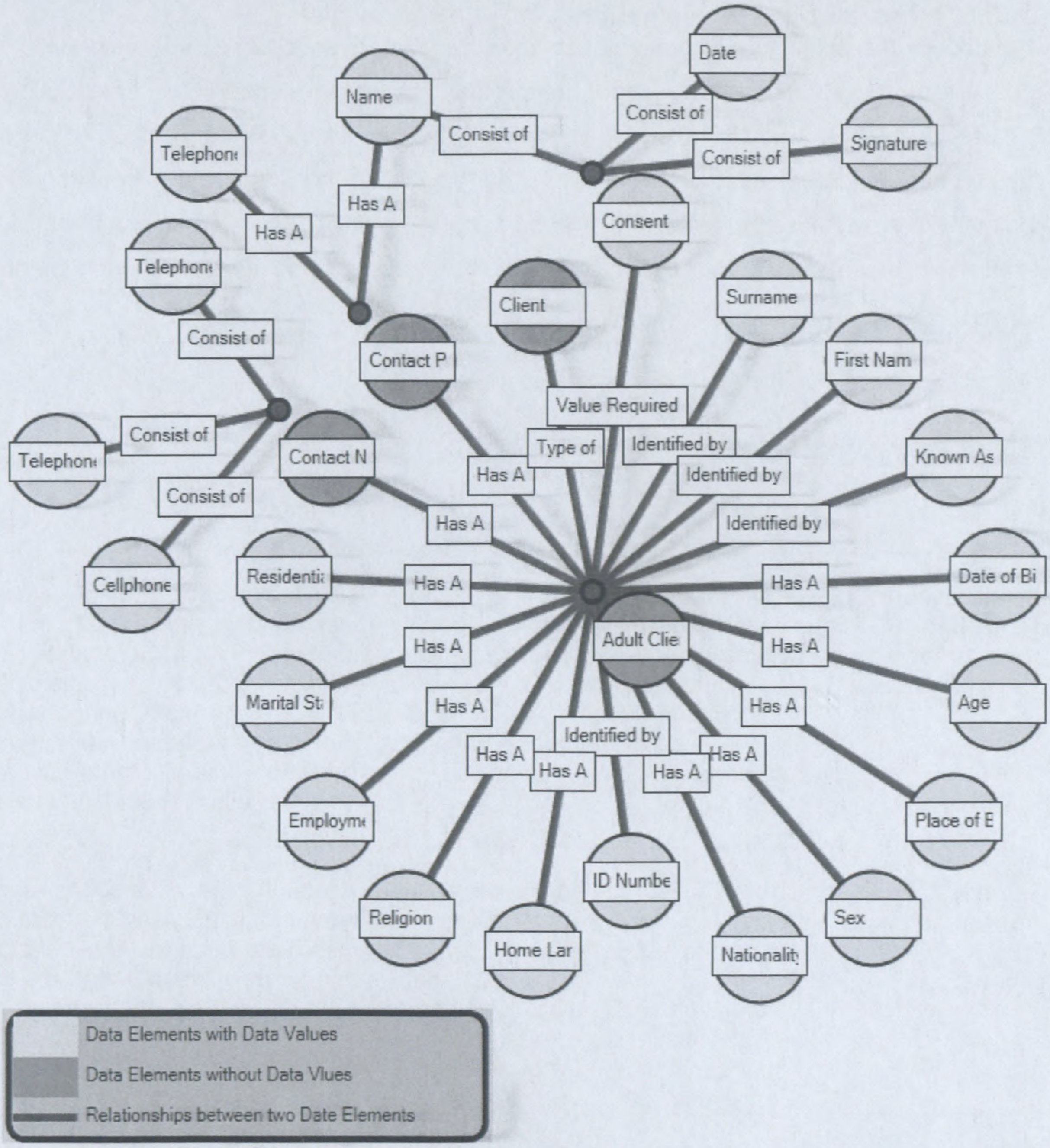


Figure 8.3: *Generated Visualisation of the example form.*

The visualisation does not simply show hierarchies of relations but shows a network of relations (where an entity can have many 'parent' and 'child' entities). The relational network clearly shows where data-elements are being reused, or where they could have been reused. An example which could probably have been cleared using additional source materials is that of name. Although not superficially obvious there are several different types of names, there is the 'first name' and 'surname' combination and then there is 'name' alone which would most likely imply both first and last name, there is also the 'known as' which is a nickname or colloquial by which a person can be known. These differences might not be of immediate importance but it does show that what constitutes a name or some form of standardisation of the name data-element was not considered.

8.1.3.4 Repository usage evaluation

The preceding sections looked at the usage of the developed repository in order to provide a context for the evaluation of the usage, the 'product', of the developed repository solution. The previous section was also intended to contextualise the evaluation the previous sections gave an overview of the example document used to evaluate the usage of the repository. In addition the previous section looked at the process of identifying the data-elements on the example document and finally the repository content produced using the example document.

This section primarily focused on the content produced from the example document but does take into consideration the contextual information detailed in the other sections.

Table 8.7: *Evaluation of the usage of the repository.*

Usage Characteristics	Positive	Negative
Identification (Process) of the data-elements (repository content)	The data-elements on the example paper form were easy to identify. Some issues might arise when data appears across a number of forms but ideally these issues would aid in the identification of data-element inconsistencies.	The identification of the data-elements is somewhat time-consuming. As more data is added it is occasionally necessary to make changes to previously captured information (as later data-elements reveal new details about previous elements).
Representation of the source material ("real world") information	Different source materials might differ but from example document used the repository is capable of capturing the identified data-elements, and the visualisation is fairly successful (needing mild changes to the layout and visual representation).	Although the repository can capture all the identified data-elements, the visualisation still needs more refinement (most likely involving the intended repository users, which fall outside the scope of research).
Comprehension (identification of data issues)	Although the visualisation component is not yet complete is does a fairly good job of visually representing the repository content, and is aids in the representing some of these issues (primarily duplications, and inconsistent relationships). The CRUD interfaces also offer some comprehension although is more laborious and time-consuming than the visualisation interface.	It is far easier for the user who entered the repository content to also identify the data issues, although the visualisation is an effective tool for comprehension it only eases the process of comprehension, some time and effort is still needed by an external party to interpret and comprehend the repository content.

The above criteria for measuring the usage of the metadata repository were not based-on the academic literature (as was the case with several of the other evaluations conducted within this chapter). The criteria used for evaluation was based-on what was perceived to be important aspects of the repository, namely: the process of identifying the data-element from the source materials and creating the repository content, how successful the repository was at representing the data-elements identified using the source materials and finally the ability of the repository to aid in the comprehension of the data-elements and the identification of the issues related to the data-elements.

Research Contribution

Chapter 9 Research Contributions

9.1 Introduction

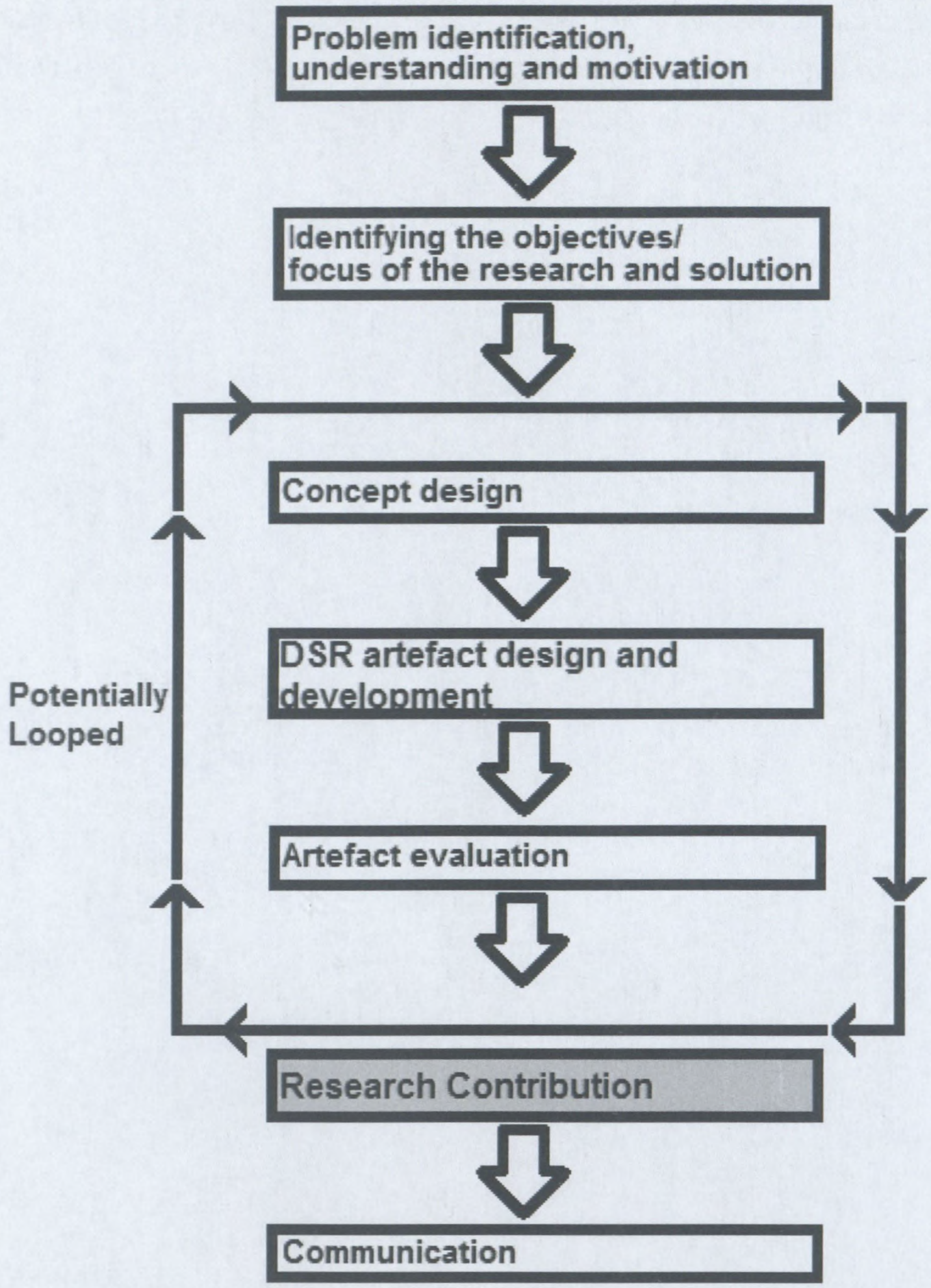


Figure 9.1: *Chapter 9 in relation to the research methodology.*

This chapter's primary focus is on the presentation and discussion of the research process and the findings of the research. It is the sixth and second to last step in the DSR methodology used in this research. The identification of the research contribution occurs after the development efforts have been completed, although preliminary identification of these contributions occurred as they arose in the previous steps. This chapter refines these preliminary contributions and presents them as a collective whole.

