



University of Bradford eThesis

This thesis is hosted in [Bradford Scholars](#) – The University of Bradford Open Access repository. Visit the repository for full metadata or to contact the repository team



© University of Bradford. This work is licenced for reuse under a [Creative Commons Licence](#).

A Dementia Care Mapping (DCM) data warehouse as a resource for improving the quality of dementia care

Exploring requirements for secondary use of DCM data using a
user-driven approach and discussing their implications for a
data warehouse

Shehla KHALID

Submitted for the Degree of
Doctor of Philosophy

Faculty of Health Studies
University of Bradford

2016

Abstract

Shehla Khalid

A Dementia Care Mapping (DCM) data warehouse as a resource for improving the quality of dementia care

Exploring requirements for secondary use of DCM data using a user-driven approach and discussing their implications for a data warehouse

Keywords

Dementia Care Mapping, DCM, Data warehouse, grounded theory, secondary use, user-driven, requirement analysis, requirement gathering, data management, secondary research

The secondary use of Dementia Care Mapping (DCM) data, if that data were held in a data warehouse, could contribute to global efforts in monitoring and improving dementia care quality. This qualitative study identifies requirements for the secondary use of DCM data within a data warehouse using a user-driven approach. The thesis critically analyses various technical methodologies and then argues the use and further demonstrates the applicability of a modified grounded theory as a user-driven methodology for a data warehouse. Interviews were conducted with 29 DCM researchers, trainers and practitioners in three phases. 19 interviews were face to face with the others on Skype and telephone with an average length of individual interview 45-60 minutes. The interview data was systematically analysed using open, axial and selective coding techniques and constant comparison methods.

The study data highlighted benchmarking, mappers' support and research as three perceived potential secondary uses of DCM data within a data warehouse. DCM researchers identified concerns regarding the quality and security of DCM data for secondary uses, which led to identifying the requirements for additional provenance, ethical and contextual data to be included in a warehouse alongside DCM data to meet requirements for secondary uses of this data for research. The study data was also used to extrapolate three main factors such as an individual mapper, the organization and an electronic data management that can influence the quality and availability of DCM data for secondary uses. The study makes further recommendations for designing a future DCM data warehouse.

Acknowledgements

Though a PhD thesis is the outcome of one's independent work and only my name appears on the cover of this thesis, a great many people have contributed to its production. It is therefore important to thank them all. I owe my gratitude to all those people who have made this thesis possible.

I would like to thank my supervisory team, Prof. Claire Surr from Leeds Beckett University, Prof. Daniel Neagu from the Faculty of Engineering and Computing (FoEC) and Prof. Neil Small from the Faculty of Health Studies (FoHS). I have been amazingly fortunate to have this professorial team from a variety of backgrounds, with expertise in their own fields, to advise, encourage, assist and supervise me during these years of creating, developing and completing this multidisciplinary work, which commenced with a vision to bring IT into health to support quality of dementia care. My supervisors gave me the freedom to explore on my own, and at the same time provided their guidance when I needed direction. Their insightful comments, constructive criticism at different stages of this study were thought provoking and they helped me to focus on my ideas and channel my thoughts. I am grateful to them for their practical advice, encouragement and moral support. I am also thankful to them, particularly Claire and Neil, for reading my drafts, commenting on my views, and helping me enrich my ideas and improving my writing skills.

I would also like to thank Jo Crossland and Sharon from the BDG (now School of Dementia Studies) to help me learn about DCM. They enhanced my understanding of primary and secondary use of DCM data through in-depth discussions.

I would also like to thank Dr Rana Tassabehji from the Faculty of Management and Law, who gave her time for useful discussions regarding the importance of user requirements in information systems and application of grounded theory for requirement analysis. Big thanks also goes to my friends Dr Katherine Ludwin and Dr Sophia Alim for reading some of my drafts and providing helpful feedback and comments.

I would also like to thank my fellow PhD students within the FoHS, especially my learning set members Colin Ayer, Rose Peacock and Fiona Meddings for their moral support, exchanging ideas, and for the stimulating discussions.

I would like to thank the University of Bradford for giving me a studentship for three years to complete this work. I am also grateful and appreciative that I was given unique opportunities by both faculties (FoHS and FoEC) to develop my technical and qualitative skills and helping me to be a researcher and produce this original piece of work. Furthermore, both the FoHS and School of Dementia Studies have given me opportunities to attend a number of national conferences and meetings during my study (appendix 13), which gave me chances to communicate with others to learn and share my study ideas, outcomes, and methodology.

I wish to express my sincere thanks to all study participants (DCM mappers) who had been very cooperative in expressing their views about their current and potential uses of DCM data.

Last and therefore the integral part of my life, my husband, children (Muzna, Maha and Sami) and my mum Nasreen, sister Naila, brothers Rizwan and Rehan have helped me stay sane through these difficult years. Their constant support, encouragement, prayers and care helped me stay focussed to complete this huge piece of work. I deeply appreciate their belief in me and I greatly value their love, support and friendship.

I dedicate this thesis to the two important men in my life, my husband Burhan and my late father Khalid Shami. Burhan, you have been a constant source of love, care, concern, support and strength all these years and I want to assure you that none of this would have been possible without your love, encouragement and especially patience. My papa witnessed the start of this study, listened to my issues, encouraged me to carry on and prayed for me every single day. He is not here today to see it complete but I know today he would have been very proud of me.

Contents

Abstract.....	i
Acknowledgements.....	iii
Contents	vi
List of Figures	xiii
List of Tables.....	xiv
Glossary of terms and definitions used within this thesis (in alphabetical order)	xv
1. Background and Introduction; Setting the Scene	1
1.1. Introduction	1
1.2. Information technology and health-related data management.....	1
1.3. Dementia	3
1.4. The Dementia Care Mapping (DCM) tool and process	6
1.5. DCM data.....	10
1.6. Defining primary and secondary use of DCM data	16
1.7. Primary and secondary use of DCM data	18
1.8. DCM data-management framework; a data warehouse approach..	24
1.9. Aims of the study	29
1.10. Main contributions of the study	29
1.11. Structure of the thesis	31
2. Secondary Uses of DCM Data	38
2.1. Introduction	38
2.2. Secondary use of DCM data for research purposes	38
2.3. Potential secondary uses of DCM data for benchmarking	40
2.3.1. Data suitability	44
2.3.2. Data comparability.....	50

2.3.3.	Data availability and quality	57
2.4.	Perceptions and approaches of benchmarking and DCM data	62
2.5.	Secondary use of DCM data and need for an effective IT solution .	69
2.6.	Summary of the chapter.....	72
3.	Data Warehousing within Healthcare: Benefits and Challenges.....	74
3.1.	Introduction	74
3.2.	General concept, process and structure of data warehousing	75
3.3.	A data warehouse repository versus traditional database.....	77
3.4.	Healthcare data management and data warehousing	80
3.4.1.	Historic value of healthcare data	86
3.4.2.	Data linkage and integration.....	88
3.4.3.	Analytics on healthcare data	91
3.5.	DCM data management framework: a data warehouse approach..	95
3.5.1.	Data-access component.....	96
3.5.2.	Data extraction, transformation and storage component.....	97
3.5.3.	Data dissemination component	109
3.6.	Data quality and security challenges	112
3.6.1.	Data quality	114
3.6.2.	Data quality requirements in data warehouses	126
3.6.3.	Data security requirements in data warehouses	142
3.6.3.1.	Ethical requirements.....	146
3.6.3.2.	Legal requirements	152
3.6.3.3.	Social requirements	157
3.7.	Summary of the chapter.....	158
4.	Requirement Analysis for a Data Warehouse.....	163
4.1.	Introduction	163
4.2.	Requirement analysis	163
4.3.	Types of requirements and terminology.....	166
4.4.	Role of the analyst	169
4.5.	Methods for gathering requirements	173
4.6.	Methods for analysing and presenting requirements	180

4.7.	Requirement analysis for a data warehouse.....	184
4.7.1.	Data-driven approach for requirement analysis.....	184
4.7.2.	User-driven approach for requirement analysis.....	188
4.8.	Significance of a user-driven approach.....	190
4.8.1.	Identifying Information requirements	192
4.8.2.	Recognising social issues	192
4.8.3.	Usability requirements.....	193
4.9.	Challenges of a user-driven approach	194
4.9.1.	Identifying the right users for the right requirements	194
4.9.2.	Users' understanding and knowledge of the new system.....	198
4.9.3.	Communication Issues between the analyst and the users....	203
4.9.4.	User-requirements expression	203
4.10.	User-driven approach within a data warehouse.....	205
4.11.	Summary of the chapter.....	216
4.12.	Aim of the study	221
4.13.	Objectives of the study.....	221
5.	Philosophical and Methodological Considerations	224
5.1.	Introduction	224
5.2.	Research philosophy	226
5.3.	Qualitative research approach	233
5.4.	Grounded theory techniques as a methodological framework	235
5.5.	Grounded theory	237
5.6.	Grounded theory process	243
5.7.	Developments or modifications within grounded theory.....	249
5.8.	Substantive area	259
5.9.	Use of research question.....	263
5.10.	Use of the literature within grounded theory	265
5.11.	The role of the researcher.....	269
5.12.	Philosophical and methodological similarities between grounded theory and requirement analysis.....	271

5.13. Summary of the chapter.....	275
6. Using Modified Grounded Theory; Data Collection and Analysis	278
6.1. Introduction	278
6.2. Sampling; recognising and identifying participants	279
6.2.1. Recruiting	286
6.2.2. Number of participants	288
6.3. Ethical approval	289
6.4. Data-collection methods	291
6.5. The development of the first interview guide.....	293
6.6. Pilot interviews and development of interviewing skills	296
6.7. Data-collection process.....	298
6.8. Audio-recording and interview-transcription process	304
6.9. Interview data-management process	306
6.10. Data analysis	310
6.10.1. Preliminary analysis of interview data	310
6.10.2. Formal analysis of interview data	313
6.10.2.1. Open coding; exploring data.....	315
6.10.2.2. Axial coding; linking categories.....	325
6.10.2.3. Selective Coding; finding a central category.....	328
6.11. Evaluating qualitative studies.....	333
6.11.1. Credibility	333
6.11.2. Dependability	338
6.11.3. Transferability.....	340
6.11.4. Confirmability	342
6.12. Summary of the chapter.....	345
7. Potential Secondary Uses of DCM Data; Users' Views	347
7.1. Introduction	347
7.2. Benchmarking.....	350
7.3. DCM teaching, training and support	360
7.4. Secondary research.....	366

7.4.1.	DCM data resource with pre-collected data	368
7.4.2.	Exploring DCM data	369
7.4.3.	Access to unpublished DCM data	379
7.5.	Key findings and contribution to knowledge.....	383
7.6.	Summary of the chapter.....	388
8.	Users' Information Requirements	391
8.1.	Introduction	391
8.2.	Data content information requirements.....	393
8.2.1.	DCM data (coding and textual).....	393
8.2.2.	Contextual data	404
8.3.	Metadata information requirements	417
8.3.1.	Provenance information	419
8.3.2.	Keyword information.....	430
8.3.3.	Ethical information.....	436
8.4.	Key findings and contribution to knowledge.....	452
8.5.	Summary of the chapter.....	453
9.	Factors Affecting the Availability and Quality of DCM Data	455
9.1.	Introduction	455
9.2.	Mapper's role	456
9.2.1.	Mental and physical presence.....	456
9.2.2.	Consulting the DCM Manual	458
9.2.3.	Mapper's Inter-rater reliability	460
9.2.4.	Mapper's training and experience	462
9.2.5.	Mapper's link with care setting	466
9.3.	Organisation's role.....	475
9.3.1.	Organisational/management support in implementing DCM ..	475
9.3.2.	Organisational support for mappers	482
9.3.2.1.	Further training.....	482
9.3.2.2.	Peer support.....	484
9.3.2.3.	Regular mapping	485
9.3.2.4.	Help from experts	485
9.3.2.5.	Mapping in pairs	486

9.4. Primary DCM data management system; limitations and requirements.....	488
9.5. Key findings and contribution to knowledge.....	507
9.6. Summary of the chapter.....	511
10. Future Work, Limitations and Conclusions.....	513
10.1. Introduction.....	513
10.2. Summary of the findings based on the study objectives.....	513
10.3. User-driven approach vs. data-driven approach.....	516
10.4. User-driven approach; reflections on the use of modified grounded theory.....	520
10.5. Recommended future work.....	523
10.5.1. Identifying, collecting and linking additional data.....	523
10.5.2. Data quality.....	525
10.5.3. Data security.....	527
10.5.4. System usability.....	531
10.5.5. DCM data-management systems for primary purposes.....	531
10.5.6. Development of the methodology.....	531
10.6. Limitations of the study.....	532
10.7. Conclusions.....	535
References.....	541
Appendix 1: Consent form (general).....	578
Appendix 2: Consent to take part in a focus group.....	580
Appendix 3: Information sheet for health and social care organisations.....	583
Appendix 4: Information sheet for DCM researchers.....	588
Appendix 5: Study participants' details.....	593
Appendix 6: General interview guide for all study participants.....	595
Appendix 7: Semi-structured interview guide (for researchers).....	596
Appendix 8: Semi-structured interview guide (for practitioners and trainers)	598

Appendix 9: Individual interview summary memo	600
Appendix 10: Memo example	601
Appendix 11: A list of some codes.....	602
Appendix 12: A list of some codes and categories	603
Appendix 13: A flow diagram showing the sequence and flow of emergence of category 'provenance data'	606
Appendix 14: A model for describing category 'contextual data'	607
Appendix 15: Publications and conference papers.....	608

List of Figures

Figure 1: A two-step process for managing DCM data for secondary use ...	26
Figure 2: A general structure of a data warehouse	77
Figure 3: A two-step approach for warehousing Dementia Care Mapping (DCM) data (taken from Khalid 2010: 66)	95
Figure 4: A fact table (data arranged at highest granularity level) and various dimension tables (taken from Khalid 2010: 73)	106
Figure 5: A fact table (data arranged at lowest granularity level) and various dimension tables taken from Khalid (2010: 79)	107
Figure 6: Research process stages	226
Figure 7: Phases of data-collection and analysis process	311
Figure 8: A coding framework.	314
Figure 9: The components of the 'paradigmatic model' and a main concept	327
Figure 10: Six high-level categories	328
Figure 11: Central category: availability and quality of DCM data for secondary uses.....	331
Figure 12: Central category: users' information requirements	331
Figure 13: Category 'potential secondary uses of DCM data' and its sub-categories	349
Figure 14: Category 'DCM data warehouse Information requirements and its sub-categories	392
Figure 15: Category 'factors affecting the availability and quality of DCM data' and its sub-categories.....	456

List of Tables

Table 1: A list of Behaviour Category Codes (taken from DCM 8 Users' Manual (2005: 17).....	10
Table 2: A list of Mood and Engagement Values (taken from DCM 8 Users' Manual (2005: 13).....	12
Table 3: An example of open coding process.	318

Glossary of terms and definitions used within this thesis (in alphabetical order)

Axial coding

The second stage of the coding process in grounded theory, which involves associating categories and sub-categories through their properties and dimensions.