This chapter does not simply present the research findings and contributions but further attempts to provide reflection and insight in regard to a number of areas touched by the research, specifically in terms of the literature, the research process and methodology and the research findings.

9.2 Summary of Findings

Some of the research findings were produced by the initial literature review conducted as part of the research process. Additional findings were produced by the ethnographic study and research area contextualisation, the development of the research and finally the design and development process itself. The findings related to both the problem and solution domain.

This section briefly categorised the findings of this research into the categories of: literature analysis, design and research methodology, problem contextualisation and artefact design and development. Within this section the findings will be briefly presented followed by a discussion.

The first subsection looks at the findings produced from the literature, the second subsection looks at the findings from the research methodology, the third at the problem contextualisation, primarily the ethnographic study in Chapter 4 and the final subsection looks at the findings from the design and development process.

9.2.1 Literature Review

This section lists and discusses the findings produced from the literature review detailed in Chapter 2.

While conducting the research the literature proved to be a valuable asset. The literature acted as a starting point for the research allowing for an initial overview of the current academic knowledge available on the various topics involved in the research. But not only did the literature serve as a starting point, it also helped to refine the methodology used during the research and it aided in the conceptualisation of both the problem domain and intended solution artefact developed as part of the DSR process. The literature also aided in defining the criteria for evaluating the various artefacts developed during the research.

This section will list and discuss the various topics found within the literature in the same order that it was listed in Chapter 2.

Many of the articles sourced on home-based care, especially within the rural and developing context, clearly speak toward the HBHC context, although very few of the articles explicitly call it HBHC, or any similar name. Only a handful of the articles provide a detailed description of the HBHC context.

A significant amount of literature does exist on the topic of rural and developing countries especially in regard to healthcare. These mostly tend to be either narrow in scope, focusing on a given solution or a given situation or broad in scope focusing more on predominate characteristics and themes.

The literature thus rarely spoke directly to the HBHC context, but it was fairly well covered indirectly. It was thus necessary during the literature review to refine the more general topic of rural healthcare in developing countries to ensure its relevance to the topic of HBHC.

Very little literature sourced on HI speaks directly towards the HBHC context. The literature on HI did however provide a thorough overview of the characteristics of HI including advantages and disadvantages.

The advantages of HI, listed in Table 2.1 in Section 2.2.1, provided a good motivation and justification for why research into the topic of HI is important and the disadvantages. Furthermore three broad categories of contribution where identified namely: advantages for the patients; advantages for the facility and advantages for the healthcare workers. Many of these advantages fell into one or more of these categories.

The disadvantages of HI, listed in Table 2.2 in Section 2.2.2, provided a number of areas for future refinement and research. These disadvantages also further provided the key link to the topic of interoperability. For the disadvantages broad categories were also defined namely: human issues, technical issues, organisational issues and cost/effectiveness issues.

The literature provided a number of examples of the usage of HI and the usage of HI in rural and under-resourced communities. Most likely this is because of the limited explicit reference to HBHC in the literature there also exists limited explicit mention of HI in terms of HBHC. To overcome this limitation Chapter 2 contains a fairly detailed discussion of HBHC and then of HI. By discussing these two topics in depth it was possible to link the two.

The topic of integration in HI was fairly diverse, although it was clear that it was still a present concern with decades of work having gone into an attempt to overcome the issues preventing integration.

This showed that there existed a wealth of knowledge to draw from when attempting to discuss or address interoperability. This also provided a fair warning that the topic of integration especially in HI is an extremely complex one that would potentially take a significant amount of effort to fully address. The topic of integration did however provide insight into a number of relating topics such as semantics, ontologies and of interoperability and integration standards.

The literature shows that there clearly is no single definition or usage of metadata. This was in part because of the broad definition given to metadata. For this reason the research had very little pre-existing knowledge to draw on in term of structural metadata that spoke directly to the research area. This however offered a clear area where the research can make a contribution, further clarifying the topic of structural metadata and providing a practical example of structural metadata in use.

Another topic which was poorly covered in the literature is the explicit mention of metadata within healthcare, especially rare was metadata within HBHC. In most of these cases metadata was mentioned briefly in the context of large clinical research systems, where metadata can help with the processing of the large volumes of data within these systems. No examples were found of metadata being used to define the data-elements in healthcare offering another possible area of research.

Most of the metadata literature which was sourced related to digital libraries where the Dublin Core was a well-vested and fairly saturated standard. However some of the authors were attempting to show the usefulness of standards such as METS within the digital library domain. Predominately the use of metadata within the digital library context was meant for the processing, cataloguing and retrieval of digital library content.

Because there was little literature on the topic of structural metadata that was useful to the research there were also few standards found which could be utilised during the research. The literature review still considered the topic of standards to show that the necessary diligence and effort was put into the research.

It quickly became apparent from the literature that metadata was a complex and multi-faceted topic where specific topics were extremely well covered and other far more vague. It did become clear that metadata existed in almost all systems with the only real difference whether it was explicit, identified as being metadata, or implicit, known by another name or not identified separately. A good example of this was with the conventional relational database model where structural metadata does exist, it defines how data is grouped (tables, rows, columns etc.) and how data relates (relations). With some knowledge of metadata it is possible to see the dichotomy of data and structural metadata within the relational database model but rarely is this explicitly pointed out.

The literature on Metadata Quality and Metadata Creation provided useful insights which aided in the research. However the sourced literature on the topic of quality criteria for metadata was limited.

Because of the broad definition and corresponding multitude of understandings of what metadata is also held a potential issue for the research, the research might be dismissed because it does not conform to one of the perceived definitions. By detailing the multitude of different types of metadata, hopefully it was successfully conveyed that there is no single definition or usage of metadata and that at best only broad categories can be defined.

The literature on repositories and semantic repositories provided an overview on the topic but not a deeper understanding or any clear guidelines to the construction of a repository or semantic repository. The literature did provide criteria for evaluating a repository which proved useful in the evaluation phase of the research.

The topic of semantics provided a good link between HI in terms of interoperability and the topic of ontologies. The literature showed the clear importance of taking semantics into account when attempting to enable interoperability and when creating and representing ontologies.

The literature on DSR did however make it clear that it was not a dominate form of research within the IS field and in part is still being debated and questioned. For this reason several methodologies as well as epistemological and ontological stances were discussed within the academic literature.

It also became clear that the literature on practical implementations of DSR is rare, especially in the healthcare context. Most of the articles are only overviews or discussions of DSR and not practical applications thereof; seemingly most of the authors are still fighting to get DSR accepted as a credible form of research within IT academia. Unfortunately although there is a fairly acceptable pool of theoretical knowledge to draw on when it comes to DSR, most of this has little practical applications, meaning that the research still needs to ensure that the chosen methodology meets the requirements of the research. Furthermore the literature on DSR does provide a number of guidelines and frameworks for conducting DSR but details on aspects of the research such as analysis are left somewhat wanting. Although this is in part because of the limited practical application found in the literature and because of varied differences between development projects.

The next section discusses the findings from the creation of the research methodology.

9.2.2 Design of the Research Methodology

This section reflects on the DSR methodology, the process of which is detailed in Chapter 3 under the heading Research methodology.

An amalgamated DSR methodology was used; discussion of several available DSR methodologies found in the literature and the process of combining these methodologies is detailed in Chapter 3 in Section 3.7.

Because of numerous similarities between the frameworks, guidelines and methodologies for DSR purposed in the literature creating a combination was simple enough, there are similarities between the frameworks and also notable differences. The various related steps are grouped together and presented in Chapter 3 with Table 3.3. The idea behind combining the different guidelines and frameworks was to ensure the quality of the research, to help ensure that all the necessary components have been included and that no misinterpretation of a given step or phase within the research exists.

Findings in terms of the analysis of available DSR guidelines and methodologies include:

- That these methodologies usually have one of two strong leanings.
  - Either they bare a strong resemblance to generic conventional research.
  - A strong resemblance to some particular software design and development methodology.
- Several strong themes exist in the available DSR methodologies including:
  - Addressing relevant research problems.
  - Design and development phases.
  - Importance of evaluation.
- Very few methodologies call for DSR to have a strong academic literature basis or for the usage of the academic literature during the research.
- Detailed discussion of rigour and relevance in terms of the research is far less prevalent in newer purposed methodologies.

- Design activities are usually combined with Development in a single phase and furthermore design activities are rarely elaborated in much detail.
- The DSR methodologies rarely make reference to a specific development method (waterfall, spiral, agile etc.) and also rarely make reference to any particular design method (co-operative design, bio-mimicry etc.).
    - The research applicability of a given development and design method is thus also omitted.

The literature provided a number of different frameworks and guidelines for conducting DSR. Most of these frameworks and guidelines seemed to have strong similarities to some of the more conventional software development life cycles (SDLC) although rarely did the authors point out these commonalities or for that matter refer to any specific software development methodology while discussing their particular framework or guideline for conducting DSR.

This lack of reference to the SDLC or any particular software development methodology seemed fairly odd since DSR in most cases involves both the design and development of components/artefact and the various available software development methodologies influence the design and development process differently. For example with the classic waterfall development methodology all design is completed up front and is rarely the design changed when development or implementation process has started. While in the agile methodology the design can be refined after the development process has started.

A lack of discussion as to the research validity of a given development and design approach is thus lacking but would otherwise have been welcomed, especially within this research as it would have provided a strong argument and justification for using a particular design and development method. This research however having used only a single design and development methodology can only mildly contribute in this regard, since no comparison can be made.

The development phase of this research was conducted using agile software developments approach. An agile approach was preferred by the researcher and further deemed to be relevant to the research being conducted. The relevance of an agile approach is because it assumes that all requirements cannot be known upfront and that as the development process continues new insights will be gained and the design and requirements will most likely have to change in order to meet these requirements. This was important because of the nature of the design-based research being conducted. DSR sees the process of design and development as a knowledge generation processes whereas a development approach such as the waterfall model assumes that all requirements can be elicited upfront and that the problem being addressed can be fully explored and understood before design and development. A development approach such as the waterfall model stands in direct contrast to the underlying idea of DSR since if all these insights can be known before development then there is no need for development to provide these insights.

Evaluation is an important part within a DSR methodology. The evaluation provides the validity of the artefacts produced during the development phase of the DSR.

In terms of evaluation the followings was discovered:

- That evaluation of the produced artefacts needs to use criteria relevant to the artefacts.
- The production of universal criteria for artefact evaluation is not feasible because of the variety of artefacts that can be produced.
- To ensure validity of the evaluation valid criteria needs to be selected. The most logical choice is to ground these criteria in the academic literature or in terms of the artefact goals.
- Evaluation in DSR is not limited to completed artefacts, proof of concept artefacts and the process of artefact creation can both hold research value.

Most of the frameworks however had very little to say about the evaluation of the actual artefacts being developed. This was somewhat concerning since the evaluation plays a large role in ensuring the validity of the designed research artefacts and, as a direct result, the validity of the underlying knowledge generated via the research artefacts. The reason for this limited availability of universal evaluation criteria is easy enough to understand when looking at the nature of software development undertakings.

There are many variables in play within any given software-development project and thus the criteria for success can be vastly different between the different projects. Furthermore software development can produce so many different types of products, ranging from cellular phone applications and video games to massive cloud based system each notably different in terms of their construction and goals. Attempting to find common, universal and relevant criteria to evaluate all these differing software development products is by no means a small task. For the most part then the most relevant criteria which can be used to evaluate a given software development product or process is one that is deeply linked to the nature of the given product or process being evaluated.

Within DSR this same issue applies when defining criteria for evaluating the DSR process itself, since DSR contains components of design and development and produces numerous and varied artefacts. The notion of creating standard and relevant criteria for evaluating all the different artefacts produced by DSR is not feasible.

An additional issue with DSR is that for the most part it does not cover the entire software development process; DSR can look at only a handful of artefacts, the subset of a system, the process of development or some aspect of the development process. Thus DSR varies in its focus and granularity meaning that the criteria for evaluation can be even more difficult to identify.

For this reason the focus of this research was primarily on the artefacts produced, not on the complete system and the users' interaction with the system. Essentially the focus of the research was around the design and development of the technical components and the underlying concepts of the system. The two primary components were identified to be the metadata schema and the repository itself, and thus evaluation criteria were found for these two components. For the metadata the NISO standard for quality metadata was used, although it was deemed to not fully be relevant to structural metadata and was slightly adapted. For the repository itself the work of Thibodeau's (2006) provided the axes for evaluation.

A qualitative evaluation was used to evaluate the repository. Qualitative data has the advantage of capturing a lot of information, implicitly and explicitly, and this was seen as being advantageous for the presentation of the research since it allowed for a more robust view of the research. This allowed the intended audience to be presented with a fairly clear picture with which to judge on their own, also helping to manage the possible personal prejudices of the author.

From the research several criteria for DSR are purposed including:

- The ability of the literature to contribute in a number of aspects.
    - Contextualising the research problem.
    - Source of possible problem solutions.
    - A source for standards and requirements for the design and development of DSR artefacts.

  - o Source of criteria for the evaluation of DSR artefacts and the DSR artefact creation process.

- The importance of firstly understanding the research area and the research problem to ensure that adequate and relevant DSR artefacts are designed and developed.

- The need to be transparent about the process involved in creating the DSR artefacts, thus aiding in justifying the design choices and ensuring research biases are clearly presented.

- The realisation that DSR and BSR methods do have an overlap allowing for DSR to make use of what is usually considered behavioural research methods.

9.2.3 Problem Contextualisation

From the literature and from personal experience of the researcher, both during the research and before, it was clear that consideration had to be given to the problem context and the solution context in order to help ensure some form of success. Simply creating a solution without grounding the logic and the design of the solution in the reality and truth can only lead to an unrealistic solution, which in turn can only lead to irrelevant and unrealistic results. For this reason actual HBHC initiatives were chosen to help understand the issues at play and aid in the contextualisation of the problem.

From the research it became clear that having an understanding of a problem is the key to solving the problem. This is important even in DSR which based-on the literature is in some cases not associated with a strong literature grounding or the usage of behaviour-science methods. This topic was discussed and justified in Chapter 3 in Section 3.6, but this was further reinforced by the research itself.

For the most part the findings produced by the research are fairly similar to what is found in the literature on HBHC such as:

- HBHC plays an important role in providing healthcare services and community support for rural areas where it is needed most urgently.

- HBHC by its nature attempts to overcome many of the issues faced by formal healthcare such as.

- Informal caregivers provide the bulk of the care services.

- Caregivers are primarily volunteers from the communities they serve.

HBHC suffers from a number of inhibiting factors such as:

- A lack of funding.
- A lack of trained medical professionals such as doctors and nurses.
- Limited and poor infrastructure such as telecommunications, electricity and transport.
- In more rural HBHC organisational and management issues are faced.
- Stigma surrounding certain diseases, and the fear of patients being associated with given disease if they receive home-based care.

The primary insight produced by contextualising the problem area related to the nature of the solution. There is no quick fix solution available and especially without firstly researching the casual factors in play, for example, there was a lack of IT based solution at the two HBHC initiatives the research looked at but simply giving them computers would not solve anything - the issue is not with the lack of computers but with the lack of trained users and in Motherwell there is also a lack of electricity.

Although more advanced solutions and innovation would conceivable hold more benefit this is however not the case, before more advanced solutions can be implemented there is a need to address the underlying issues faced by HBHC.

It is first necessary to address the issues at the ground, to create a stable foundation on which more advanced and complex solutions can be built.

The contextualisation helped to define the intended solution. The issues of data inconstancy appear across HBHC initiatives and along with a lack of computer skills and in some cases a lack of infrastructure to support computer based solution at the facility level of HBHC it was clear that the solution had to exist outside the HBHC initiative. The intent of the solution artefact was thus to serve as a tool which can be used by some form of external body to gather structural metadata, semantics and source materials.

The following section will focus on the design and development of the DSR artefacts.

9.2.4 Artefact Design and Development Process

Design and development are two related but still separate processes. Attempting to develop without a pre-existing design can be difficult and error prone for the design offers a "blue print" which is instantiated during the design process.

On the other hand assuming that all design can be done up front and assuming that the process of development is simply the process of following a set number of steps (the design "blue print") can itself also be error prone. This is likely because it is rare that all possible aspects of the solution can be imagined and planned for before the development start. However upfront design does have the value of provided a solid foundation which can then be further refined during the development process.

The ideal for design and development then would be that a sufficient amount of upfront design is first undertaken, understanding the problems being addressed and possible solutions. After the upfront design has been completed and as the development process starts, mild changes are allowed to be made as further analysis and research continues and new insights are produced. Of course this needs to be managed correctly as making too many changes to the scope and specification leads to scope-creeping, leading to an issue of development effort having to try and meet an ever changing target. The goal then is not to fundamentally change the scope and specification but to refine and flesh-out.

This was the case with the development component of this research. An understanding of the problem being addressed was required to aid in the development of a meaningful and relevant solution and the literature grounding along with the ethnographic study provided this understanding of the problem. With a fairly comprehensive understanding of the problem area, the process of defining the objectives and requirements of the solution could be undertaken, aiding to ensure that the objectives were aimed at address relevant issues. The solution was designed based-on the objectives. When a relevant solution was settled upon (the semantic metadata repository), development was undertaken in which the specifications were further developed and refined using new insight obtained from the design and development as well as from additional ethnographic and literature inquiries.

Figure 9.2 demonstrate the contributions of the literature review and the ethnographic study (the contextualisation). The figure shows how the literature review and ethnographic study contributed to the various phases in of the DSR methodology and attempts to show the importance of understanding the research problem, research context and the research area (the ethnographic data) and the value of conducting DSR on a strong foundation of academic literature / knowledge.

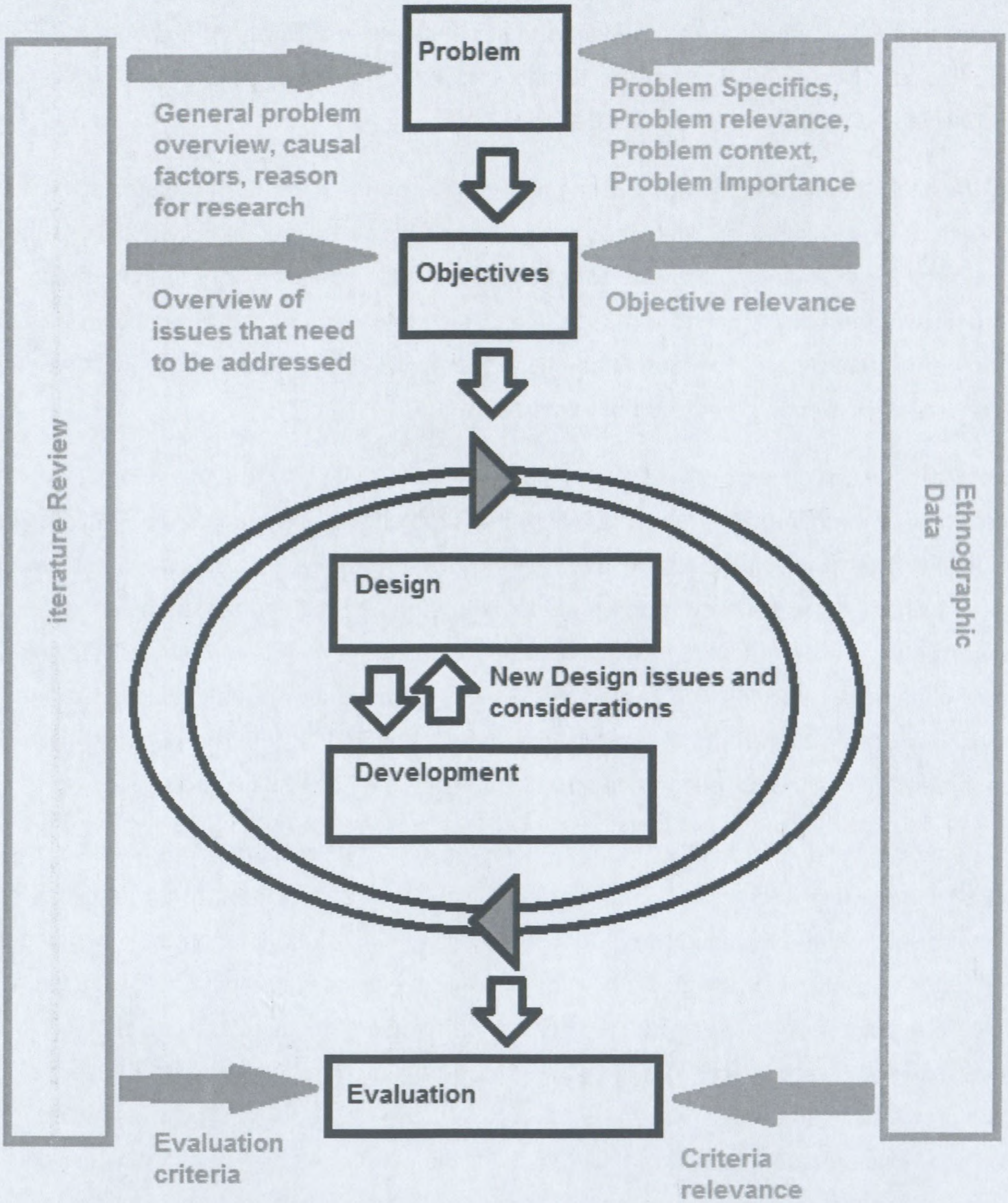


Figure 9.2: *Contributions of the literature and ethnographic data to the design and development.*

Prior to the design and development the solution had a rough outline with several specifications amongst which was that because of the limited available skills and resources at the HBHC initiatives, the solution is intended for a third party to use, ideally some form of governing body. Because the solution is aimed at a third party with the needed skills and infrastructure, it followed that the user will be trained in the usage of the solution. This lessens some of the UI design requirements since standard computer terminology can then be used and the interface development requires less focus on HCI aspects.