Aggregated data

Data that has been summarised

Code

Name or label given to a piece of text during the coding process

Concept

A building block of the theory - a basic unit of the theoretical framework.

Category

Merged codes based on a similar concept

Core-category

The central category formed through linking the developed categories and sub-categories.

Data mining

Data mining is a process of extracting valuable data patterns and knowledge from a large amount of data.

Data model

A data model is diagrammatic representation of entities (e.g. mapping participant, mapper, mapping session) and their relationships.

Data warehousing

A process of integrating data from various sources into a single repository, called a data warehouse, for its use for various secondary purposes.

Data warehouse

A central data repository that contains integrated and historical data taken from various data sources arranged in an easily accessible format.

Database

A single source data repository where data is managed (arranged) for storage and easy retrieval.

Data-driven approach

An approach involved extracting information (requirements) from the current and existing working systems and documents as requirements for designing the data warehouse.

DCM data management framework

A diagrammatic representation of a proposed way of managing DCM data within various data repositories by various users.

Dementia Care Mapping (DCM)

An observational tool and practice development process, used to assess and improve the quality of care of people with dementia in formal dementia care settings, such as care homes, day centres and hospital wards.

Extraction Transformation Loading (ETL)

The process of extracting data from various sources, transforming it into a consistent and user acceptable format and loading it into the data warehouse.

Grounded theory

A research methodology, which supports the development of theory or an integrated framework of concepts that can be used to explain or predict a phenomenon.

Historic data

Data collected over time

Information system

A computer-based system that involves storing and processing data into a required format in order to provide specific information for the system users.

Integrated data

Data that is collected from various sources.

Mapper

An individual trained to use the DCM tool and process.

Mapping

The process of observation within DCM for a specific time period.

Multidimensional analysis

Looking at data from various angles (e.g. time, location).

Multidimensional data

A set of data that provides the view of the situation from various angles (directions).

OLAP (online analytical application)

An on-line computer-based application used to query or retrieve data held in a data warehouse.

Open coding

The first stage within the grounded theory coding process, which involves looking at raw data and giving names (codes) to the pieces of text based on some underlying concept.

Requirements

Pieces of information that are obtained from various sources in order to understand the features/functionality of the new system.

Requirement analysis

The process of obtaining, understanding and amalgamating requirements to develop mutually required features/ functionality within the system.

Secondary use of DCM data

The reuse of DCM data, collected from various individual mappers and organisations, for purposes different from those for which it was originally collected.

Substantive area

The area where the identified theory can be used or implemented or an area where the identified findings are applicable.

User

The potential users of the DCM data warehouse who are using DCM data currently or have expectations of using it in the future

User-driven approach

An approach for designing a data warehouse, which involves understanding the needs/features/functionality of the system through the eyes of the people who will potentially be using it (users) in order to identify requirements.

1. Background and Introduction; Setting the Scene

1.1. Introduction

The aim of this chapter is to provide a detailed background and introduction to various aspects of this multidisciplinary study. The chapter begins with briefly introducing the role of information technology in healthcare data management. It then describes the concept of dementia and further explains Dementia Care Mapping (DCM) as a method for improving the quality of care in formal dementia care settings. The chapter then goes on explaining the DCM data and defining its primary and secondary uses before highlighting the need of a technical solution such as a data warehouse for managing data for secondary uses.

The chapter continues by briefly explaining the previous work, highlighting the gaps in the field and identifying the need of further work in terms of bringing user views and perceptions for designing a data warehouse for DCM. The chapter ends by enlisting the main contributions to the study and providing summaries of the remaining chapters of the thesis.

1.2. Information technology and health-related data management

Information technology (IT) can provide efficient, integrated and reliable methods to collect, maintain and transfer patient health information in a secure manner (Hoppszallern 2012; Department of Health 2013a; National Information Board 2014). The contributions of IT to the development of health-related data and systems include improvements in the quality, efficiency and security of data, which in turn support better decision-making within healthcare services, research and policy (Detmer 2000; Wisniewski et

al. 2003; Department of Health 2013b). Health-related data can be used for direct clinical use (primary use), i.e. individual patient care, and for non-direct clinical use (secondary use), i.e. research or decision-making purposes (Safran et al. 2007; Innovative Medicines Initiatives 2014). This is the reason why, globally, serious initiatives have been taken to store health-related data in electronic formats in order to make it digitally connected, fluid, interoperable and accessible for primary and secondary use purposes (Health Care Reform 2010; Health and Social Care Information Centre 2015a). These purposes involve enhancing the wellbeing of patients through the provision of good, cost-effective and informed care (Health Canada 2001; European Commission 2004; Neupert 2009; Health Information and Quality Authority 2012).

Information technology's contribution to managing health-related data effectively and innovatively, to facilitate its secondary use, is becoming a priority for many governments. Information systems are being introduced in the health sector for the following reasons: the integration of diverse data to support analysis; the facilitation of electronic data transfer and sharing across sectors and organisations; and the improvement of quality, efficiency, safety, security and collaboration through research and decision-making (Raghupati and Tan 2002; Hoppszallern 2012). The UK government has supported initiatives for using IT-based solutions to manage health-related data for secondary purposes, such as research (Medical Research Council 2011). Major UK national programmes include the Research Capability Programme (RCP) (National Institute of Health Research 2006) in England, Secure Anonymised Information Linkage (SAIL) (Ford et al. 2009;

Administrative Data Liaison Service 2012) in Wales and the Scottish Health Information System (SHIP 2012). These programmes are responsible for building infrastructure, supporting the linkage of National Health Service (NHS) health data with non-NHS health data to enable its secondary use for research, and improving public safety and healthcare.

It is also emphasised by the International Medical Informatics Association (IMIA) (2012) that using health-related data for secondary uses can provide enormous benefits for all types of clinical and health services and for social and public-health research. Accumulated and aggregated health data provide value for a broad range of research, quality, public health and commercial applications, for example Online Analytical Processing (OLAP) (Ponniah 2001) and Decision Support Systems (DSS) (Sartipi et al. 2007). OLAP and DSS applications use aggregated or summary data to process it further for secondary uses, for example performing online queries on the data for complex analysis and processing data for making decisions based on trends and patterns.

Despite the global and national efforts in introducing IT to transform healthcare by utilising health-related data for secondary uses, there remain areas where its contributions are minimal, specifically in managing dementia care related data electronically, including implementing innovative systems or methods for facilitating its secondary use (Khalid 2009).

1.3. Dementia

Dementia is defined as “a decline in mental ability which affects memory, thinking, problem solving, concentration, communication and perception”

(Mental Health Foundation 2012). It is an umbrella term for a range of disease processes, including Alzheimer's disease, vascular dementia and dementia with Lewy bodies that cause damage to the brain cells, all processes, which are likely to affect a person's communication skills. Age is the main factor in developing dementia, with the majority of those with the condition being over the age of 65; however, younger people can be affected as well (Department of Health 2009). Currently, there are 46.8 million people estimated to be living with dementia worldwide (Prince et al. 2015). An increase of more than 10 million since 2010. The numbers of people with dementia globally are estimated to be 74.7 million in 2030 and more than 131.5 million in 2050 (Prince et al. 2015). There are 9.9 million new cases of dementia every year, indicating one new case every four seconds, which is 30% higher than the annual number of new cases estimated for 2010 by the World Health Organisation in their 2012 report (Prince et al. 2015). Currently about 800,000 people with dementia live in the UK and the number is expected to double in the next 30 years (Department of Health 2015a).

People with dementia are high users of health and social care services. In England alone, there are approximately 670,000 people with dementia, of which one third live in residential care settings, while two thirds of care-home residents are currently estimated to have dementia (Department of Health 2013b). Further, an estimated 25% of acute hospital beds in England are occupied by people with dementia (Alzheimer's Society 2009). According to the Department of Health (2013b), more than £19 billion is spent each year on dementia care within formal dementia-care settings. There are also indirect costs as more family carers provide care and support for people with dementia on a daily basis.

With the aging population bringing about an increasing prevalence of dementia and associated economic, social, health and personal costs, dementia has been made an international health priority (Knapp et al. 2007; World Health Organization and Alzheimer's Disease International 2012; Prince et al. 2015) and a national priority in the UK (Knapp et al. 2007; Department of Health 2012; Department of Health 2015b). Governments, health and social-care providers, dementia-specific organisations and charities therefore, promote not just medically driven research and interventions for the treatment and cure of dementia, but also social and psychological support for improving the quality of life and the quality of care for those living with dementia.

Assessing and improving the quality of formal dementia care is not a simple or short-term initiative. The needs of people with dementia are varied and often

highly complex. As dementia progresses, people with the condition become more dependent and often require 24-hour care and support from multidisciplinary teams, including GPs, nurses and care staff, through services providing assessment, treatment, outreach, respite and social care. Therefore, providing good-quality care requires skilled and trained paid staff, a high-quality care setting and coordinated services (Knapp et al. 2007). In a 2012 report, the Alzheimer's Society highlighted the fact that 'unacceptable variations' were being seen in the quality of care provided to people with dementia across all formal care settings (Alzheimer's Society 2012), a challenge also identified by the Department of Health (2013b). The reasons for this include: the complex needs of people with dementia; the need for trained and skilled staff to understand the needs and aspirations of people with dementia, maintaining their dignity and self-respect; and ineffective coordination between professionals and services that provide formal dementia care. Many people with dementia who are being cared for, or who are living in formal care settings may have difficulties communicating their needs, or experiences of care and therefore may have limited or no voice in speaking out for improved care quality. Since emphasis has been given nationally to improving the quality of care and the quality of life of people with dementia in formal dementia care settings (Department of Health 2013b), there is a need for a systematic approach to facilitate this.

1.4. The Dementia Care Mapping (DCM) tool and process

Based on Kitwood's (1997) person-centred philosophy that dementia care should focus on improving a person's individual and social wellbeing. Dementia Care Mapping (DCM) (Bradford Dementia Group 1997; Bradford

Dementia Group 2005) was devised as an observational tool and process to assess and improve the quality of care of people with dementia in formal dementia care settings such as care homes, day centres and hospital wards (Bradford Dementia Group 2005). DCM is recommended to assess the quality of life of people with dementia by the National Audit Office (2007) and NICE/SCIE (2007).

DCM is used both as a tool within research studies as an outcome measure and a practice development process for assessing and improving dementia care quality within formal dementia care settings (Brooker and Surr 2005; Brooker and Surr 2010). As a practice development process, DCM is conducted in a cycle of five phases: briefing the staff; observing people with dementia (mapping participants); capturing and analysing the information (DCM data); further reporting and feeding back the findings to staff; and making action plans and setting targets for improving the care (Brooker 2005). The DCM guidelines suggest that the five-phase cycle is repeated after every three to six months to assess the existing levels of care and set new targets for the quality of life and the quality of care provided for people with dementia (Bradford Dementia Group 2005).

Within a research context, DCM has been used as an outcome measure for assessing behavioural patterns, levels of wellbeing, the quality of interactions with staff and quality of life of people with dementia (Brooker 2005). It has been used to assess the efficacy of staff training, care quality and culture change projects (Lintern et al. 2002) as well as the efficacy of a range of interventions, including aromatherapy (Ballard et al. 2002), intergenerational

activity programs (Jarrott and Bruno 2003), horticultural therapy (Gigliotti et al. 2004)

and reminiscence programs (Brooker and Duce 2000). Where DCM is used as a research outcome measure the cyclic practice development process is not usually followed, just the mapping and data analysis stages are completed. The analysis of data for research purposes is also usually different in nature to that used for practice development purposes. In research, the DCM data is likely to be collected alongside a range of other data related to the specific research questions and outcomes being investigated.