The development component of the research followed an agile methodology which was centred on the use of an evolutionary prototyping development approach. Within an evolutionary prototype an initial prototype is created and refine. This prototype is however not a throw-away. As the prototype is refined and further developed the prototype itself becomes the eventual system.

Software development projects can be undertaken in two ways, development of vertical groupings of functionality or development of horizontal groupings of functionality. With the vertical method functionality and components of the solution are grouped and developed together, i.e. in the case of the solution artefact all the needed functionality for capturing the institution information (forms, classes, database-tables etc.) was developed in unison. With a horizontal method the functionality is organised in layers and developed as such, for example, firstly a database design is created, followed by the business logic layer or the various interfaces, then the UI etc. With horizontal development it can be top-down, where high-level components such as the UI is created first and then work continued down to the supporting layers such as business logic or databases. Alternatively with horizontal development it can be bottom-up which is simply the reverse, the lower level components are created first such as the database and work continues creating the various components which rely on these low-level components such as the UI.

Thus within the evolution-prototyping within this research an iteration of development was an additional vertical slice of functionality, or a sub-set of a vertical slice, depending on the size and complexity of the intended functionality which is added. Following an Agile approach each iteration of development was preceded by a research component (when required) and concluded with a small evaluation of the component developed within the iteration. Since the research team was fairly small it resulted in limited coordination and logistical issues.

However because this was a research undertaking there was a limit to how detailed the initial objectives and specification could be. Offering certain ambiguity in the design provided opportunity for experimentation and reflection in the development process, which otherwise would have run the risk of making the development effort "mechanical" and overly rigid. An example was in the definition of the database design, a rough idea was created and implemented, and it was evaluated several times and refined in order to ensure that the optimal design was used. Within each refinement the original ethnographic data and objectives were used to ensure that the refinements were not unnecessary or beyond the scope of the objectives.

Presumably of course commercial software development does not have this luxury offered to academic software development. Commercial software has strict requirements, such as time and cost, and these factors usually remove any notion of experimentation or refinement via iteration. Academic software development especially using a DSR such as experimentation and ambiguity is ideal, without the strict time constraint the researcher can experiment with different possible solutions, rejecting or refining as needed.

Ideally both commercial and academic development would have real users or usages. This is important because simply developing software for the purpose of development is pointless. Even a weak link to some real world problem such as for example attempting to process a set number of records more efficiently as a number of real world usages thus provides a much needed level of relevance.

Relevance and rigour are two important topics in academic development. Relevance of course is provided by linking the design and development effort to a real world problem, which provides the much needed context and requirements for the development. Rigour in conventional research is related to the validity and in most cases can be provided by recreating a given phenomenon in a controlled environment or by increasing the size and variety of the sample pool. This conventional rigour is difficult to transplant to academic development.

A potential solution is to link rigour of the process of creating a given software artefacts as well as their ability to solve a given problem. By using rigorous method an argument can be made that the artefacts are rigorous and by evaluating the artefacts in terms of their abilities to solve a given problem, a further argument for rigour can be made.

As for what criteria to use to ensure that the designs and developments are rigorous, it is wholly based-on the nature of the artefact and the overall objectives, which in turn is based-on the real world link that exists between the academic development effort and the addressed problem.

Within this research this real world link is elaborated and detailed using the ethnographic study. The ethnographic study is used to clarify the research problem and to create the research objectives, which then in turn lead to the design and development. But the ethnographic study also provides some of the criteria for evaluating the developed artefacts, but the primary evaluation criteria originated from the academic literature. By using criteria from the academic literature a certain amount of validity is assured and in this research the criteria were further refined using the ethnographic data.

9.2.5 Contributions of the DSR artefacts

This section details the contributions that originate from the DSR artefacts which were designed and developed as part of the research.

The design and development was aimed at creating a semantic metadata repository and the research was aimed at understanding what is required to create a semantic metadata repository. The topic of artefacts within DSR is detailed in Chapter 2 in Section 2.7.3. Within this research a number of artefacts were identified including: the solution artefact which is the prototype semantic metadata repository and the metadata model.

In terms of long term management, in regard to the repository the following requirements were identified:

- Repository content needs to be easily managed, both on a low granularity (an individual data-element) and on a high granularity (groups of related data-elements).
- If a large amount of content is expected within the repository the repository needs to have facilities that allow a user to easily interpret and understand a large amount of content in a fairly short period.
- Limit the content complexity if the content needs to be human managed.
- An understanding of how frequently the changes will occur.

In a situation where long term management is required within a repository solution along with frequent access to the content, it is important to provide means by which a user can easily gain an understanding of the content (in a situation where the user is not familiar with the repository content). In the semantic metadata repository a visualisation interface was constructed to visually represent the repository content. This aided in providing the users with a deeper understanding of the data-elements and their relationships which would otherwise have been time-consuming and difficult if the same information was presented in a number of plain lists.

Being able to easily understand and interpret the content is important but also is the ability for the user to manage the content at either a low of high granularity. An issue which was faced in the repository solution was the lack of ability to manage data at a higher level of granularity. The visualisation allowed for issues and inconsistencies to easily be identified but occasionally these issues occurred in several elements and each had to be altered individually.

A number of advantages exist in using the metadata to capture the details of the data-elements:

- Metadata works on a meta-level, meaning that more frequent changes to the data values does not necessarily need to be reflected on the meta-level.
- The metadata, especially if used to create semi-ontologies, is easy to visualise.
- Metadata can easily capture the data-elements and their relationships.
- Additional elements (such as data-element source information) can easily be added without directly influencing the metadata structures.
- The metadata itself can range in complexity, either being very minimalist or verbose. The more complex the metadata the less general applicable but the more potentially useful for a certain set of purposes.
- Semantics, which otherwise exist implicitly, can be captured and take on a meaningful role.

It was seen as being ideal to keep the metadata as minimalistic as possible. The value in the solution repository was not in the amount of data captured within a given metadata-element (corresponding to a data-element) but rather in how the data-elements relate to each other. Increases in the complexity of a metadata-element would potentially limit the usefulness. But the level of complexity of a given metadata-element is dependent on the particular context and intended usages. The primary focus was to keep the metadata-elements as generic and overall usable as possible and thus minimalist metadata-elements were chosen.

In terms of Evaluation three criteria originated from the academic literature, specifically: orientation, coverage, collaboration. These provided adequate broad topics for discussion but in terms of evaluation it was also deemed necessary to demonstrate the actual usage of the repository and demonstrate the outputs.

Chapter 10 Conclusion

10.1 Overview of Chapters

A generic chapter structure was used as a basis consistent of: introduction, literature review, research design, research process, research findings, discussion and conclusion. The generic chapter structure was further refined to align it with the chosen research methodology, which is detailed in Chapter 3 Section 3.7 and visualised in Figure 3.3.

In the thesis consists out of a total of 10 chapters. The first chapter the introduction provides the basic overview of the thesis; it details the research problem, motivation and research question. Although the content of the first chapter is only summarised and is further detailed later in the thesis, especially in chapter 3, the introduction chapter still offers a good starting point.

The second Chapter presents the literature review in accordance with the generic thesis structure. The literature review presents and discusses the academic literature which was deemed relevant to the research. The literature review fills multiple roles, the literature on HBHC and HI provides the need, justification and motivation for the research as provided in the academic literature. Further the literature review provides a source for the technological due-diligence providing a solution to design and development issues and questioned. Finally the literature also provided details and information for the research.

The third chapter provides the research design and details of the research process and surrounding topics. The third capture provides a detailed description of the research problem, the research question and objectives, the research methodology, ethical concerns of the research, research rigour, the scope of the research and how research data will be gathered and evaluated.

Following the DSR methodology the fourth Chapter corresponded to the first phase, that of Problem Identification, Understanding and Motivation. In order to achieve to identify, understand and motivate, an ethnographic study was conducted in the HBHC context. Although the ethnographic study to a lesser degree continued throughout the research to simplify its presentation, it was collectively included in the fourth chapter.

The fifth chapter corresponded to the second phase of the DSR methodology, the identification of the Research Focus and Objectives. This chapter makes use of the findings from the literature review and the ethnographic study to provide general objectives for both the research and the DSR artefacts that were developed as part of the research.

Chapter 6 encompasses 2 related phases of the DSR methodologies, specifically the conceptual design phase and the design and development phase. Chapter 6 details the process of creating initial solution concepts, creating more concrete designs and the process of instantiating the designs.

Chapter 7 demonstrates the instantiation of the design created and detailed in Chapter 6. Primarily the demonstration of the instantiation involved detailing and showing the UI components of the solution, going through the various UI and discussing the purpose of a particular interface. To provide a detailed demonstration of the instantiation the chapter also covers some of the more technical aspects such as the Code Classes and Code Class inheritance.

Chapter 8 corresponds to the fifth phase in the DSR methodology, the evaluation. The evaluation involves evaluating the artefacts created during the research process using criteria originating in the academic literature, but supported by the ethnographic findings. The evaluation only encompassed a handful of artefacts including: the overall solution artefact, the semantic structural metadata model and the usage and product of using the solution artefact.

Within Chapter 9 the overall research is reflected on and the primary research contributions are identified. Chapter 9 corresponds to the second last phase of the DSR methodology, research contribution.

10.2 Research Questions Answered

This Section revisits the research questions and attempts to provide a deeper understanding of how these questions were addressed in the research and why.

Each subsection encompasses one of sub-questioned, detailed in Table 3.1 in Chapter 3 Section 3.2.

10.2.1 What are the contextual implications for designing a repository of HBHC data-elements?

To better understand the contextual implication which influenced the design and development, the literature provided a good starting point but was not sufficiently detailed and relevant enough and thus an ethnographic study was conducted to support the findings of the literature review.

The ethnographic study provided not only the context for the design process but also provided a deeper understanding of the HBHC and of the care data-elements, information which in turn heavily aided in focusing and guiding the design process.

From this it is concluded that the contextualisation and using an ethnographic study were the most effective means of addressing this question.

10.2.2 How are the care data-elements and the relationships between them appropriately represented

In order to understand how care data-elements and the relationships between care data-elements can be understood, it was necessary to first gain a deeper understanding of these care data-elements and their particular context. This understanding of the care data-elements was provided by the previous question, detailed in 10.2.1.

From this understanding of the data-elements this question was addressed by using the academic literature. The academic literature provided a vast resource of potential solutions, which were narrowed using the ethnographic study and contextualisation.

The literature presented the topic of ontologies. Ontologies, detailed in Chapter 2 Section 2.6 provided a means of capturing concept and their relationships, further metadata which was detailed in Chapter 2 Section 2.3 and especially structural metadata provided a means of effectively capturing data for a single care data-element and its immediate relationships, which collectively would form a pseudo-ontology. In terms of offering optimal understanding of the data-elements it was also necessary to add a component of semantics, detailed in Section 2.5.

The repository concept, detailed in Chapter 2 Section 2.4, was added to provide a single means of storage and access to the ontology and the various data-elements which constitute it.

From the above it was concluded that an ontology was the appropriate means of representing the care data-elements and relationships, that semantic structural metadata was the appropriate means of capturing a single data-element (and collectively creating ontology).

10.3 How can the appropriate design considerations for a semantic metadata repository be identified within a research process

In the absence of pre-existing knowledge of a subject it is necessary to gain this knowledge through research. The research needs to use the most relevant and adequate methods to gain this knowledge. This is the question being addressed here.

In order to understand the design considerations of a semantic metadata repository it was deemed that the most appropriate means is to actually design a semantic metadata repository. But clearly differences exist between the research process and design and development process.

The literature offers a means of bridging the gap between research and design / development, specifically DSR which is detailed in Chapter 2 Section 2.7. However the literature presented numerous different DSR-based methodologies, but none met the requirements exactly.

From the above the research concluded that the best means of discovering the appropriate design considerations was to use a DSR based methodology which was an amalgamation of the existing methodologies, taking from these methodologies the most relevant and meaningful phases, as detailed in Chapter 3 Section 3.7.

10.3 Future Research

By defining and limiting the overall scope and focus of the research it ensures that the research does not grow too large and beyond the control of the researcher. In such a situation where the scope is too broad there is the possibility that too many questions might arise and too many variables might need to be taken into account. A limited scope provided necessary focus for the research and ensured that key priorities could be achieved.

The scope of the research was specifically focused toward understanding what was required to design an adequate semantic metadata repository within the home-based healthcare. The research specifically focused on the technical components. It was decided to focus only on the technical components of the research because the human components such as the user interface or the user's interaction with the system would have made the scope too wide to manage easily because these aspects could easily constitute their own research undertaking.

For that reason possible future research could look into the design of the user interface and user interaction for a semantic metadata repository. Ideally this potential path of research could draw on the current research allowing it to focus primarily on the design aspects of the user interface or user interaction components.

Because the focus research was on the overall design aspects the research only purposed a possible semantic structural metadata model to generate semi-ontologies for representing HBHC care-data-elements. From the academic literature, detailed in Chapter 2 Section 2.3.1 and Section 2.3.2, it was clear that semantic structural metadata is less prominent than other forms of metadata meaning that there is ample room for additional research.

Another avenue of potential future research is in either refining the semantic structural metadata model purposed in this research or even creating a new model for capturing semantic structural metadata within the HBHC domain or seeking potential usages outside of the HBHC domain.

Visualisation of the semantic metadata repository content was seen as being an advantage in that it eased the management, representation and comprehension of potentially large and complex repository content. The research however did not attempt to purpose an optimal means of visualising the repository content nor did it take into any greater consideration the various factors which could influence the visualisation of the repository content, such as: content type, content's level of detail, user characteristics, technological limitation etc. The visualisation component of the research only focused on the most basic information (data-elements, relationships, limited semantics etc.) excluding visualising most of the semantic and contextual information.

Future research can thus be done in terms of what is the most effective means of visualising the content of the semantic metadata repository as well as gaining an insight into the various factors which would influence the visualisation.

While conducting the research, various potential opportunities for future research in terms of DSR became apparent.

DSR has a component of design and development which all the sourced articles on the topic of DSR attest to, as listed in Table 3.3 in Section 3.7 of Chapter 3 and discussed in Section 2.7 of Chapter 2. However none of the sources make reference to any particular type of design and development methodology such as agile development, the waterfall model, spiral model etc. Thus little research exists which looks at the advantages and disadvantages of using a particular design and development methodology to conduct DSR and therefore this can be a potential opportunity for future research.

While conducting the literature review several methodologies, guidelines and frameworks for DSR were sourced, as detailed in Section 3.7 of Chapter 3. Some of these methodologies were refinements of those that came before but others were separate with no direct tie to previous methodologies. An amalgamated methodology was used in this research drawing the most relevant and meaningful aspects from the various methodologies in the literature. However this seems to indicate that more research can be done in terms of defining a standard DSR methodology, or more likely defining specific DSR methodology for specific research applications.

References

Al-Khalifa, H.S., Davis, H.C., 2006. The evolution of metadata from standards to semantics in E-learning applications. In Proceedings of the seventeenth conference on Hypertext and hypermedia. Odense, Denmark. ACM, pp. 69–72. Available at: http://portal.acm.org/citation.cfm?id=1149956 [Accessed February 12, 2011].

Allen, R.B., Schalow, J., 1999. Metadata and data structures for the historical newspaper digital library. Proceedings of the eighth international conference on Information and knowledge management - CIKM '99, pp.147-153. Available at: http://portal.acm.org/citation.cfm?doid=319950.319971 [Accessed June 25, 2010].

Ammenwerth, E., Graber, S., Herrmann, G., Burkle, T., König, J., 2003. Evaluation of health information systems–problems and challenges. International Journal of Medical Informatics, Vol. 71(2-3), pp.125–135. Available at: http://linkinghub.elsevier.com/retrieve/pii/S138650560300131X [Accessed February 12, 2011].

Andriessen, D., 2006. Combining design-based research and action research to test management solutions. In 7th World Congress Action Research. Groningen. Available at: http://www.weightlesswealth.com/downloads/Andriessen+DBR+and+Action+Research.PDF [Accessed February 12, 2011].

Aumüller, D., 2005. SHAWN: Structure helps a wiki navigate. In Proceedings of the BTW-Workshop WebDB Meets IR. Citeseer. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.112.3012&rep=rep1&type=pdf [Accessed February 10, 2011].

Bassiliades, N., Kokkoras, F., Vlahavas, I., Sampson, D., 2003. An intelligent educational metadata repository. Intelligent Systems, Techniques and Applications, Vol. 4, pp. 297–337. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.16.4909&rep=rep1&type=pdf [Accessed May 9, 2011].

Benjamins, V.R., Fensel, D., Pérez, A.G., 1998. Knowledge management through ontologies. In Proceedings of the Second International Conference on Practical Aspects of Knowledge Management. Basel, Switzerland: Citeseer, pp. 30. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.45.5505&rep=rep1&type=pdf [Accessed February 10, 2011].

Bernstein, P.A., Haas, LM., Jarke, M., Rahm, E., Wiederhold, G., 2000. Panel: Is generic metadata management feasible?, In Proceedings of the 26th International Conference on Very Large Databases. Cairo, Egypt, pp. 660-662.

Bertelsen, O.W., 2000. Towards a design-oriented epistemology. Scandinavian Journal of Information Systems, Vol. 12, pp.15-27.

Beyer, M., Kuhn, K.A., Meiler, C., Jablonski, S., Lenz, R.., 2004. Towards a flexible, process-oriented IT architecture for an integrated healthcare network. In Proceedings of the 2004 ACM symposium on Applied computing. Nicosia, Cyprus: ACM, pp. 264–271. Available at: http://portal.acm.org/citation.cfm?id=967958 [Accessed February 12, 2011].

Bharwani, S., 2006. Understanding Complex Behaviour and Decision Making using Ethnographic Knowledge Elicitation Kools (KnETs). Social Science Computer Review, Vol.

24(1), pp. 78. Available at: http://ssc.sagepub.com/content/24/1/78.abstract [Accessed February 10, 2011].

Billings, J.A., 1998. What is palliative care? Journal of Palliative Medicine, Vol. 1(1), pp. 73–81. Available at: http://www.liebertonline.com/doi/abs/10.1089/jpm.1998.1.73 [Accessed February 20, 2011].

Bird L., Heard S., J.W., 2005. Knowledge management in health care. Journal of health care finance, 31(3), pp.10–14.

Bowes, M., 2006. Forms, Applications, and Issues in Structural Metadata. Sciece Direct, pp.85-121. Available at: http://www.sciencedirect.com/science/article/B86P9-4PCJ07M-H/2/b39decdb13c8f02fc07179c656d2e928 [Accessed March 12, 2010].

Brandt, C.A., Gadagkar, R., Rodriguez, C., Nadkarni, P.M., 2004. Managing complex change in clinical study metadata. Journal of the American Medical Informatics Association, Vol. 11(5), pp. 380. Available at: http://jamia.bmj.com/content/11/5/380.full [Accessed February 12, 2011].

Brandt, C.A., Morse, R., Matthews, K., Sun, K., Deshpande, A.M., Gadagkar, R., Cohen, D.B. Miller, P.L., Nadkarni, P.M., 2002. Metadata-driven creation of data marts from an EAV-modeled clinical research database. International journal of medical informatics, Vol. 65(3), pp. 225–241. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1386505602000473 [Accessed February 12, 2011].

Bretherton, F.P., Singley, P.T., 2002. Metadata: A user's view. In Scientific and Statistical Database Management, 1994. Proceedings., Seventh International Working Conference on. IEEE, pp. 166–174. Available at: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=336950 [Accessed February 12, 2011].

Bryant, A., 2008, The future of information systems – Thinking Informatically, European Journal of Information Systems, Vol. 17, pp. 695–698.

Casanave, C., 2007. Designing a semantic repository: integrating Architectures for reuse and integration. In W3C workshop on eGovernment and the web. pp. 1–5. Available at: http://www.w3.org/2007/06/eGov-dc/papers/SemanticRepository.pdf [Accessed May 31, 2011].

Chen, H., Finin, T., Joshi, A., 2003, An Ontology for Context-Aware Pervasive Computing Environments. Special Issue on Ontologies for Distributed Systems, Knowledge Engineering Review.

Chetley A, 2006, Improving health, connecting people: the role of ICTs in the health sector of developing countries, The World Bank, Washington. Available at: http://www.healthlink.org.uk/PDFs/infodev_frame.pdf, [Accessed August 14, 2010].

Chandrasekaran, B., Josephson, J.R., Benjamins, V.R., 1999. What are ontologies, and why do we need them? Intelligent Systems and Their Applications, IEEE, Vol. 14(1), pp. 20–26. Available at: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=747902 [Accessed May 9, 2011].

Cho S., Mathiassen L., Nilsson A., 2008, Contextual dynamics during health information systems implementation: an event-based actor-network approach. European Journal of Information Systems, Vol. 17, pp. 614-630.