Only trained individuals can use the DCM tool. Every year individuals from many countries, usually from health and social-care and research backgrounds, are trained. Trained individuals are called mappers and the observations carried out using the DCM tool are called mapping. The mapping is usually conducted by one or more mappers, depending on the number of participants being observed, with each mapper typically continuously observing five to eight participants for a specific time period (Bradford Dementia Group 2005). The length of mapping is variable, depending on the mapping purpose, ranging from short maps of 30 minutes to longer maps of up to six-hours. During observations, DCM data is recorded as codes that reflect participants' behaviour, mood, engagement levels and type of interaction with staff in every five-minute period (time-frame). These coding frames will be explained in detail in the next section.

1.5. DCM data

DCM provides a means of gathering data, which constitute both quantitative and qualitative types of information. The quantitative information is presented in four types of coding frames: Behaviour Category Codes (BCC); Mood and Engagement (ME) values; Personal Detractions (PDs); and Personal Enhancers (PEs). Behaviour Category Codes (BCC) (see Table1) are described as one of 23 different domains that represent a range of mapping participants' behaviour recorded as letters A to Y (except H, M), i.e. A for articulation (when the participant is engaged with other living things, either human or animal, using verbal or non-verbal communication) (Bradford Dementia Group 2005). There is a 'Z', which is used for behaviours that fit no other category.

Table 1: A list of Behaviour Category Codes (taken from DCM 8 Users' Manual (2005: 17).

Code	Memory Cue	General Description of Category
A	Articulation	Interacting with others verbally or otherwise - with no obvious accompanying activity.
B	Borderline	Being engaged but passively.
C	Cool	Being disengaged, withdrawn.
D	Doing for self	Self-care.
E	Expressive	Expressive or creative activity.
F	Food	Eating or drinking.
G	Going back	Reminiscence and life review.
I	Intellectual	Prioritising the use of intellectual activities.
J	Joints	Exercise or physical sport.
K	Kum and go	Walking, standing or moving independently.
L	Leisure	Leisure, fun and recreational activities.
N	Nod, Land of	Sleeping, dozing.

O	Objects	Displaying attachment to, or relating to, inanimate objects.
P	Physical	Receiving practical, physical or personal care.
R	Religious	Engaging in a religious activity.
S	Sexual expression	Sexual expression.
T	Timalation	Direct engagement of the sense.
U	Unresponded to	Attempting to communicate without receiving a response.
V	Vocational	Work or work-like activity.
W	Withstanding	Repetitive self-stimulation of a sustained nature.
X	X-cretion	Episode related to excretion.
Y	Yourself	Interaction in the absence of any observable other.
Z	Zero option	Fits none of the existing categories.

Within the same five-minute time-frame, the mappers also record the participants' mood and engagement levels, known as a Mood and Engagement (ME) value (see Table 2). The ME values are expressed on a six-point scale ranging from extreme distress (-5) to extreme positive mood and engagement (+5) (Bradford Dementia Group 2005). So, for example, a participant engaging in a positive conversation would be coded as A+3. Over a six-hour map, up to 72 time-frames of data may be coded for each participant.

Table 2: A list of Mood and Engagement Values (taken from DCM 8 Users' Manual (2005: 13).

Mood	ME values	Engagement
Very happy, cheerful. Very high positive mood.	+5	Very absorbed, deeply engrossed/engaged.
Content, happy, relaxed. Considerable positive mood.	+3	Concentrating but distractible. Considerable engagement.
Neutral. Absence of overt signs of positive or negative mood.	+1	Alert and focused on surroundings. Brief or intermittent engagement.
Small signs of negative mood.	-1	Withdrawn and out of contact.
Considerable signs of negative moods.	-3	
Very distressed. Very great signs of negative mood.	-5	

DCM is the only tool that captures not only information about the behaviour and associated mood and engagement of people with dementia, but also the quality of interaction they have with the care staff. Together these play a significant role in indicating the quality of care they receive within care settings (Bradford Dementia Group 2014). The quality of staff interaction with people with dementia is recorded through Personal Enhancers (PEs) and Personal Detractions (PDs) (Bradford Dementia Group 2005). PDs are examples of staff behaviour that have the potential to undermine the personhood of people with dementia (Kitwood 1997) and can have an impact on their overall wellbeing. PEs are associated with interactions between the participant and the care staff that have the potential to enhance their wellbeing or personhood. There are 17 different types of PD and PE that may be coded as and when they occur. During each mapping a large amount

of qualitative notes are written by mappers to give a context to the formal coding frames and additional

information on the environment of the care setting, such as the noise levels, temperature and overall ambiance where the mapping is taking place.

Alongside DCM data, a limited amount of additional information is also collected as part of the mapping process. This includes the date and time of mapping, location of mapping, mapper and participants' name/id, as collected within the DCM data collection sheets (e.g. DCM raw data sheets) (Bradford Dementia Group 2005). A number of research studies, however, report the collection and use of detailed information including: participant characteristics such as age, gender, ethnicity, length of stay in health and social care facilities (Kuhn et al. 2005; Barnes 2013); depression levels (Kuhn et al. 2004); and cognitive state (Kuhn et al. 2004; Kuhn et al. 2005). Information is also collected about care setting characteristics such as type, location (Willemse et al. 2011) and size (Kuhn et al. 2002); and staff-related information such as staff ratios and training (Innes and Surr 2001). This additional information is not routinely collected in a standard mapping process although it may be available from other sources such as electronic patient records.

Following processing, the DCM data help to identify areas for potential improvements in existing practices and future planning of care. The DCM 8 user's Manual (2005) provides some examples of how a mapper can undertake basic processing and calculations of the data, such as the percentage of time spent in each BCC and ME and the total number of PDs and PEs. Further processing involves calculating the average of all ME values over the observation period at an individual and group level (known as the well- or ill-being, or WIB, score). This provides an indicator of the

relative level of well- or ill-being experienced during the map by an individual
or by the group

as a whole. The richness of the DCM data also permits other analyses to be completed for individuals or groups of participants. For example, by selecting specific data items and by combining BCC and ME analysis to produce particular indices such as agitation and distress levels, withdrawn behaviour, passive engagement and opportunities for activity and engagement (Brooker and Surr 2005).

The wider uses of DCM as a tool and process produce large amount of raw data, which, following further processing, can be used for both primary and secondary purposes. In the absence of any clear distinction between what constitutes primary or secondary use of DCM data, the following section will define these terms within the context of this study.

1.6. Defining primary and secondary use of DCM data

The terms 'primary use' and 'secondary use', for DCM data are defined based on how these terms are used in relation to healthcare data. First, I will elaborate on the general meaning of these terms and their definition in relation to healthcare data. Drawing upon these definitions, I will then define the primary and secondary use of DCM data.

The word primary means 'first' and the word secondary means 'second'. The use of these words in relation to data is usually related to how, when and who collected and used (analysed) the data (Boslaugh 2007). For example, when the collector (either an individual or an organisation), who gathered data with a specific purpose in mind, used/analysed the data for the *first* time after its collection, the process is referred to as the primary use of data

(Boslaugh 2007). However, when the same data is used for the *second* time, usually for

a different purpose, and by a user who was not involved in its collection, the process is referred to as the 'reuse or secondary use of data' (Boslaugh 2007).

Based on the above definitions, the British Medical Association (2007: 2) defines the term 'primary use' of healthcare data as when "health professionals primarily collect patient data to provide direct care and treatment to an individual patient or specific patient population". Their definition of secondary use of data is when "patient data is also used for other activities that contribute to health and social care services more generally, such as conducting medical research and managing health services". Using this concept, 'primary use' will be defined as the use of DCM data collected by an individual or organisation for its original, specific purpose. Secondary use, however, will be defined as the reuse of DCM data, collected from various individual mappers and organisations, for purposes different from those for which it was originally collected. Based on these definitions, the next section explores the primary and secondary uses of DCM data.

1.7. Primary and secondary use of DCM data

In both a practice development and research context, the common uses of DCM is reported at a local level, that is, the individual mapper or an organisation collecting the data uses it within the setting or research project, for the purpose it was originally collected, thus referring to its primary uses. These uses are reported for assessing and improving the quality of care provided at an individual, setting and organisational level. The raw DCM data provide individual level information, which is used by care staff to assess,

plan and monitor care via care plans (Packer and Jeffries 1997). This individual level data can be accumulated and analysed, along with other detailed

information about the care setting, to identify factors that may influence the quality of care provided at a setting level. For example, studies show the use of DCM data for setting level care-quality monitoring and improvement and identification of staff training and development needs (Younger and Martin 2000; Lintern et al. 2002). Further, the individuals' DCM data taken from many care settings can be accumulated to identify potential areas for improvements at an organisational level through, for example, care-quality audits (Brooker et al. 1998; Younger and Martin 2000).

The re-analysis of DCM data for secondary purposes can produce new knowledge that can inform future improvements in the DCM method/tool and provide suggestions for providing quality dementia care within formal dementia-care settings. There exist some examples in the literature that demonstrate the secondary use of DCM data for research purposes, such as Fossey et al. (2002), who conducted secondary analysis of DCM data from three earlier studies in order to examine the psychometric properties of DCM as a research outcome measure. Also Innes and Surr (2001) undertook a cohort analysis of data from two separate care home based studies to describe standards of dementia care in care home settings in the UK. A detailed analysis of these studies is provided in Chapter 2. In addition to the secondary use of DCM data for research, the literature also indicates that, if managed effectively, DCM data can be used for benchmarking. A detailed investigation of the potential secondary uses of DCM data for benchmarking is provided in Chapter 2. Despite the widespread primary use of DCM and indications of its potential for secondary use (Innes and Surr 2001; Fossey et

al. 2002; Khalid 2010; Khalid et al. 2010) there has, to date, been a limited effort to develop a

technical solution that can facilitate the reuse of large amounts of complex and information-rich DCM and related datasets. This constitutes a major barrier to secondary use of DCM data.

Managing DCM data to make it available for reuse through using effective technological methods/solutions, is in line with the UK's national policy of making health-related data open, transparent and available for secondary uses (Health and Social Care Information Centre 2012) and using technology and data for good health and social-care provision (National Information Board 2014). Further, specifically in relation to dementia, this is also in line with the Department of Health's call for collecting and using datasets that can provide an understanding of dementia care provision at local and national level for sharing experiences and good practice among care providers (Department of Health 2013a).

Moreover, dementia care related data made available for secondary use could potentially form part of current dementia research initiatives. In 2013, a global dementia summit brought many countries together to pledge to improve dementia care (Department of Health 2013c). For this purpose, the major action plan was to commence initiatives for integrating data to enhance the opportunities for dementia research (Department of Health 2013c). In response to this, a recent report published by the Organisation for Economic Co-operation and Development (OECD) (Deetjen et al. 2015) proposes a big-data solution for dementia research. It defines big data as both medically driven and non-medically driven data for dementia research. The term 'non-medically driven data' here refers to data about people's

lifestyles, diets and food choices. The OECD report authors assert that the medical and non

-medical dimensions collectively could support improvements to the care of people with dementia (Deetjen et al. 2015). While the current focus of the big-data proposal is to enhance research to support better dementia diagnoses, as well as cures and treatments, dementia care related data could also provide a potential resource for understanding, comparing and assessing the existing care provision for people with dementia. However, to facilitate this dementia care related data needs to be available in electronic and integrated formats in order to be considered part of such big-data initiatives. Currently there remain few mechanisms for achieving this, and none is currently available for storage of DCM data. Therefore, the need for information systems, such as a data warehouse, is required to integrate data from various dimensions, thereby providing opportunities for potential secondary use for in-depth analysis, decision-making and research (Post et al. 2013).

1.8. DCM data-management framework; a data warehouse approach

In order to maximise the potential of data being generated through DCM by using it for potential secondary uses, a sustainable and consistent data management solution is required. In a previous study, I proposed a data management framework using a data warehouse approach for managing the secondary use of DCM data (Khalid 2010). A detailed and critical review of this study is provided in Chapter 3. A data warehouse is a type of information system which provides an electronic repository that stores and links data taken from various sources and enables its retrieval for secondary use (Stolba 2007). The detail about how a data warehouse works is also presented in Chapter 3. My 2010 study was the first, and is currently the

only, technical solution proposed for managing DCM data for secondary purposes. It adopts a two

-step conceptual approach (Figure 1). A rationale for why a data warehouse and two-step approach was proposed is discussed in Chapter 3.

The first step involves storing DCM data taken from national and international mappers within a web-based data repository. This data repository is called the DCM international database; its purpose is to enable national and international mappers to input their DCM data into an online database system, which will store their own collected mapping data over time and also generate basic reports based on completion of analysis (Khalid 2009).



Figure 1: A two-step process for managing DCM data for secondary use

The second step involves processing DCM data into a data warehouse for the purpose of long-term storage and for reusing the data for secondary purposes, for example: complex analysis and reporting on the DCM data; secondary research; benchmarking the quality of care; and data-mining for identifying trends and patterns of good dementia care (Khalid 2010). While my previous study initiated the important and novel work for proposing a solution for managing DCM data for secondary uses, it only went so far as assessing and proposing a technical architecture for the data management needs of DCM data. The DCM data warehouse still needs to be designed, developed and implemented through what is known as a development life-cycle (Thakur and Gosain 2011).

Within the development life cycle, designing a data warehouse is a fundamental and important step towards its successful use and acceptance by users (Browning and Mundy 2001; List et al. 2002; Schaefer et al. 2011). The design process of a data warehouse involves producing a data model, which is a structural representation of data. A data model is designed based on specified requirements. The requirements refer to information obtained from various sources such as existing systems, documents or potential users of the system, which illustrate 'what the system can do' and 'how it can be done'. Understanding requirements within the design process is referred to as requirement analysis, which is a process of obtaining, synthesising and analysing the requirements into an explanation that can support the design and development of a workable and acceptable system (Abai et al. 2013). A detailed view of what constitutes requirements and how these are gathered for designing and developing a data warehouse is presented in Chapter 4.

In order to demonstrate data management across a two-step framework, within my previous study (Khalid 2010), an attempt to design a data model for a DCM data warehouse was made. It was a technical effort where the existing system of Excel programme¹ and the design of the DCM international database (Khalid et al. 2010) were analysed to obtain the requirements for designing a future DCM data warehouse. This technically focussed approach of gathering requirements for designing a data warehouse is called a data-driven approach, and is one of the two main approaches to designing a data warehouse. These two approaches are critiqued in Chapter 4. Based on these

¹ Excel programme is provided by the BDG to support mappers' basic analysis of some of the DCM data such as BCC and ME.

requirements, a basic data warehouse design (e.g. a data model) was proposed, which was validated by using simulated DCM data. The main aim was to assess the DCM data management framework and validate the data flow from the DCM international database to the DCM data warehouse and its uses for data-mining (Khalid 2010). This study therefore made the first successful attempt at showing the technical workability of a DCM data warehousing approach. However, the study was limited in that it did not involve potential users in the design process for example through gathering their requirements for designing the DCM data warehouse.

As will be explored in further detail in Chapter 5, data warehouses are information systems that are represented as a combination of people, technology and organisations (Iivari and Hirschheim 1996). These three aspects influence the data warehouse design and the requirement analysis process immensely. Users play an important role in identifying the requirements that inform the type and structure of the data (data model) that goes into the warehouse, the processing of the data, which provides valuable information, and the retrieval of that information for specific purposes (Lindgaard et al. 2006). Therefore, a growing body of literature (Kappleman 1994; Raab 1998; Teixeira et al. 2012) suggests and emphasises the involvement of users for identifying the system requirements, which could be designed and developed further. This approach is called a user-driven approach for designing a data warehouse. A detailed rationale of using a user-driven approach for a DCM data warehouse is provided in Chapter 4.

Further, as will be discussed in detail in Chapter 4, the literature suggests using data-driven and user-driven approaches together to provide a broader

and more detailed picture of requirements for a data warehouse (Golfarelli 2010; List et al. 2002). While my previous study (Khalid 2010) used a data-driven approach for designing data models for a DCM data warehouse, it was limited in that it did not seek to include potential, future DCM data warehouse users' views regarding their requirements for the type of data they need storing within the data warehouse that would facilitate their potential secondary use of the DCM data. Such information is vital in designing a data model for a DCM data warehouse and it is important to consider, while developing and implementing the system, so that is not just technically operative but user-accepted as well (Schaefer et al. 2011).

1.9. Aims of the study

This study aimed to explore requirements for the secondary use of DCM data within a data warehouse using a user-driven approach. While this study began with the broader aim, study objectives were set following a literature review and presented at the end of Chapter 4.

1.10. Main contributions of the study

While there is a substantial amount of literature available to demonstrate effective primary use of DCM data, there is very little known within the field regarding its secondary use and the relevant concerns and issues about this from the potential users' perspectives. This study introduces the concept of and motivations for the secondary use of DCM data and provides original contributions to knowledge by identifying user-identified requirements that are significant regarding the secondary use of DCM data. Further, there is also limited evidence in the literature to demonstrate how users' views can

be translated into specific requirements that enhance the existing technical data

model of a DCM data warehouse. To fill these gaps in the literature and to build on my previous study (Khalid 2010), this multidisciplinary study has identified user views and perceptions regarding their potential secondary use of DCM data and interpreted these as requirements for a future data warehouse. As a result, this study identifies user driven requirements, which provide a crucial, different perspective from a data-driven approach.

In addition to this, the study has also contributed in identifying the prospective users for a future data warehouse based on their perceived potential uses of DCM data. This shows various types of users' intention to utilise the system in the future and provides motivations for designing and developing a data warehouse for DCM.

In the absence of any existing knowledge of users' views on the secondary use of DCM data, this study employed an exploratory methodology in the form of a modified grounded theory. The details of the rationale and use of this methodology is provided in chapters 5 and 6. From a methodological perspective, this study has contributed by demonstrating the use of modified grounded theory for identifying views and perceptions of the user group which has limited technical knowledge and further interpreting them into a set of data warehouse requirements.

1.11. Structure of the thesis

The rest of the thesis is organised into the following chapters.

Chapters 2, 3 and 4 provide background literature and together set the objectives of this study, which are presented at the end of Chapter 4. Next, chapters 5 and 6 provide a detailed view of the chosen methodology and its

application within this study. Chapter 7, 8 and 9 then present the study findings according to the objectives of the study and discuss these within the context of the relevant literature. Finally, Chapter 10 summarises the key findings, underlines the future research and practical work for a future DCM data warehouse and concludes the study. Summaries of all chapters are presented next.

Chapter 2: Secondary uses of DCM data

This is one of the three chapters which provides background literature. This chapter explores the potential secondary uses of DCM data and argues the potentiality of DCM data for secondary uses and the need for an effective IT based system to support this. This provides a context for Chapter 3, where a data warehouse concept is discussed as an effective solution for managing the secondary uses of healthcare data.

Chapter 3: Data warehousing in healthcare: benefits and challenges

This chapter also discusses relevant background literature, focussing on data warehousing within healthcare as an effective data management solution for secondary use of data. This chapter introduces the concept of warehousing and its general process and structure. It further illuminates the growing use of data warehouses within healthcare for managing data that is collected from various sources for secondary purposes. Further, the benefits of warehousing healthcare data are explored in detail. This leads to an argument in favour of the need to warehouse DCM data. A detailed critique of my previous work is provided next, where I proposed a data management

framework constituting a two-step warehousing approach for DCM data. This chapter then discusses

the limitations of my previous study with an emphasis on involving user views for a data warehouse. Further, data quality and data security are discussed as the most reported challenges to warehousing healthcare data. The main argument of this chapter is that, while my previous study initiated novel work designing a DCM data warehouse using a data-driven, technical approach, there is still a need to involve potential users to explore their requirements for designing and developing a user-acceptable system.

This argument leads to the next Chapter, where user involvement in the context of designing a data warehouse is discussed in detail.

Chapter 4: Requirement analysis for a data warehouse

This is the last of three chapters critiquing the background literature to the study. This chapter focuses on providing a rationale for a user-driven approach for gathering requirements for a DCM data warehouse and argues the need for an appropriate methodology for this purpose. This follows the next chapter, which discusses the chosen methodology for requirement analysis for a DCM data warehouse.

Chapter 5: Philosophical and methodological considerations

User-driven approaches for requirement analysis are underpinned by specific philosophical assumptions. This chapter presents the philosophical considerations arising from this study, which harmonise with the user-driven approach to requirement analysis. Further, it provides a detailed explanation and a critical rationale for using a modified grounded theory methodology for this study.

The theoretical explanation of the methodology provides the underpinnings for Chapter 6 where it is applied to the specific data collection and analysis procedures used in this study.

Chapter 6: Using modified grounded theory: data collection and analysis

This chapter provides a rationale for the research methodology, data-collection and analysis methods used. It also presents and discusses the rationale and techniques used for participant sampling, recruitment and study data management as well as ethical issues associated with this research. A detailed explanation of data collection and analysis is also provided. This chapter ends by providing a detailed explanation of the criteria used to evaluate the quality of data collection, analysis and findings of this qualitative study. The outcome of the chosen methods is presented as findings of this study in the next chapters.

Chapters 7, 8 and 9: Findings and discussion

These chapters present and discuss the findings according to the objectives of the study. Chapter 7 explores users' views and perception of their potential secondary uses of DCM data, arguing that there could be three potential uses of a DCM data warehouse. Chapter 8 then presents and analyses the users' information requirements for the use of DCM data for research purposes. Chapter 9 presents and discusses some of the main issues extrapolated from the study that can potentially influence the secondary use of DCM data.

Chapters 7, 8 and 9 also locate the study findings within the existing literature for explanation and verification purposes and discusses their implications for

a DCM data warehouse for further work. Each chapter also highlights the main contributions to knowledge this thesis makes.

Chapter 10: Future work, limitations and conclusions

This chapter summarises the main study findings and reflects on the use of a user-driven approach. It further makes recommendations for future research and technical work associated with the design of a DCM data warehouse, presents limitations of the study and its conclusions.

2. Secondary Uses of DCM Data

2.1. Introduction

The aim of this chapter is to investigate the existing and potential secondary uses of DCM data. The chapter begins by providing examples of empirical work that demonstrates current secondary use of DCM data within the context of research. It continues by further exploring benchmarking as a potential secondary use of DCM data and the practicality of DCM data for this purpose. Finally, the chapter summarises the argument that, while DCM data can potentially be used for secondary purposes, it requires effective IT-based solutions for data management, for which it is important to obtain the user requirements.

2.2. Secondary use of DCM data for research purposes

There is limited evidence regarding the secondary use of DCM data and of literature reporting the re-analysis of DCM data for secondary purposes. In this section, I will critique the limited number of such studies, which demonstrate the potential secondary use of DCM data for research purposes.

Fossey et al. (2002) collected DCM data from three earlier studies. This included: Cohort A - 123 participants of a longitudinal nursing-home intervention study; Cohort B - 24 continuing-care participants in the placebo group of an aromatherapy intervention study; and 30 participants from a randomised neuroleptic withdrawal study within the same cohort group. Fossey and colleagues re-analysed the data to examine the psychometric properties of DCM as a research outcome measure. They re-analysed the

DCM data from the three research studies to examine the association between

various DCM data items and demonstrated that DCM has good internal consistency. Internal consistency refers to the extent to which different data items of a tool/method, measure the same general construct of the tool (such as activities, WIB score and social withdrawal within DCM). This study suggests that previously collected DCM data can potentially be re-analysed to extract useful knowledge about the DCM tool, such as providing evidence of its efficacy as a research outcome measure.

While Fossey and colleagues' study (2002) demonstrate the potential of DCM data for secondary use (e.g. research purposes), it is not known whether the authors did experience any concerns or issues related to data quality. Further, the study is limited in terms of reporting the consent- and privacy issues while reusing the patient-related data. Similarly, Innes and Surr's (2001) study, where data was used taken from two separate care home based studies to describe standards of dementia care in care home settings in the UK, does not explain any details that clarify the quality and ethical status of the DCM data for secondary uses. The literature reports that data quality and security issues are two major challenges in the reuse of healthcare data. A detailed review of these challenges is provided in Chapter 3.

Next, I will explore a further potential secondary use of DCM data mentioned in the literature, with the aim of examining the practicality of DCM data for such a purpose.

2.3. Potential secondary uses of DCM data for benchmarking

In the literature, one of the potential secondary uses of DCM data highlighted is benchmarking quality of care within care settings (Brooker 2005).

Benchmarking in healthcare is defined as “the continuous and collaborative process of measuring and comparing the outcomes of key work processes with those of the best performers in evaluating organisational performance” (Lovaglio 2012: 2). Benchmarking involves a process of comparison, including the following steps: Identifying the area (e.g. process, event, or outcome) of comparison; developing key performance indicators² based on which the comparison will be made; and establishing or identifying a point of comparison (e.g. a set target or benchmark). Following these steps, a plan of action is developed to improve the identified area of concern. According to Lovaglio (2012: 2), two types of benchmarking can be used to evaluate quality performance by an organisation. Internal benchmarking is “a process (which) involves identifying best practices within an organisation, to compare best practices within the organisation, and to compare current practices over time”. External benchmarking is “a process (which) involves using comparative data between organisations to judge performance and identify improvements that have been proven to be successful in other organisations”

There are two important components to these definitions; the first is the necessity of choosing the right indicator relevant to the purpose of comparison (Goldstein and Spiegelhalter 1996). The second component is the need to ensure that data is suitable, comparable, consistent and most of all available before its collection to feed indicators (Nolte 2010; Ettorchi-Tardy et al. 2012). While indicator selection is the key issue in benchmarking

² A key performance indicator is defined as ‘a summary statistical measurement on an institution or system, which is intended to be related to the ‘quality’ of its functioning’ (Goldstein and Spiegelhalter 1996).

(Nolte 2010), feeding these indicators with effective data is also a major criterion for

successful benchmarking (Campbell et al. 2003). The effectiveness of data is assessed by characteristics such as suitability, availability, quality (completeness and accuracy) and comparability. All these are reported as major data requirements for benchmarking within healthcare (Nolte 2010). This section will next examine the practicality of DCM data for benchmarking in the light of these data requirements (e.g. suitability, availability, comparability and quality).

The choice of indicators and the related data requirements are also influenced by organisations' perceptions of benchmarking (Raleigh 2010), which consequently influence the approaches used for it, the selection of indicators and the relevant data to feed these indicators. This section will end with examining various perceptions and approaches for benchmarking within healthcare with a specific aim to analyse their applicability for DCM data.

I begin, however, by investigating various data requirements for benchmarking and position of DCM data in this context.