Chong Q., Marwadi A., Supekar K., Lee Y., 2003, Ontology Based Metadata Management in Medical Domains, Journal of Research and Practice in Information Technology (JRPIT), Vol. 35(2), pp. 139 - 154.

Chou, C.C.H. & Goethals, A., 2009. Document Metadata : document technical metadata for digital preservation. Library, pp.1-13.

Cleary, J., Holmes, G., Cunningham, S.J., Witten, I.H., 1996. MetaData for database mining. In Proceeding IEEE Metadata Conference. Citeseer. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.30.688&rep=rep1&type=pdf [Accessed February 12, 2011].

Cundiff, M.V., 2004. An introduction to the metadata encoding and transmission standard (METS). Library Hi Tech, Vol. 22(1), pp. 52–64. Available at: http://www.emeraldinsight.com/journals.htm?articleid=861404&show=abstract [Accessed February 12, 2011].

Darmoni, S.J., Thirion, B., 2000. A standard metadata scheme for health resources, Journal of the American Medical Informatics Association, Vol. 7(1), pp.108. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC61461/ [Accessed February 12, 2011].

Davidson, E., Reardon, J.L. 2005. Organizing vision for IT healthcare: Analysis of the discourse surrounding electronic health records. Paper presented at the Academy of Management Annual Meeting, Honolulu, HI.

Davies, J., Studer, R., Sure, Y., Warren, P.W., 2005. Next generation knowledge management. BT Technology Journal, Vol. 23(3), pp.175–190. Available at: http://www.springerlink.com/index/K3X867H827314810.pdf [Accessed May 31, 2011].

Delen, A., 2010, Service Design Challenges in Home Based Healthcare In the Western Cape: A Case Study, Masters Thesis, Cape Peninsula University of Technology.

Devedzic, V., 2002, Understanding Ontological Engineering. Communications of the ACM, Vol. 45, pp. 136-144.

Dixon, D.R., Dixon, B., 1994. Adoption of information technology enabled innovations by primary care physicians: model and questionnaire development. In Proceedings of the Annual Symposium on Computer Application in Medical Care. American Medical Informatics Association, pp. 631-635. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2247957/ [Accessed February 12, 2011].

Dogac, A., Kabak, Y., Laleci, G., 2004, Enriching ebXML Registries with OWL Ontologies for E_cient Service Discovery", in Proceedings of RIDE'04, Boston, USA.

Dushay, N., 2001. Using structural metadata to localize experience of digital content. Arxiv preprint cs/0112017, pp.1-23. Available at: http://arxiv.org/abs/cs.DL/0112017 [Accessed February 12, 2011].

Dushay, N., 2002. Localizing experience of digital content via structural metadata. In Proceedings of the second ACM/IEEE-CS joint conference on Digital libraries - JCDL '02. New York, New York, USA: ACM Press, pp. 244. Available at: http://portal.acm.org/citation.cfm?doid=544220.544274 [Accessed July 5, 2010].

Ellis, T.J., Levy, Y., 2010. A Guide for Novice Researchers: Design and Development Research Methods. Citeseer. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.170.2962&rep=rep1&type=pdf [Accessed February 12, 2011].

Evans, R., Stone, D., Elwyn, G., 2003. Organizing palliative care for rural populations: a systematic review of the evidence. Family Practice, Vol. 20(3), pp. 304. Available at: http://fampra.oxfordjournals.org/content/20/3/304.abstract [Accessed February 10, 2011].

Eysenbach, G., 2000. Consumer health informatics. Bmj, Vol. 320(7251), pp.1713. Available at: http://www.bmj.com/content/320/7251/1713.short [Accessed March 21, 2011].

Eysenbach, G, 2003, The Semantic Web and healthcare consumers: a new challenge and opportunity on the horizon?, International Journal of Healthcare Technology and Management, Vol 5(3-5), pp. 194-212.

Euzenat J, 2001, Towards a Principled Approach to Semantic Interoperability, Workshop on Ontologies and Information Sharing, IJCAI'01, Seattle USA.

Farquhar, A., Fikes, R., Pratt, W., Rice, J., 1995. Collaborative ontology construction for information integration. Technique Reports of Knowledge Systems Laboratory, Department of Computer Science, KSL-95, Vol. 63. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.50.5303&rep=rep1&type=pdf [Accessed May 9, 2011].

Fraser, H.S.F. et al., 2005. Refereed papers Implementing electronic medical record systems in developing countries. Informatics in Primary Care, 13, pp.83-95. Available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.135.1906 [Accessed February 12, 2011].

Garde, S., Knaup, P., Hovenga, E.J.S., Heard, S., 2007. Towards semantic interoperability for electronic health records: Domain knowledge governance for openEHR archetypes. Medicine, Vol. 46(3), pp. 332–343. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.83.4127&rep=rep1&type=pdf [Accessed May 31, 2011].

Gartner, R., 2002. METS: Metadata Encoding and Transmission Standard. JISC Techwatch Report TSW, pp. 2–5. Available at: http://www.jisc.ac.uk/media/documents/techwatch/tsw_02-05.pdf [Accessed February 10, 2011].

Gillman, D. and Appel, M. (1997). The Statistical Metadata Repository: an electronic catalog of survey descriptions at the U.S. Census Bureau. IASSIST Quarterly, Vol. 21(2), pp. 34-51.

Godby, C.J., Smith, D., Childress, E., 2003. Two paths to interoperable metadata, Citeseer. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.76.8133&rep=rep1&type=pdf [Accessed February 12, 2011].

Greenberg, J., 2008. Metadata Extraction and Harvesting: A comparison of Two Automatic Metadata Generation Applications. The Medical clinics of North America, Vol. 6(4), pp. 59-

82. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2245891/ [Accessed February 10, 2011].

Gregor, S., 2002. Design theory in information systems. Australian Journal of Information Systems, 10(December), pp.14–22. Available at: http://dl.acs.org.au/index.php/ajis/ article/viewPDFInterstitial/439/399?ads= [Accessed February 12, 2011].

Gregor, S., Jones, D., 2007. The anatomy of a design theory. Journal of the Association for Information Systems, Vol. 8(5), pp. 312–335. Available at: http://seal.tst.adfa.edu.au/~ejl/ Portal/Systemsplanning/SP/pages/gregor_07_design_theory.pdf [Accessed February 12, 2011].

Grimson, J., 2001, Delivering the electronic healthcare record for the 21st century, International Journal of Medical Informatics, Vol. 64(2-3), pp.111–127. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1386505601002052 [Accessed May 9, 2011].

Harper, R.H.R., O'Hara, K.P.A., Sellen, A.J., Duthie, D.J.R.,1997, Toward the paperless hospital? British journal of anesthesia, Vol. 78(6), pp. 762. Available at: http://bja.oxfordjournals.org/content/78/6/762.short [Accessed February 12, 2011].

Hart, T., 1971. The inverse care law. The Lancet, Vol. 297(7696), pp. 405–412. Available at: http://linkinghub.elsevier.com/retrieve/pii/S014067367192410X [Accessed February 12, 2011].

Hartmann, J, 2006, ONTHO`LOGY. An Ontology Metadata Repository, In Demo and Poster Proceedings of ESWC'2006, Budva, Montenegro.

Hasan, H., 2004. Information systems development as a research method, Journal of Information Systems, pp. 4-13. Available at: http://dl.acs.org.au/index.php/ajis /article/viewPDFInterstitial/142/122?ads= [Accessed February 12, 2011].

Hatala, M., Richards, G., 2002. Global vs. community metadata standards: Empowering users for knowledge exchange. The Semantic Web—ISWC 2002, pp. 292–306. Available at: http://www.springerlink.com/index/X3ENK5QMBCGWT4MW.pdf [Accessed February 12, 2011].

Haux, R., 2006. Health information systems - past, present, future, International journal of medical informatics, Vol. 75(3-4), pp. 268-281. Available at: http://www.ncbi.nlm.nih.gov /pubmed/16169771 [Accessed July 20, 2011].

Holland C.P., Light B., Wills K., 2001, ERP and best of breed: a comparative analysis, Business Process Management Journal, Vol. 7(4), pp. 216-224.

Iivari, J., 2007. A paradigmatic analysis of information systems as a design science. Scandinavian Journal of Information Systems, Vol. 19(2), pp. 39. Available at: http://www.hec.unil.ch/yp/DRIS/Articles/Iivari07.pdf [Accessed February 12, 2011].

Heflin, J., Hendler J., 2000, Semantic interoperability on the web, In Extreme Markup Languages'2000, Available At: http://www.cs.umd.edu/projects/plus/SHOE/pubs/extreme 2000.pdf [Accessed March 15, 2011]

Hersh W., 2004, Health care information technology: progress and barriers. JAMA. 2004, Vol. 292(18), pp. 2273–2274.

Hevner, A.R. et al., 2004. Design science in information systems research. Mis Quarterly, 28(1), pp.75–105. Available at: http://www.jstor.org/stable/25148625 [Accessed May 31, 2011].

Hodge G., 2001, Metadata made simpler: a guide for libraries, NISO – Press, Availability at: http://www.ncsi.iisc.ernet.in/raja/is214/214-2001-2002/Metadata_Bklt.pdf [Accessed February 10, 2011].

Hongoro, C., McPake, B., 2004. How to bridge the gap in human resources for health. Lancet, Vol. 364, pp.1451–1456. Available at: http://info.worldbank.org/etools/docs/library/206847/HONGORO.pdf [Accessed February 20, 2011].

Huang, J., Jennings, N., Fox, J., 1994, Cooperation in distributed medical care, In Proceedings of Second International Conference on Cooperative Information Systems, Toronto, Canada, Citeseer, pp. 255–263. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.37.5500&rep=rep1&type=pdf [Accessed March 21, 2011].

Huynh T.N., Mangisengi O., Tjoa A.M. 2000, Metadata for Object–Relational Data Warehouse Proceedings of the Second International Workshop on Design and Management of Data Warehouses, DMDW'2000, Stockholm, Sweden.

Hudson, P., 2003. Home-based support for palliative care families: challenges and recommendations. Medical Journal of Australia, Vol. 179(6), pp. 35. Available at: http://www.mja.com.au/public/issues/179_06_150903/hud10296_fm.pdf [Accessed February 20, 2011].

Iivari, J., 2007. A paradigmatic analysis of information systems as a design science. Scandinavian Journal of Information Systems, Vol. 19(2), pp. 39. Available at: http://www.hec.unil.ch/yp/DRIS/Articles/Iivari07.pdf [Accessed February 12, 2010].

Ingenerf, J., Reiner, J., Seik, B., 2001, Standardized terminological services enabling semantic interoperability between distributed and heterogeneous systems, International journal of medical informatics, Vol. 64(2-3), pp. 223–240. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1386505601002118 [Accessed May 31, 2011].