2.3.1. Data suitability

The suitability of data for benchmarking requires that the identified indicators reflect what needs to be measured. According to Campbell et al., (2003: 817), an indicator is a “retrospectively measurable element of practice performance for which there is evidence or consensus that it can be used to assess quality of care provided and hence change it”. Indicator selection and development for healthcare quality have been reported as a complex and challenging activity in benchmarking (Campbell et al. 2003). The major issue

in this regard is associated with defining care quality, as it is considered a multidimensional

concept (Lovaglio 2012), and then with assessing what needs to be measured to reflect care quality (Raleigh 2010). Further, the validity and reliability of any measurable indicator is also highlighted as a main requirement for benchmarking (Campbell et al. 2003). Raleigh (2010: 7), argues, “indicators should be seen as screening tools that prompt further investigation, rather than as definitive markers of quality”. Therefore, it is important that indicators should be carefully selected and interpreted. They should be fit for purpose and should not be interpreted wrongly, as this can be damaging for an organisation with a low ranking against that benchmark, stigmatising for the staff of that organisation and, most important of all, misleading for the patients (Raleigh 2010).

In the context of DCM, a number of questions are raised while assessing the practicality of DCM data for benchmarking. They include the type of indicators DCM might provide for assessing and improving the quality-of-care within formal dementia care settings and whether these indicators are valid and reliable for the specified purpose. I will explore these questions next.

A number of empirical studies have demonstrated the validity and practicality of DCM indicators for measuring the quality-of-life of people with dementia and the quality-of-care provided to individuals and groups and within care settings (Innes and Surr 2001; Fossey et al. 2002; Brooker 2005). Based on this evidence, DCM is recognised as a useful tool for care quality monitoring and improvement in key national policy and guidance. For example, the National Audit Office (National Audit Office 2010), in its report *‘Improving Dementia Services in England’*, endorsed DCM as a method for measuring

the quality of life of people with dementia. Likewise, the UK's Audit Commission

(Audit Commission 2000), in its *'Forget Me Not'* report on mental health services for older people, featured DCM's role in improving quality of care within formal dementia care settings.

Brooker (2005), in her literature review, also maintained that DCM is a unique method in that, unlike other quality-of-care and quality-of-life methods, it produces rich data that reflects the elements of both the quality-of-life of an individual with dementia and the quality-of-care provided within care settings. The WIB score is a measure of the average well- or ill-being of people with dementia observed using DCM, which is then interpreted further to understand quality of life and assess potential improvements over time (Bradford Dementia Group 2005). Further, DCM also provides an indicator of the type of activities provided within care settings, which may then be interpreted to assess the variations in activities across various types of healthcare. For example, the percentage of time spent in some of the activity codes (such as A, D, E, F, G, I, J, K, L, O, P, R, S) can provide an indication of the potential for positive engagement of people with dementia with their environment. According to the DCM Manual (Bradford Dementia Group 2005), positive engagement is one of the key determinants of quality-of-life. Similarly, PDs and PEs can also provide indications of the quality-of-care within a care setting. The DCM Manual highlights other indicators that could be extracted from DCM data, such as the levels of agitation and stress, withdrawal and passive engagement experienced by people with dementia. Together, the BCC and WIB can also provide an indication of the type of activities that can either improve individual or group-level wellbeing or that can contribute to their ill-being (Bradford Dementia Group 2005).

While there is evidence that DCM provides indicators as a measure of quality-of-life and quality-of-care (Brooker et al. 1998; Innes and Surr 2001; Kuhn et al. 2002; Brooker 2005), the applicability of these indicators as benchmarks needs exploration. The evidence from the literature shows that the WIB score is used as an indicator to assess changes in care over time (Brooker et al. 1998). For example, using the WIB score, a benchmark is set at the baseline and further mappings are conducted to assess changes in scores and thus care over time. While such a benchmark might not reflect best practice, it gives an indication against which further improvements can be assessed and signifies the potential for use of DCM data for internal benchmarking. For external benchmarking, however, a benchmark can be a calculated average of the group WIB scores of various organisations that show best practices of care and this can be used to draw a line between high- (above average) or low- (below average) performing organisations. While there is a lack of empirical evidence regarding the use of DCM indicators for identifying best practices in care, Brooker (2005) provided an example how this can be done by combining group WIB scores taken from studies (n=39) which were reporting data on various types of care settings such as day-care centres and long-term care. She found that the WIB score was higher in all day-care settings as compared to long-term care, as the mean group WIB score across all day-care centres was reported as 1.94, while for long-term care it was 0.9. While Brooker (2005) suggested the potential of the group WIB score as an indicator against which to benchmark, there is a dearth of any further research that attempts to assess the practicality or use of this or other DCM indicators for benchmarking.

In summary, benchmarking indicators need to be a valid measure for assessing and improving quality in order to be used effectively for this purpose. Evidence shows that some items of DCM data, such as the WIB score can be used for assessing changes in care over time, and are thus effectively used as a method for internal benchmarking. While there are suggestions in the literature that the mean WIB score across similar types of care settings can be used to indicate differences in care across setting types, there is no further evidence of the applicability of this data for identifying best practice or benchmarking. For this reason, further research is required to examine the applicability of DCM data for external benchmarking.

While data suitability ensures that the indicator is valid and reliable for specific use for benchmarking, the collected data needs to be comparable if used for benchmarking, as will be explored next

2.3.2. Data comparability

Comparability of data requires that indicators are comparable across organisations or over time based on similarities in functions, processes, methods and outcomes (Nolte 2010). Data comparability is reported as a major challenge when the aim is to compare organisations, especially when comparing internationally (Nolte 2010; Kossarova et al. 2015). Comparison should be made on a like-for-like basis but collecting such data is challenging as various types of care settings provide care differently, have different ways of collecting data and collect different types of data that might lead to similar outcomes (Nolte 2010). Therefore, as Nolte (2010) asserts, many confounding factors can influence comparisons of what may seem to be

similar settings, services or organisations. Where there is a lack of comparative data, the

process of case-mix adjustment³ can be applied to make the data comparable across organisations (Kritsotakis et al. 2008).

In the context of DCM, one of the important questions is whether DCM does or can provide comparable data that can be used for benchmarking. There is ample evidence that indicates users of DCM believe it is able to provide comparative data. For example, DCM data has been used for internal benchmarking or comparison purposes at individual and group levels across various types of care settings to assess changes in care over time. Brooker et al. (1998) collected DCM data from various types of UK-based care settings, within one NHS Trust, over a three-year period. Comparisons were made to examine changes in care outcomes across nine units (two day hospitals, two respite-care units, four continuing-care units and one assessment ward) over three cycles of mapping. While this study demonstrated a good example of comparing and assessing care changes over time, it has been criticised for not recruiting similar mapping participants in all DCM cycles (Cooke and Chaudhury 2012). However, arguably, while only 25% of study participants remained part of all three cycles, Brooker et al. (1998) study showed that many other important factors were taken into account for credible comparisons. For example, all recruited units had similar models and philosophies of care. This means that all units had single-bedroom accommodation and mixed-sex participants with separate sleeping, washing and toilet facilities and all staff were known by their first names. Further, the staffing ratio, the length of maps and patients' profiles were similar

³ A case-mix adjustment is a statistical process of adjusting for the differences between organisations and patient characteristics, thus allowing a fairer comparison.

supporting effective comparisons (Brooker et al. 1998). This demonstrates the inherent difficulties of comparing service settings over time due to the ever-changing patient/resident population within hospitals and care homes. Therefore, a consideration for benchmarking DCM is the sample representativeness, but also the additional data collected alongside DCM, which permits assessment of comparability or issues which may impact on this.

DCM data has also been used for comparison purposes while assessing changes in care across a number of settings. An American study, assessing quality-of-care in assisted-living (AL) facilities by Kuhn et al. (2002), provides an example of this. Using DCM data, Kuhn et al. compared the quality-of-care provided in both dementia-specific (n=7) and non-dementia (n=3) AL facilities. They used group WIB scores to compare care quality within both types of care settings. While considering specific features of each type of care setting, Kuhn and colleagues also indicate the feasibility of using DCM data to compare AL facilities with nursing homes or day-care centres to assess variations in care. However, this requires additional data (e.g. care settings' characteristics) collected alongside DCM data to compare similar settings.

The above studies indicate that DCM data can be used for comparison purposes. While all the studies used DCM to make comparisons either across time or between care settings, their underlying purpose was learning and care improvement rather than ranking units against each other, for example by suggesting some were providing good care and others poorer

levels of care. This is called comparative benchmarking within healthcare, as will be discussed in detail later.

Data comparability is the key to the benchmarking process and one of the major criteria of comparability is to see if data is consistent (Nolte 2010). This means that the collected data should show similar results when collected at different points in time, or within different organisations. Brooker (2005) found in her literature review that various studies had shown a similarity in some of the DCM data collected from organisations providing similar types of care (day-care centres or long-term care centres). Based on this finding, she proposed using DCM indicators (e.g. group WIB score and BCC profiles) for benchmarking. However, she further suggested that confounding factors, such as those related to the participants (i.e. dependency levels) and care-setting characteristics (i.e. type and size), needed to be taken into consideration.

Studies also suggest that residents' dependency levels may have an impact on the wellbeing of people with dementia. A study conducted by Edelman and colleagues (2004) with participants from special care facilities, assisted living facilities and adult day care centres, found that low mean individual WIB scores are associated with both high levels of cognitive impairment and increased activities of daily living (ADL) dependency. Similarly, another study (Thornton et al. 2004) reported significant correlations between WIB scores and individuals' total dependency levels and cognitive and behavioural functions. They found that wellbeing levels of those individuals living in continuing care settings and day hospitals are significantly higher, as they had lower dependency levels and fewer cognitive and behavioural issues. Chenoweth and Jeon (2007) also reported an association between lower WIB score and reduced physical function. The DCM Manual also highlights

that several aspects could influence the comparability of DCM data when used to

assess the change over time (Bradford Dementia Group 2005). For example, participants' dependency levels and their change over time and data being collected by different mappers are two major aspects, which need careful consideration.

In summary, there is evidence demonstrating the use of DCM data for comparison purposes across time and across organisations. Further, there is some evidence that DCM can produce consistent data across settings offering similar types of care, thus suggesting the potential of using DCM data for comparability.

While data suitability and comparability are important requirements, data needs to be available for benchmarking and to be of a certain level of quality.

2.3.3. Data availability and quality

Data need to be available on a regular basis for measurement purposes so that benchmarking can be associated with continuous improvement in quality (Meissner et al. 2006). What is of an excellent standard today might show a shift in expected performance tomorrow. The reference point, or benchmark, should be reviewed regularly (Kay 2007). Therefore, data needs to be collected on a regular basis to set the benchmarks that reflect already achieved improvements. Nolte (2010) asserted that data availability could be a major challenge for benchmarking, particularly when the aim is to collect data at an international level, as each country may have its own method of or tools for collecting data. This can have impact on data availability as well as on data quality and comparability for benchmarking (Kossarova et al. 2015).

Within healthcare, both routine and purposeful⁴ data is used for benchmarking. A number of studies demonstrate the practicality and suitability of routine data for this purpose. For example, a US-based study by Earl et al. (2005) used administrative data from Medicare to compare the intensity of end-of-life care for cancer patients. Hermann et al. (2006) and Meehan et al. (2007) compared existing mental-health indicators taken from several healthcare organisations and verified the usability of routine data for benchmarking. However, both studies recommended considering the use of case-mix adjustment processes for fairer comparisons. While routine data is appreciated for its regular availability, issues related to the quality and comparability of such data for benchmarking have been identified (Powell et al. 2003). In order to conduct effective benchmarking, where data is suitable, complete, accurate, available and comparable, the collection of a standardised and purposeful dataset is encouraged (Nolte 2010). One example of this is the Care Quality Commission's (CQC) project for monitoring health and social-care services across England (Care Quality Commission 2016). While ensuring the quality, regular availability and comparability of data, CQC used an extensive list of indicators to collect care-monitoring data (both quantitative and qualitative) from various sectors, including NHS acute trusts, GP practices and trusts providing mental-health services. While purposeful data collection provides a degree of control of quality and comparability of data, it also requires effective collaboration and planning for making data available for benchmarking (Ellis 2006).

⁴ Data collected specifically for benchmarking.

Data need to be complete and correct in order to provide information for a specified indicator (Campbell et al. 2003). Poor-quality data can raise a number of issues including the misinterpretation of the indicators (Raleigh 2010). Data quality issues are mostly apparent in routinely collected data, as data is not collected specifically for benchmarking. Using such data for benchmarking can be risky. For example, when routinely collected data is compared statistically, whether across time or with other care providers, variations in data are revealed. According to Powell et al. (2003: 122), “naturally, such variations imply ranking: that the measure reflects quality and that variations in the measure reflect variations in the quality”. However, if the data is not of good quality, such variation could be misleading in indicating ranking or change, when it may be reflective simply of variations or inaccuracies in collection.

In the context of DCM, data need to be available for benchmarking purposes. The availability of data is related to its collection by mappers and organisations and then its storage in electronic databases for benchmarking analysis. There is the opportunity for using previously collected DCM data, if it is stored in electronic databases. Based on the concept and data model of the DCM international database (Khalid et al. 2010), the University of Bradford has developed a web-based DCM database application called the arc|hive DCM database (Surr et al. 2015) that can be used as a potential resource for quality DCM data. This database provides some inbuilt data validation, as will be discussed in detail in Chapter 3. However, it still does not permit checks for quality to be conducted beyond only of actual DCM

codes. Further, while this system can make DCM data available more readily, it is not known how

frequently the system is being used. In addition, it only contains DCM data and not additional data that might be required for effective benchmarking. While the arc|hive DCM database can technically support data availability for benchmarking, the collection of DCM data however depends on individual mappers and organisations and their regular use of DCM.

Furthermore, DCM data needs to be of good quality for benchmarking purposes. Within the context of the primary use of DCM data, the literature shows that data quality is associated with the mapper's reliability score (e.g., inter-rater reliability; IRR) (Brooker et al. 1998; Thornton et al. 2004). The criteria for what constitutes quality DCM data for secondary analysis purposes is, however, unknown.

While in this section, the data-availability and data-quality requirements are highlighted in relation to benchmarking, the literature also underlines the significance of these requirements for other secondary uses of data (e.g., research purposes) (Weiskopf and Weng 2012). More details of data-quality and availability requirements for the secondary use of data within research context are presented in Chapter 3.

In summary, the literature highlights two important aspects of benchmarking: the indicator selection and the effectiveness of data to feed the indicator. The data is effective if it is suitable, comparable, available and of good quality to feed the indicators for benchmarking. Examining the practicality of DCM data for benchmarking within the context of the data requirements, there is evidence that DCM provides some valid indicators (such as the WIB score) for measuring quality-of-care and quality-of-life and that DCM can produce a

consistent set of data within organisations and comparable data across time. While such characteristics make DCM data suitable for internal benchmarking, there is still limited evidence that support the practicality of DCM data for external benchmarking. To do this would mean assessing the practicality of DCM indicators for comparisons, especially for identifying the best practices in care. While suitability and comparability of data are important requirements for benchmarking, the literature also indicates that DCM data should also be of good quality and available for benchmarking purposes. However, little is known about either of these areas with the existing literature.

Within any organisation, the data requirements for benchmarking also depend, in part, on the perceptions of benchmarking, which in turn influence the approaches to benchmarking, as discussed next.

2.4. Perceptions and approaches of benchmarking and DCM data

Initiated in the 1990s, benchmarking within healthcare has evolved over time. Benchmarking was used initially for the comparison of performance outcomes. A particular focus was on competitive analysis and on the aim of achieving a pre-set statistical target average by organisations to know if they were above or below average (Ellis 2006). Whilst this competitive view was interesting for the public, governments and commissioning bodies, it did not inform the organisations about how to improve their processes and performance. The desire for improvements led healthcare organisations to evolve the benchmarking concept to a point where the emphasis is now on understanding and analysing the best processes in the sector for producing

high levels of performance (Ellis 2006), thus making possible continuous quality improvements within organisations.

The modifications in the concept have also influenced the approaches used for benchmarking and the type of data collected for this purpose. Ettorchi-Tardy et al. (2012) put it as moving from quantitative to qualitative approaches, thus referring to shifting from objective to subjective types of data and methods of benchmarking. Based on the type of data and its use, benchmarking approaches are categorised as competitive, comparative and collaborative (Ettorchi-Tardy et al. 2012). Competitive benchmarking consists of measuring organisations' performances by comparing them with those of existing competitors, where the aim is to reach a statistical number to show whether the organisations are meeting the highest levels of performance achievement (Kay 2007). Whilst this approach gives care-provider organisations the satisfaction that they have achieved a pre-set target, it is usually criticised for ignoring patient experiences (Kay 2007), as this approach does not encourage the collection and analysis of the data that explains the processes of improved performance. Ellis (2006) asserts that it misses the main aim of benchmarking, which is to improve the level of performance by acquiring in-depth understanding of how others have improved their performance, an aim that is best met by focusing on more qualitative approaches. DCM data can be used for this type of benchmarking, as it provides quantitative data such as WIB scores that can be used for statistical benchmarks.

Comparative benchmarking focuses on how similar functions are performed by different organisations, including those that are performing best (Kay 2007). While this approach demands collecting sufficient information to inform reasons for improvements within an organisation, Ettorchi-Tardy et al.

(2012) maintains that it dilutes the sense of competition and focuses on learning from

each other's good practices within different organisations. Therefore, the aim is to collect statistical as well as qualitative data for a detailed analysis to understand how others have improved their processes. This is considered a positive approach to benchmarking, which reflects organisations' need to learn from others (Kay 2007). DCM data can potentially be used for comparative benchmarking, however, additional data is required to show the whole process of change. Currently, data is collected as DCM codes and qualitative information that might explain the reasons for the recorded codes but no additional information is systematically collected to explain the process about if and how care quality was improved. This would therefore need to be built into future DCM data collection and storage processes.

Collaborative benchmarking consists of the process of sharing knowledge about a particular activity/process, with the purpose of improving a specific area of care (Mosel and Gift 1994). This approach focuses on effective and healthy collaboration among organisations and ensures that the benchmarks reflect the views of both patients and their carers (Mosel and Gift 1994), thus providing as much data as possible for effective comparisons. The collaborative approach focuses on qualitative methods of data collection for identifying patient experience. A good example of this approach can be seen in the UK-based 'Essence of Care' project, which produced benchmarks of good and effective communication between patients and care staff (Department of Health 2010). While the above-mentioned approaches suggest collecting various types of data that are determined by the approach to benchmarking within the organisation (Ettorchi-Tardy et al. 2012), the

main issue, however, concerns the quality, availability, suitability and comparability

of data pertinent to meeting the users' perceptions of benchmarking (Nolte 2010). This type of benchmarking can be useful for DCM data, as organisations can together set an aim to improve quality of care and thus collaborate with each other to gather data on similar processes for comparison purposes. Effective collaboration is the key to this approach.

In the context of DCM, the choice of benchmarking approach should complement the underlying philosophy of DCM use. The literature indicates that DCM has always been used for the purpose of care improvements for individuals and groups within formal dementia care settings (Capstick 2003). The aim to date, therefore, has been to learn how care can be improved over time either at individual, group or organisational level. The similar approach to learning and care improvements can be used for benchmarking. This can also be reflected for the choice of approaches that facilitate learning from good practice and care improvement processes such as comparative and collaborative benchmarking. However, it is important to collect as much information as possible alongside DCM data for such types of benchmarking. It is not clear from the current literature, whether such information is or would be available.

As argued above, while DCM data can potentially be used for benchmarking, the choice of any approach to benchmarking depends on how users perceive the use of DCM data for this purpose. There is a gap in the literature associated with explaining why DCM data is required for benchmarking and what users' perceptions are about using DCM data for this purpose. Such requirements can give an indication of the types and characteristics of data that would need to be stored within the warehouse, when such a system is

designed. Finding user requirements for their potential secondary use of DCM data is the main aim of this study. Chapter 7 presents users' perceived potential secondary uses of DCM data and their implications for a future data warehouse.

2.5. Secondary use of DCM data and need for an effective IT solution

So far, I have argued the potentiality of DCM data for secondary uses such as benchmarking and that there exist some examples that demonstrate the potentiality of DCM data for further analysis within research context. This section discusses some of the additional motivations for the secondary use of DCM data and the need for an effective data management system for this purpose.

Every year more and more individuals and organisations are trained to use DCM, which may increase the use of DCM and thus the amount of data generated. These larger amounts of data increase DCM's potential uses for secondary purposes. Further, mappers are spread around the globe and there is evidence that DCM data is being collected from various types of care settings such as residential facilities (Lai et al. 2004; Chenoweth et al. 2009; Barnes 2013) and hospital wards (Woolley et al. 2008). The use of DCM is also promoted in non-dementia care settings, such as neuro-rehabilitation wards (Westbrook et al. 2013), assisted-living facilities (Zimmerman et al. 2005) and intellectual-disability residential services (Jaycock et al. 2006; Finnamore and Lord 2007). While there is yet a lack of knowledge on how regularly DCM is used at national and international level, its use in various types of settings and patient groups shows that it can provide rich and multi-purpose data. Such a rich dataset can allow one to see the possibilities for its

secondary use if it is accumulated over a period of time and collected from various mappers and organisations.

As mentioned previously in Chapter 1, a large amount of rich DCM data can also be part of existing and future initiatives/efforts to improve the quality of dementia care by utilising the existing datasets. However, to facilitate the secondary uses, either for benchmarking, research or any user-required purposes, DCM data need to be collected in an electronic, standardised and integrated format. This is also to deal with the issue which Sandra and Gramon (2007: 95) state is one of the major barriers in using healthcare data for secondary purposes, that of “locating existing data”. The integration of DCM data within a specific resource can deal with this issue. In order to support the secondary use of healthcare data, a number of studies suggests collecting data in a standardised format, developing appropriate ethical and legal frameworks to support data-sharing (Safran et al. 2007), collecting additional information alongside patients’ healthcare data for in-depth analysis, re-defining technical architectures and communicating and promoting the opportunities and benefits of secondary data (Health Industries 2009).

A DCM data resource for secondary uses can potentially be a solution to providing access to integrated and historic DCM data without expending effort on conducting DCM method. This suggestion is based on the observation that DCM is criticised for being a time- and resource-consuming tool compared with other dementia-care improvement tools (Beavis et al. 2002; Edvardsson and Innes 2010). Studies have highlighted that DCM training is expensive (Edvardsson and Innes 2010). Furthermore, conducting

observations is an intensive process that requires time and dedication from the mappers

(Beavis et al. 2002). Moreover, implementing DCM within care practice is also resource consuming (Beavis et al. 2002; Cooke and Chaudhury 2012), requiring leadership and managerial skills (Bradford Dementia Group 2014). A DCM data resource can provide an opportunity to explore data to identify new insights to suggest aspects of care that might be improved (Khalid 2010). However, DCM data require management at the point of collection and storage in order to render them shareable and then available in a data resource.

Considering the importance and usefulness of DCM data, Brooker (2005) also emphasised the management of DCM data in order to fully exploit their richness and highlighted the need for innovative, efficient and reliable IT solutions.

2.6. Summary of the chapter

This chapter began by analysing the empirical studies that provide examples of the secondary use of DCM data for research purposes. The lack of reporting of any related issues and concerns regarding the secondary use of DCM data within these studies was noticed. The chapter then went on to explore a potential secondary use of DCM data for benchmarking the quality of dementia care. This exploration revealed that data effectiveness is the key to a successful benchmarking process. The chapter then outlined the characteristics of effective data, such as suitability, availability, quality (completeness and accuracy) and comparability and referred to these as key data requirements for benchmarking. It then examined the practicality of DCM data in the light of data requirements for benchmarking. This commenced with a discussion of the suitability of DCM data for

benchmarking, including exploring potential DCM indicators, their usability and likely effectiveness for assessing and improving the quality of care within formal dementia-care settings. The chapter argued that, while sufficient evidence exists to indicate that DCM could provide some key quality indicators (e.g., the WIB score) and that there exists good evidence to indicate that it can be used effectively to assess changes in care overtime, further research is required to examine the potential applicability of DCM data for benchmarking.

The chapter then moved on to arguing that there is evidence demonstrating the use of DCM data for comparison purposes across time and across organisations, thus suggesting its use for comparability for benchmarking. Further, this chapter also highlighted the significance of data availability and its quality for benchmarking as well as for other secondary uses such as research purposes. The chapter then turned to arguing the case for establishing user perceptions of benchmarking, as these influence the type of benchmarking by giving an indication of the types and characteristics of data needed to be stored within the warehouse.

The chapter ended by highlighting the significance of integrated DCM data for secondary uses and argued the need for an effective data-management system for such a purpose. Data warehousing, as a technical solution for managing DCM data for secondary purposes, is examined in Chapter 3.

3. Data Warehousing within Healthcare: Benefits and Challenges

3.1. Introduction

This chapter has two aims. The first is to highlight the significance of data warehousing within healthcare, which will be followed by a detailed review of the previous study I undertook (Khalid 2010), where this was proposed as an IT-based solution for managing DCM data for secondary uses. The second aim is to emphasise the significance of users' views and perceptions regarding secondary use of DCM data, which can inform requirements for various aspects of a future data warehouse. In order to achieve both aims, the chapter examines the concept of data warehousing, its role, structure and functions in managing healthcare data and associated benefits and challenges. While doing so, it details and critiques my previous study. As indicated in Chapter 1 this chapter will develop the argument of the importance of recognising potential users' views about various aspects of a DCM data warehouse, thus making the case for adopting a user-driven approach. The main argument of this chapter is that, while my previous study proposed a successful technical structure for a DCM data warehouse, in order to design a user accepted warehouse, there is a need to identify specific user-identified functions (uses), benefits and challenges that may be used to inform identification of clear user-driven requirements for the warehouse's future design and development.

3.2. General concept, process and structure of data warehousing

Sahama and Croll (2007) identify that defining the concept of a data warehouse largely depends on the background and views of the definer. This is the reason why there are several definitions of a data warehouse, which reflect individual's own views, understanding and perceptions. For example, Kimball (2008) views a data warehouse as a *system* that extracts, cleans, transforms and delivers source data into dimensional data store (a data repository) and then supports and implements queries and analysis for the support of decision-making. While Kimball focuses on the functionality of a data warehouse, Inmon (2012a) emphasises the characteristics of a warehouse and focuses on data storage aspect and views a data warehouse as a *repository* that stores integrated data in support of management's decision-making process (Inmon 2012a).