Inmon, W.H., O'Neil, B.K., Fryman, L., 2008, Business Metadata: Capturing Enterprise Knowledge, Morgan Kaufmann, Burlington, US

Jacucci, E., Shaw, V., Braa, J., 2006. Standardization of health information systems in South Africa: The challenge of local sustainability. Information Technology for Development, Vol. 12(3), pp. 225–239.

Jones D. M., Bench-Capon T.J.M., Visser P.R.S., 1998, Methodologies for ontology development. In Proceeding IT & KNOWS Conference of the 15th IFIP World Computer Congress, Budapest, Chapman-Hall.

Kaur, H., Wasan, S.K., 2006. Empirical study on applications of data mining techniques in healthcare. Journal of Computer Science, Vol. 2(2), pp.194–200. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.108.8981&rep=rep1&type=pdf [Accessed May 31, 2011].

Keller, R.M., Berrios, D.C., Carvalho, R.E., Hall, D.R., Rich, S.J.,Sturken, I.B., Swanson, K.J., Wolfe, S.R, 2004, Semantic Organizer: A Customizable Semantic Repository for Distributed NASA Project Teams. In proceeding of Third International Semantic Web Conference, Hiroshima,Japan. pp. 767–781

Kerhervé, B. & Gerbé, O., 1997. Models for metadata or metamodels for data? In Proceedings of 2nd IEEE Metadata Conference, Silverpines, Maryland, USA, Citeseer, pp. 1-12. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.83.1104& rep=rep1&type=pdf [Accessed February 12, 2011].

Kim, W., 2005. On metadata management technology: status and issues. Journal of Object Technology, Vol. 4(2), pp. 41–47. Available at: http://citeseerx.ist.psu.edu/viewdoc/download ?doi=10.1.1.97.2314&rep=rep1&type=pdf [Accessed February 12, 2011].

Kim, J., Schwarm, S.E., Ostendorf, M., 2004. Detecting structural metadata with decision trees and transformation-based learning. In Proceeding of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, HLT/NAACL'2004, Boston, Massachusetts, pp. 137–144. Available at: http://acl.ldc. upenn.edu/hlt-naacl2004/main/pdf/100_Paper.pdf [Accessed February 12, 2011].

Kiryakov, A., Popov, B., Terziev, I., Manov, D., Ognyanoff, D., 2004. Semantic annotation, indexing, and retrieval. Web Semantics: Science, Services and Agents on the World Wide Web, Vol. 2(1), pp. 49–79. Available at: http://linkinghub.elsevier.com/retrieve /pii/S1570826804000162 [Accessed May 31, 2011].

Klaas, N.P., 2007. Factors influencing the retention of nurses in the rural health facilities of the Eastern Cape Province. Available at: http://uir.unisa.ac.za/dspace/handle/10500/544 [Accessed February 20, 2011].

Kuechler, B. & Vaishnavi, V., 2008. On theory development in design science research: anatomy of a research project. European Journal of Information Systems, 17(5), pp.489–504. Available at: http://www.palgrave-journals.com/ejis/journal/v17/n5/abs/ejis200840a.html [Accessed February 12, 2011].

Phokion Kolaitis, 2005, Schema mappings, data exchange, and metadata management. In Proceedings of ACM PODS, pp. 61–75, Baltimore, Maryland, USA.

Kuhn K.A., Guise D.A. 2001, From hospital information systems to health information systems: problems, challenges, perspectives. Methods of information in medicine, Vol. 40, pp. 275–287.

Lewin, S., Dick, J., Zwarenstein, M., Aja, G., Van Wyk, B., Bosch-Capblanch, X., Patrick, M., 2009. Lay health workers in primary and community health care. Cochrane Database of Systematic Reviews, Vol. 1(3). Available at: http://apps.who.int/entity/rhl/reviews/ CD004015.pdf [Accessed February 20, 2011].

Lewy T., Glaser H., Shadbolt N., 2005, A Framework for Reference Management in the Semantic Web, Technical Report, University of Southampton, Available at: http://eprints.ecs.soton.ac.uk/11539/ [Accessed May 31, 2010].

Li, Q., Shilane, P., Noy, N.F., Musen, M.A., 2000. Ontology acquisition from on-line knowledge sources. In Proceedings of the AMIA Symposium. American Medical Informatics Association, pp. 497. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2244130/ [Accessed May 31, 2011].

Litsios S., 2004, Primary Health Care, WHO and the NGO Community, Society for International Development, Vol. 47(2), pp. 57-63.

Ludwick DA, Doucette J., 2008, Adopting electronic medical records in primary care: lessons learned from health information systems implementation experience in seven countries. International Journal Medical Informatics, Vol. 78(1), pp. 22–31.

March, S.T., Smith, G.F., 1995. Design and natural science research on information technology. Decision Support Systems, 15(4), pp.251–266. Available at: http://linkinghub.elsevier.com/retrieve/pii/0167923694000412 [Accessed February 12, 2011].

March, S.T., Storey, K.G., 2008, Design Science in the Information Systems Discipline: An introduction to the special issue on design science research, MIS Quarterly, Vol. 32(4), pp. 725-730.

Marenco, L., Tosches, N., Crasto, C., Shepherd, G., Miller, P.L., Nadkarni, P.M., 2003. Achieving evolvable Web-database bioscience applications using the EAV/CR framework: recent advances. Journal of the American Medical Informatics Association, Vol. 10(5), pp. 444. Available at: http://jamia.bmj.com/content/10/5/444.full [Accessed February 12, 2011].

Marshall, C.C., 1998, Making Metadata: a study of metadata creation for a mixed physical-digital collection. In Proceedings of the 3rd ACM conference on Digital libraries. ACM, Pittsburgh, PA, USA, pp. 162–171. Available at: http://portal.acm.org/citation.cfm?id=276675.276693&type= series [Accessed February 12, 2011].

Masseroli, M., Visconti, A., Bano, S.G., Pinciroli, F., 2006, He@lthCo-op: A web-based system to support distributed healthcare co-operative work. Computers in biology and medicine, Vol. 36(2), pp.109-127. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16389072 [Accessed October 21, 2010].

Mrtenson C., Horndahl A., 2008, Using semantic technology in intelligence analysis. In Skvde Workshop on Information Fusion Topics (SWIFT 2008), pp. 4-6, Skvde.

Martinez A., Villarroel V., Seoane J., del Pozo F., 2005, Analysis of Information and Communication Needs in Rural Primary Care in Developing Countries, IEEE Transactions on Information Technology in Biomedicine, Vol. 9(1), pp. 66 – 72.

Materstvedt, L.J.,Clark, D., Ellershaw, J., Forde, R., Gravgaard, A.M.B., Muller-Busch, H.C., Sales, J.P., Rapin, C.H., 2003. Euthanasia and physician-assisted suicide: a view from an EAPC Ethics Task Force. Palliative Medicine, Vol. 17(2), pp. 97. Available at: http://pmj.sagepub .com/content/17/2/97.full.pdf [Accessed February 20, 2011].

Mead CN, 2006, Data Interchange standards in healthcare IT –computable semantic interoperability: Now possible but still difficult do we really need a better mousetrap?, Journal Healthcare Information Management, Vol. 20(1), pp. 71-78.

Mödritscher, F., Hoffmann, R. and Klieber, W., 2007, Integration and semantic enrichment of explicit knowledge through a multimedia, multi-source, metadata-based knowledge artefact repository, Proceedings of the International Conference on Knowledge Management (I-Know 2007), Graz, pp.365–372.

Moss L.T., 2007, Critical Success Factors for Master Data Management, Cutter IT Journal, Vol. 20(9), pp. 7-13.

Musen MA, 1998, Modern Architectures for Intelligent Systems: Reusable Ontologies and Problem-Solving Methods, Journal of the American Medical Informatics Association, 1998(Symposium supplement), pp. 46-54.

Musen MA., 2001, Creating and Using Ontologies: What Informatics is all About, In: Proceedings of Medinfo 2001, pp. 1514, London.

Najjar J., Wolpers M., Duval E., 2006, Attention Metadata: Collection and Management, Distributed Systems Institute – Knowledge Based Systems. Hannover Germany, Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.61.7010&rep=rep1&type=pdf [Accessed July, 2010].

Neches, R., Fikes, R.E., Finin, T., Gruber, T., Patil, R., Senator, T., Swartout, W.R., 1991. Enabling technology for knowledge sharing. AI magazine, Vol. 12(3), pp. 36. Available at: http://www.aaai.org/ojs/index.php/aimagazine/article/viewArticle/902 [Accessed May 31, 2011].

Niehaves, B., 2007. On epistemological diversity in design science: New vistas for a design-oriented IS research. In Proceedings of the 28th International Conference on Information Systems, Montreal, Canada: Association for Information Systems. Montreal, Canada: Citeseer, pp. 1-13. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.86.8058&rep=rep1&type=pdf [Accessed February 12, 2011].

NISO Press, 2004, Understanding Metadata, National Information Standards, Available at: www.niso.org/publications /press/UnderstandingMetadata.pdf [Accessed March 3, 2010].

Noy, N.F., McGuinness D.L., 2001, Ontology development 101: A guide to creating your first ontology. Technical Report KSL-01-05, Stanford Knowledge Systems Laboratory. Available at: http://citeseer.ist.psu.edu/viewdoc/download;jsessionid=82435ADD2D96C0EAB244025FA293FA9C?doi=10.1.1.136.5085&rep=rep1&type=pdf [Accessed March 26, 2010].

Obrst, L. 2003. Ontologies for Semantically Interoperable Systems. Proceedings of the Twelfth ACM International Conference on Information and Knowledge Management (CIKM 2003), Ophir Frieder, Joachim Hammer, Sajda Quershi, and Len Seligman, eds.,New Orleans, LA, November 3-8, New York: ACM, pp. 366-369.

Onyancha, I., Keizer, J., Katz, S., 2001. A Dublin Core application profile in the agricultural domain. In Proceedings of the International Conference on Dublin Core and Metadata Applications, DC-2001, Tokyo, Japan, Citeseer, pp. 185–192. Available at: http://citeseerx.

ist.psu.edu/viewdoc/download?doi=10.1.1.143.8213&rep=rep1&type=pdf [Accessed February 12, 2011].

Padarath, A. et al., 2003, Health Personnel in Southern Africa : Confronting maldistribution and brain drain. Production, (1), pp.1-41. Available at: http://www.queensu.ca/samp/migrationresources/braindrain/documents/equinet.pdf [Accessed June 13, 2010].

Päivärinta, T., Tyrväinen, P., Ylimäki, T., 2002. Defining Organizational Document Metadata: A Case Beyond Standards. In Proceedings of 10th European Conference on Information Systems (ECIS). Gdańsk, Poland: Citeseer, pp. 1154-1163. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.106.9745&rep=rep1&type=pdf [Accessed February 12, 2011].

Palmonari, M., Viscusi, G., Batini, C., 2008. A semantic repository approach to improve the government to business relationship. Data & Knowledge Engineering, Vol. 65(3), pp. 485–511. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0169023X08000037 [Accessed May 31, 2011].

Pantel P., Pennancchiotti M., 2008, Automatically harvesting and Ontologizing Semantic Relations, Proceeding of the 2008 Conference on Bridging the Gap between Text and Knowledge - Selected Contributions to Ontology Learning and Population from Text. IOS Press.

Park, J.R., 2009. Metadata quality in digital repositories: A survey of the current state of the art. Cataloging & Classification Quarterly, Vol. 47(3), pp. 213–228.

Park, Y.R., Kim, J.H., 2006. Metadata registry and management system based-on ISO 11179 for cancer clinical trials information system. In Symposium A Quarterly Journal In Modern Foreign Literatures. American Medical Informatics Association, pp. 11197-11197.

Patridge, J., Namulanda, G., 2008. Describing environmental public health data: implementing a descriptive metadata standard on the environmental public health tracking network. Journal of Public Health Management and Practice, Vol. 14(6), pp. 515. Available at: http://journals.lww.com/jphmp/Abstract/2008/11000/Describing_Environmental_Public_Health_Data_.3.aspx [Accessed February 12, 2011].

Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S., 2007. A design science research methodology for information systems research. Journal of Management Information Systems, Vol. 24(3), pp. 45–77. Available at: http://mesharpe.metapress.com/index/276818W6PN4T5483.pdf [Accessed February 12, 2011].

Pouloudi, A, 1999, Information technology for collaborative advantage in healthcare revisited. Information and Management, Vol. 35, pp. 345-356

Prahallad, L., Thammishetty, I.,Raghavendra, E.V., Ambati, V., 2006. A tool for Entering Structural Metadata in Digital Libraries. In International conference on Universal Digital Library - 2006. Alexandria, Egypt. Available at: http://www.ulib.org/conference/2006/28.pdf.

Purgathofer, P., 2006. Is informatics a design discipline? Poiesis & Praxis: International Journal of Technology Assessment and Ethics of Science, 4(4), pp.303–314. Available at: http://www.springerlink.com/index/37hj74n5t03g5717.pdf [Accessed February 12, 2011].

Rehman, M., Anwer, M.H., Iftikhar, N., 2005. Universal Metadata Definition. Citeseer, Vol. 5(April), pp.144-146. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.96.5447&rep=rep1&type=pdf [Accessed February 12, 2011].

Rizvi, S.A.M., Imdadi, N., 2008. Ontology integration paradigms for Automatic Semantic Integration Incorporating Semantic Repositories. In Proceedings of Second National Conference on Challenges & Opportunities in Information Technology, Mandi Gobindgarh, Punjab, pp. 291-294. Available at: http://www.rimtengg.com/coit2008/proceedings/MS13.pdf [Accessed May 31, 2011].

Rowe, M., 2008. Information and communication technology in health: a review of the literature. Journal of Community and Health Sciences, Vol. 3(1), pp. 68–77. Available at: http://repository.uwc.ac.za/xmlui/handle/10566/61 [Accessed February 12, 2011].

Schorlemmer, M., Kalfoglou, Y., 2003. On semantic interoperability and the flow of information. In Proceedings of the ISWC'03 Workshop on Semantic Integration, Sanibel Island, Florida, USA, Citeseer, pp. 80–86. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.10.7533&rep=rep1&type=pdf [Accessed May 31, 2011].

Smail-Tabbone, M., Osman, S., Messai, N., Napoli, A.,Devignes, M.D., 2005, BioRegistry: a structured metadata repository for bioinformatic databases. Computational Life Sciences, (i), pp. 46–56. Available at: http://www.springerlink.com/index/2h80xagt0kumu5m0.pdf [Accessed February 12, 2011].

Sourouni AM., Kourlimpinis G., Mouzakitis S., Askounis D., 2009, Towards the Government transformation: An ontology-based government knowledge repository, Journal Computer Standards & Interfaces , Vo.l 32(1-2), pp. 44-53

Srivastava, D., Velegrakis, Y., 2007. Intensional associations between data and metadata. In Proceedings of SIGMOD/PODS '07 International Conference on Management of Data, Beijing, China, pp. 401–412. Available at: http://portal.acm.org/citation.cfm?id=1247526 [Accessed February 12, 2011].

Stead, W.W., Miller, R.A., Musen, M.A., Hersh, W., 2000, Integration and Beyond: Linking Information from Disparate Sources and into Workflow. Journal of the American Medical Informatics Association, Vol. 7(2), pp.135-148. Available at: http://www.jamia.org/cgi/content/abstract/7/2/135 [Accessed September 12, 2010].

Stjernswärd, J., Foley, K.M., Ferris, F.D., 2007, The public health strategy for palliative care. Journal of pain and symptom management, Vol. 33(5), pp. 486-93. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17482035 [Accessed September 16, 2010].

Schadow, G., Mead, C.N., Walker, D.M., 2006, The HL7 reference information model under scrutiny. Studies in Health Technology and Informatics, Vol. 124, pp. 151-156. Available at: http://books.google.com/books?hl=en&lr=&id=Owo85gdfJawC&oi=fnd&pg=PA151&dq=The+HL7+Reference+Information+Model+Under+Scrutiny&ots=xE-q9qeqv-&sig=9doqryVFBrwJCAAv4QJWS9wn55A [Accessed February 12, 2011].

Schneider, H., Hlophe, H., van Rensburg, D., 2008. Community health workers and the response to HIV/AIDS in South Africa: tensions and prospects. Health policy and planning, Vol 23(3), pp.179-87. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18388133 [Accessed September 6, 2010].

Schuurman, N., Leszczynski, A., 2006. Ontology-based metadata. Transactions in GIS, Vol 10(5), pp. 709–726. Available at: http://www3.interscience.wiley.com/journal/118566640/abstract [Accessed February 12, 2011].

Schoeman, S., Smuts, C.M., Faber, M., Van Stuijvenberg, M., Oelofse, A., Laubscher, J.A., Benadé, A.J.S., Dhansay, M.A., 2010. Primary health care facility infrastructure and services and the nutritional status of children 0 to 71 months old and their caregivers attending these facilities in four rural districts in the Eastern Cape and KwaZulu-Natal provinces, South Africa. South African Journal of Clinical Nutrition, Vol 23(1), pp. 21-27. Available at: http://ajol.info/index.php/sajcn/article/view/52744/0 [Accessed March 6, 2011].

Smith, B., Ceusters, W., 2006. HL7 RIM: An incoherent standard. Studies in health technology and informatics, Vol 124(August), pp.133–138. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.145.4676&rep=rep1&type=pdf [Accessed February 12, 2011].

Strasser, R., 2003. Rural health around the world: challenges and solutions. Family practice, Vol 20(4), pp. 457. Available at: http://fampra.oxfordjournals.org/cgi/content/abstract/20/4/457 [Accessed February 20, 2011].

Sun, K., Nadkarni, P.M., Brandt, C.A., 2005. Metadata Driven Generation of Reports for Clinical Studies. In Symposium A Quarterly Journal In Modern Foreign Literatures. American Medical Informatics Association, pp. 2005-2005.

Taylor T.A., Parish J.R., Roderer N., 2009, Career Opportunities in Library and Information Science, Infobase Publishing, New York.

Thibodeau, K., 2006. What constitutes success in a digital repository ? In Joint Conference on Digital Libraries (JCDL) Workshop on Digital Curation and Institutional Repositories. Chapel Hill, NC. Available at: http://www.ils.unc.edu/tibbo/JCDL2006/Thibodeau-JCDL Workshop2006.pdf [Accessed February 5, 2011].

Thirion, B., Douyère, M., Soualmia, L.F., Dahamna, B., Leroy, J.P., Darmoni, S.J., 2004. Metadata element sets in the CISMeF quality-controlled health gateway. In Proceedings of the 2004 international conference on Dublin Core and metadata applications: metadata across languages and cultures. Dublin Core Metadata Initiative, pp. 1–12. Available at: http://portal.acm.org/citation.cfm?id=1383412 [Accessed February 12, 2011].

Tomasi, E., Facchini, L.A., Maia, M.D.F.S., 2004. Health information technology in primary health care in developing countries: a literature review. Bulletin of the World Health Organization, Vol 82(11), pp. 867-874. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2623061&tool=pmcentrez&rendertype=abstract [Accessed February 3, 2011].

Togno, J., Topps, D., 2000. Policy on using information technology to improve rural health care, Advanced Learning Technology: Design and Development Issues, pp. 260-261.

Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=890627 [Accessed March 2, 2011]

Timpf, S., Raubal, M., Kuhn, W., 1996. Experiences with Metadata. In Proceedings of the 7th International Symposium on Spatial Data Handling SDH'96 Advances in GIS Research II, Delft, The Netherlands, pp. 12B31-12B43. Available at: http://citeseerx.ist.psu.edu/viewdoc/ download?doi=10.1.1.19.5663 &rep=rep1&type=pdf [Accessed March 2, 2011].

Van Ginneken, A.M., 2003. Considerations for the representation of meta-data for the support of structured data entry. Methods of information in medicine, Vol 42(3), pp. 226–235. Available at: http://www.schattauer.de/de/magazine/uebersicht/zeitschriften-a-z/methods/ contents/current-issue/issue/special/manuscript/282/download.html [Accessed June 12, 2010].

van Zyl, I., 2011. Mutual Isolation and the Fight for Care An Ethnography of South African Home- based Healthcare Contexts. Journal of Health Informatics in Developing Countries, Vol. 5(1), pp.15-37.

Wagner, H., Weibel, S., 2005. The Dublin Core Metadata Registry: Requirements, Implementation, and Experience, Journal of Digital Information, Vol. 6(2).

Wang, S., Wang, H., 2010. Towards Innovative Design Research in Information Systems. Journal of Computer Information Systems, Vol. 51(1), pp.11-18. Available at: http://www. iacis.org/jcis/pdf/Wang_Wang_51_1.pdf [Accessed August 10, 2010].

Wache, H., Vogele, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H., Hubner, S., 2001. Ontology-based integration of information-a survey of existing approaches. In IJCAI-01 workshop: ontologies and information sharing. Citeseer, pp. 108–117. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.142.4390&rep=rep1&type=pdf [Accessed February 12, 2011].

Wouters, B., Barjis, J., Maponya, G., Maritz, J. Mashiri, M., 2009. Supporting home based health care in South African rural communities using USSD technology. In Information Systems. San Francisco, California, USA: Association for Information Systems. Available at: http://researchspace.csir.co.za/dspace/handle/10204/3933 [Accessed June 12, 2010].

CAPE PENINSULA
UNIVERSIT[illegible]HNOLOGY