Harjinder and Rao (1996) argue that data warehouse is a *process* of data integration from various source systems, including historic and external data, to facilitate structured queries, analytical reports and decision-making. When viewing a data warehouse as a process, the term 'data warehousing' is also used within the literature. Barquin (1996) refers to a data warehouse as a collection of techniques that provides a systematic and pragmatic *approach* to solve user problem in accessing information that is distributed in different systems within an organisation.

Despite of viewing a data warehouse as a system, repository, process and approach, all definitions share a common understanding about a data warehouse. By synthesising the above definitions, it could be argued that a data warehouse (or data warehousing) is a concept. The concept that refers

to the processes of data transferring from various sources into a data warehouse repository (e.g. extraction and transformation component), data storage into a dimensional format (e.g. data storage component) and further data dissemination to the users using a variety of analytical tools/applications (e.g. data dissemination component). A general structure of data warehousing is shown in Figure 2. While these components define the technical and management aspects of the data, data warehousing also includes a governance structure that ensures the availability, quality, interoperability, accessibility and security of data (Ponniah 2001). According to Ponniah (2001), a good data governance structure/plan deals with both technical and organisational aspects of a data warehouse, thus ensuring its success.

In order to implement a data warehousing concept, a range of technologies and applications are available for data extraction from the sources, its transformation and storage and then retrieval by the users. While it is important to consider the right technologies and applications for a data warehouse during its development, it is out of the scope of this study to discuss the applicability and suitability of specific technologies and applications for a DCM data warehouse. The present study instead focusses on the higher-level concept of data warehousing in terms of identifying user views and discussing them as requirements that can inform the overarching design of a DCM data warehouse and not the subsequent technologies and applications that are then used to develop it.

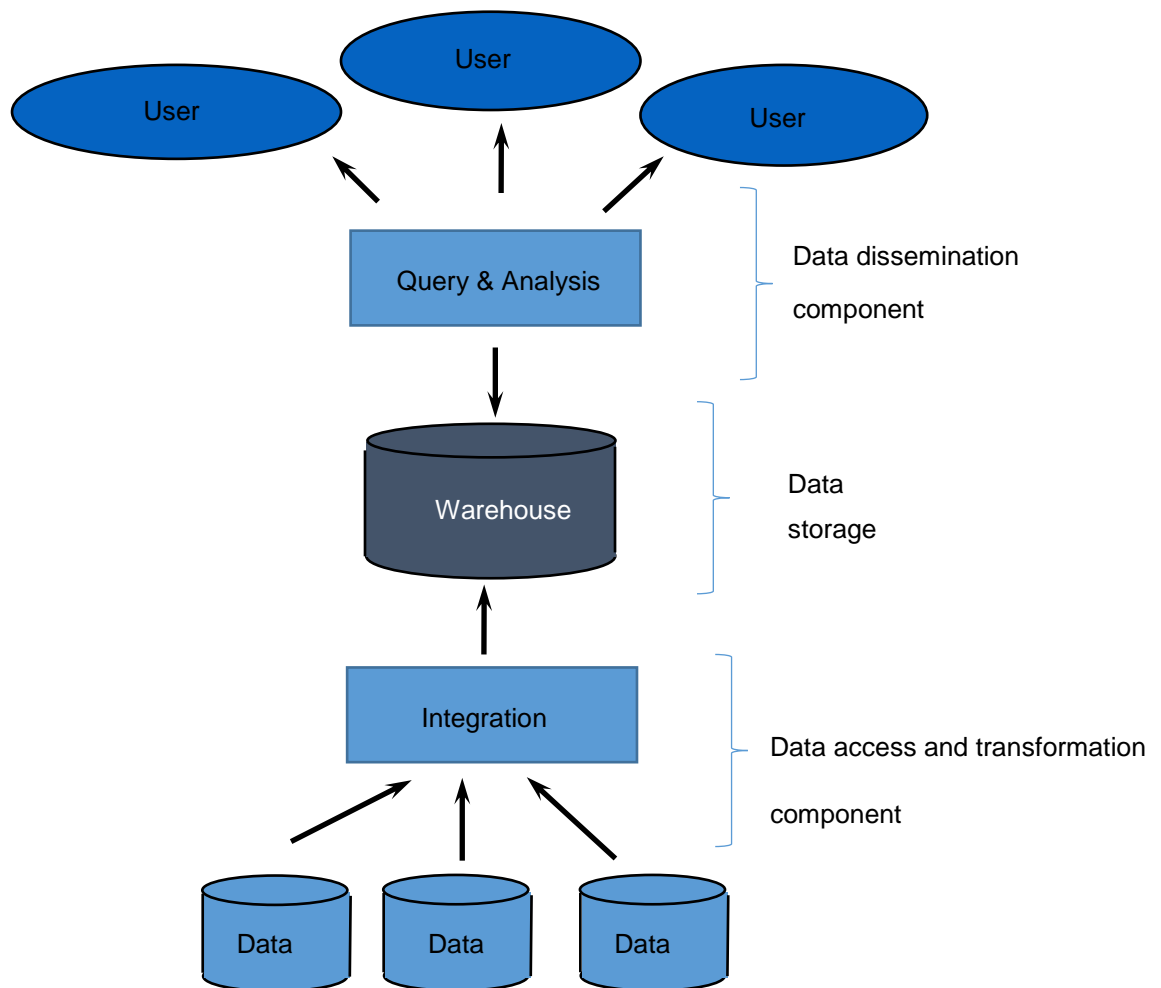


Figure 2: A general structure of data warehousing

A data warehouse is usually confused with a traditional database. While both act as data storage repositories, there exist some significant differences, which are important to highlight. These are discussed next.

3.3. A data warehouse repository versus traditional database

A data warehouse is a type of database however, its concept, functions and underlying technologies vary from traditional databases. Batini et al. (1986) maintain that a traditional database is the collection of operational or transactional data (e.g. data that is collected, used, and manipulated for everyday use) that is needed to make possible the daily operations or transactions of the organisations. For example, in healthcare, these systems

are used to register new patients and execute pre-defined operations, such as booking appointments, adding, deleting and updating patient's details, retrieving patient's previous appointment or diagnosis history etc. One example of such a clinical information system is SystemOne, which stores patients' healthcare records within England (SystemOne 2016). These clinical information systems are used by the healthcare professionals to support their primary use of data to improve care delivered to individual patients.

In the context of DCM, the traditional databases are those systems, which are used for collecting, storing and analysing DCM data for primary purposes. For example, the arc|hive DCM database and Excel programme provided by the University of Bradford. Both primary data management systems are designed to undertake basic calculations on the collected DCM data. However, the Excel programme does not provide the facility to input the entire DCM data collected during one mapping (Khalid 2009). Only mapping participants' BCC and ME values, collected during one mapping, can be stored and processed using this system. The input of qualitative notes, PDs and PEs is not possible. The arc|hive DCM database is a newly purpose built web-based database application for managing DCM data for primary purposes (Surr et al. 2015). This means individual mappers and organisations are able to input, store and analyse all coding frames and qualitative notes from their own mapping in one place.

Data warehouse is a database that stores 'derived data', which is collected from various operational databases for retrospective uses such as, reporting and research. A data warehouse differs from traditional databases as it:

- Stores data that is usually collected over a long period of time (historical data)
- Stores data based on user demands, which is collected and integrated from various sources (integrated data)
- Contains read only data (non-volatile data), which is updated on planned periodic basis (can be daily, monthly or annually, based on user requirements).
- Stores data that is organised so that system is optimised for answering complex queries from users and applications.

According to Sheta and Eldeen (2013), transactional databases are designed to answer 'who' and 'what' questions from the data, which are usually asked on a regular basis and do not consist complex queries. Therefore, a transactional database is usually designed using a relational data model⁵, which ensures efficient data entry and storage. However, a data warehouse contains data that is organised to answer 'what-if', 'what-next' and why questions for analytical processing. Therefore, a data warehouse is usually designed using a dimensional data model⁶, which supports multidimensional analysis and complex querying and reporting needs (Kimball et al. 2008). The significant differences between both a database and a data warehouse also require designing these repositories differently, using different

⁵ A relational data model represents a detailed level (for example, individual patients' details) data stored in various linked tables (Codd, 1970). Each table stores data that represent a single entity and its related information called attributes, such as a patient (entity) table with information (attributes) about their name, age, gender, address and patient identification number etc.

⁶ A dimensional data model provides structural arrangement of the data in a way that enables data users to look at a large number of inter-dependent aspects of the data from different analytical angles to support analysis and decision making activities (Chaudary and Dayal 1997). For this purpose, a dimensional model is represented in a star shaped format (star schema), which is recognised as an effective form of data representation in most data warehouses (Gray and Watson 1998).

data structures and focusing on different types of requirements. Chapter 4 covers the discussion regarding data warehouse requirements.

Next, I will explore the role of data warehousing within healthcare data management.

3.4. Healthcare data management and data warehousing

Data management is defined as “a group of activities relating to the planning, development, implementation and administration of systems for the acquisition, storage, security, retrieval, dissemination, archiving and disposal of data” (IGGI 2005: 6). As was also highlighted in Chapter 1, it is an uncontested view that information technology is playing an important role in providing effective and robust solutions for healthcare data management (Jensen et al. 2012; Wager et al. 2013; Yang et al. 2015). These solutions include tools for collection, integration, storage, and retrieval of data for primary and secondary use with the aim of enhancing the quality and cost-effectiveness of care (Yang et al. 2015).

The major use of technology within healthcare was the transformation of healthcare information in digital formats, which then led to the development of clinical information systems. Clinical information systems are used to collect, store and analyse data that is collected about patients’ diagnosis, treatment and care. An example of such data are Electronic Healthcare Records (EHR), which provides comprehensive information about patients’ healthcare (Kalra and Ingram 2006). Alongside clinical information, healthcare providers also collect administrative data including patients’

demographic details and financial and claim records in some healthcare systems (Safran et al. 2007).

The development of these systems were an initiative taken by many governments to improve patient's healthcare experience and its efficient delivery (Sheikh et al. 2011; Greg 2013). These systems are referred to as operational systems, which are designed and optimised to perform primary functions of improving patient care at local level within individual care settings. Whilst the primary reason was to facilitate effective use of healthcare data in individual care provider organisations, data stored in digital formats also make it easy for sharing and transferable across care settings for the purposes of secondary uses (Yang et al. 2015). The 2013 report, commissioned by the Department of Health to provide a review of the potential benefits from the better use of information and technology in health and social care, emphasises that information about patients should be recorded in formats that maximise the usability and value of information for primary and secondary purposes (Department of Health 2013a). I will now focus on discussing the role of IT for secondary use of healthcare data.

A significant amount of research and practical evidence contribute in emphasising the need and importance of the secondary use of healthcare data. Some of the purposes are to: inform areas of quality improvements and decision making at organisation level (Yang et al. 2015); detect general medical problems (Jensen et al. 2012); allocate resources (Kuo 2011); treat patients and find cures for various types of diseases; and predict epidemics (Yang et al. 2015). The secondary use of healthcare data involves conducting analysis on data that is collected and integrated taken from various operational data sources. In order to make this process efficient and easy, there require

effective and robust data management solutions/systems with the specific aim of its potential secondary uses.

The research and practical evidence shows that a range of technologies and systems have been designed and developed for effective management of healthcare data to facilitate its secondary use. Some of the main examples are Clinical Data Repositories (CDR) and Decision Support systems (DSS) (Gray and Watson 1998). These systems embed the concept of warehousing data that is taken from various operational systems (such as clinical, administrative and financial systems) and stored in a format that facilitates secondary uses such as, reporting, data mining and research.

Healthcare is embracing the concept of warehousing for its everyday generated data to extract knowledge for clinical decision-making, better integrated care, operations, resource planning, and medical research (del Hoyo-Barbolla and Lees 2002; HiMSS 2009; Chen 2012; Mohammed and Talab 2014). A growing body of literature indicates that warehousing may offer a robust solution for data management for secondary uses in a number of healthcare areas. For example, data warehousing has been successfully used to manage information related to infection control data for assessing quality performance (Wisniewski et al. 2003), clinical data for financial analysis (Silver et al. 2001) and estimates for cost and frequency of adverse drug events (Einbinder and Scully 2002). Further, data warehousing has also been used to facilitate the reuse of rehabilitation data to support clinical benchmarking (Completo et al. 2012); specifying and detecting clinical phenotypes for quality improvements (Post et al. 2013); the interrogation of

medical bioterrorism surveillance data to facilitate pattern and anomalous situation recognition

(Berndt et al. 2003); clinical data mining (Zhou et al. 2010); and secondary research using patients' administrative data concerning their admissions, outpatient appointments and accident and emergency visit data within NHS hospitals in England (Health and Social Care Information Centre 2015b).

Next, I will explore some of the determinant factors for using data warehouses in healthcare. These are the value of historical data for a range of analysis, data linkage and integration and further availability of large amounts of healthcare data for analytics such as data mining.

3.4.1. Historic value of healthcare data

A data warehouse provides an electronic repository for historic data collection to support its further analysis and reporting. It has been argued that healthcare data become more valuable as they become older and historic (Inmon 2012b). For example, historic healthcare data can identify patterns and trends of diseases in relation to various treatments and patients' lifestyles (Yang et al. 2015). Further, data collected from individual patients' care records can be integrated overtime to see trends and patterns that have occurred longitudinally. For example, the General Practice Research Database (2010) has collected and integrated anonymised data from patient records from nearly 6000 general practices nationwide since 1987 on a regular basis. Run by the UK Medicines and Healthcare products Regulatory Agency of the Department of Health, the data from the GPRD is used for a variety of medical and public health research purposes. This includes investigating: side effects of various type of medicines; causes of disease and medical disorders and associated link factors; outcomes of treatments

and areas of unmet medical needs; and treatments or services that work best (GPRD 2010). However, such data

linkage and integration require robust technical data management solutions, such as data warehousing.

3.4.2. Data linkage and integration

A data warehouse provides a platform to facilitate data linkage and integration activity. As healthcare is provided in several types of care provider organisations (such as local GP practices, hospitals, care homes), it is widely accepted that integration or linkage of healthcare data can produce new insights and knowledge, which can be used for a range of secondary purposes. For example, by using the Data Linkage and Extract Service, in England, the Department for Transport linked their road traffic injury database to the HES (Hospital Episode Statistics) (Health and Social Care Information centre 2015c) database in order to produce a rich dataset to explore accident circumstances and medical diagnosis (Administrative Data Liaison Service 2012). At a global level, in Sweden, a study that recruited 12-million persons' (blood donors and recipients) linked datasets and its findings contributed towards negating the myth that blood transfusion can transmit Cancer (Edgren et al. 2007).

From a technological perspective, two main approaches are used for linking and integrating healthcare data for secondary uses. One is the hub-and-spoke repository architecture while the other is point-to-point information exchange architecture (Stolba 2007). Both of these data integration approaches support different technical architectural ways of collecting, integrating, storing and using data. Within the point-to-point architectural approach each data provider develops and maintains its own databases

locally; data are then integrated when needed on request (Stolba 2007).

Using point-to-point information

exchange architecture for data management, various interfaces are designed and developed between data providers and data users for communication and data-sharing purposes. However, there are many issues with this latter approach, including data consistency and quality. Further, it is cost and time consuming to interpret various data sets into required formats whenever the integrated data is needed, which can cause consistency issues because the same data is exchanged between systems many times (Stolba 2007). It is therefore challenging to manage such systems and govern data within it.

The hub-and-spoke architectural approach on the other hand provides a solution for collecting data from various data providers and storing them in a repository before disseminating them to a variety of users. This approach provides a number of users with access to timely and consolidated data. For example, within the National Health Service (NHS) in the UK, Hospital Episode Statistics (HES) is a data warehouse that stores administrative data of patients' admissions, outpatients' appointments and Accident and Emergency details within NHS hospitals in England (Health and Social Care Information Centre 2015b). The relevant data is collected about each patient (demographic, clinical and administrative) during her/his stay within the hospital and submitted by the hospital to the HES to claim payments for the care that is delivered. HES collects this information from all NHS hospitals in England and is designed to be used for secondary purposes, i.e. non-clinical purposes. A wide range of users can have access to the HES data, such as local commissioning organisations, healthcare-provider organisations, researchers and commercial healthcare bodies, national bodies and

regulators and patients and carers (Health and Social Care Information Centre 2015c).

A range of systems have been designed and developed based on the above mentioned two approaches for data integration, which are based on the common objective of supplying healthcare providers with integrated and good-quality data to improve the quality of healthcare through decision making and research (Stolba 2007; Chisholm 2008). However, considering the issues with point-to-point data exchange architecture, many have suggested the hub-and-spoke architectural approach as the preferable approach (Chisholm 2008; Stolba 2007).

3.4.3. Analytics on healthcare data

Data analytics is a major activity associated with healthcare data, which is increasing in volume on a daily basis. The main reason is the use of information technology that has eased the way healthcare data is collected and thus has contributed to increasing its volume. Further, the integration of this data produces large amounts of data sets. Chen (2012) argues that such a huge amount of data is useless until it is further explored for identifying meaningful information. Abidi (2001) also adds that healthcare generates 'rich data' but 'poor knowledge'. In order to explore healthcare data for meaningful information and useful knowledge, a range of analytical tools are available, which when applied to data stored within the warehouses can provide useful outcomes. One of these common sets of tools for analysis purposes is data mining tools. Data mining tools are used for extracting information about the trends and patterns within the data for better decision-

making (Zhou et al. 2010). The data mining process within healthcare produces insights into the

integrated data that are not usually obvious in individual datasets prior to their integration (Yang et al. 2015).

Chen (2012) and Abidi (2001) emphasize the need of warehousing healthcare data so that various analytical tools including data mining tools can be implemented to explore data to retrieve useful information. A range of examples is available in the literature to demonstrate the applicability of this. For example, a study applied data mining tools to the data obtained from NIHRD (National Health Insurance Research Database; NHIRD) datasets from Taiwan's national health insurance database. The authors studied the comorbidity of ADHD (Attention Deficit/Hyperactivity Disorder) and found associations with psychiatric disorders in ADHD children (Chen et al. 2003). Within the UK, an example of administrative healthcare databases is the General Practice Research Database (GPRD 2010). These databases store data collected from the operational systems used within various general practices over a period of time to facilitate research studies. A DCM data resource with integrated and historic data can also enhance the potential of DCM data for various types of analytics such as data mining and reporting (Khalid 2010). However, this requires a data warehouse solution.

So far, I have explored the significance of data warehousing within healthcare, which provides a technical platform to store historic and integrated data for various analytical purposes. For the same reasons, a concept of data warehousing for DCM data was also proposed in my previous study. This study is examined in detail next with the aim to describe the proposed functions/uses and structure of the data warehouse. Various

technical features will be highlighted that were proposed and then developed
for the purpose of

that specific study. Further, the significance of user views are also highlighted in relations to various aspects of data warehouse, which will underline the limitations of my previous study.

3.5. DCM data management framework: a data warehouse approach

As was discussed in Chapter 1, in a previous study I proposed a two-step framework for warehousing the DCM data for secondary uses. A two-step framework includes three main components, as shown in Figure, 3 and are explained next.

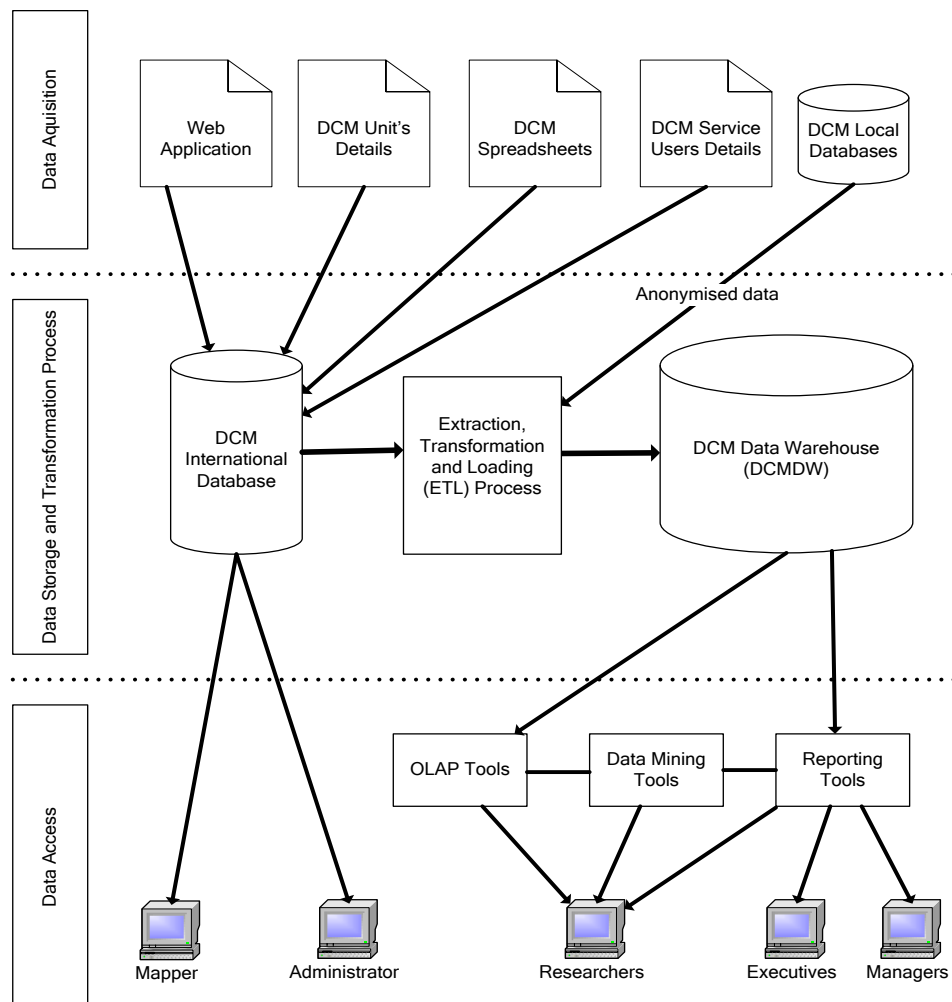


Figure 3: A two-step approach for warehousing Dementia Care Mapping (DCM) data (taken from Khalid 2010: 66)

3.5.1. Data-access component

This component deals with data sources which provide data for storage purposes within a warehouse repository. Within DCM data warehousing, a DCM international database was proposed as the main data source for the DCM data warehouse (Khalid 2010). There were two main reasons that influenced this recommendation. The first was that the DCM international database application is the only electronic system to date that will be collecting DCM data from national and international mappers. Second, the DCM international database is specifically designed for storing only quality DCM data, as the interface is tailored to deal with data quality issues (Khalid 2010). For example, in DCM, some coding combinations of BCC and ME are not permitted (e.g. B-1). In the DCM database, wrong combinations of codes cannot be entered as, at the interface and within the database, rigorous data-validity aspects were considered while designing and developing the first prototype (Khalid 2009). The University of Bradford's DCM database, the 'arc|hive', is built on the concept and the first prototype of the DCM international database (Khalid 2009) and therefore it holds the same characteristics in terms of dealing with data quality issues.

The extraction of quality DCM data, collected and stored from various individual mappers and organisations, would make it possible to conduct secondary analysis on the richer and wider DCM data. The process of extraction and integration is a process of data transfer between two databases. This process is explained next.

3.5.2. Data extraction, transformation and storage component

Extraction and transformation process

For warehousing, data is extracted from a range of identified data sources, transformed into a compatible format and loaded into the warehouse repository for further use (Skoutas and Simitsis 2007). As mentioned above, within the DCM data-management framework, the primary source of the data warehouse will be the DCM international database or the arc|hive database. However, DCM data from a variety of other sources could also be acquired within the data warehouse in order to provide data to support a range of users' needs. At the data transformation stage, the quality of the extracted data is checked again for anomalies, irregularities and incompleteness. Once the data is considered quality data, it is transformed into various formats before being loaded into the warehouse. The formats are determined by user requirements of information from the system. My previous study performed an extraction and transformation process for moving data from the DCM international database to a DCM data-warehouse repository. While efforts were made to transfer from one database structure (a relational model of the DCM international database) to another (a multidimensional model of a data warehouse), the process was relatively less complex, as the extraction and transformation process was established between two database structures for demonstration purposes. According to Takecian et al. (2013), this process becomes complicated when data need to be acquired from various sources which might not share the same structural format, nor the same physical location. In these cases, issues of data quality and security can emerge, as

will be explored later in this chapter. User requirements might identify data that need to be collected

from sources other than the DCM international database and therefore issues of quality, security, linkage and, most of all, availability can emerge.

DCM data warehouse repository

The function of a data warehouse is to store DCM data, which is primarily taken from the DCM international database, spreadsheets, flat files and other databases, in a format that provides easy and efficient accessibility for potential secondary uses. As with any other database, the data within a warehouse is stored within tables. However, these tables are called fact and dimension tables (Ponniah 2001). The fact table contains facts that can be multiple and that are referred to as important data attributes within a table. These data attributes have numerical or calculated values (Browning and Mundy 2001). Dimension tables, linked to the fact table, express the details and justification of the facts for rich and detailed analysis. For example, in the DCM data warehouse, the fact about the wellbeing of people with dementia can be explored from various dimensions related to the particular gender, area, care setting, age group and mappings.

Gallo et al. (2010) argue that the entities (such as people with dementia, care settings and mapping sessions in the context of DCM), about which data is collected, usually have many characteristics and, as asserted by Moody and Kortink (2000), the data is addictive and users usually want to explore all these characteristics and require a multidimensional view of data. A dimensional data model containing fact and dimension tables can provide a multidimensional view of data. These views are usually acquired through

user requirements and enable users to analyse the data from different angles. For

example, the user can ask for information such as the well-being or ill-being score of an individual service user at unit level, organisational level or even national and international level in a specific time period, which can be specified as days, weeks, months, years or decades. The user can also ask for details of all maps undertaken by a particular mapper or within a particular care setting, specified in terms of type and size, and in a particular area specified as local, national or international. While assumptions have been made that users might have such requirements for a multidimensional view of DCM data, there is no evidence concerning whether or how users perceive the use of DCM data for secondary purposes. To develop a user-accepted data warehouse, users' views and perceptions of their potential data use need to be obtained.

Parmanto et al. (2005) assert that, within healthcare, data usually need multi-level analysis. Therefore, it is important to find the right granularity for different levels of analysis. This means identifying at exactly what level the analysis of the data is required. For example, if it is required to provide a summary view of all patients who have been diagnosed with a specific disease, the data is required at the lowest granular level, which means an aggregated view of all patients' data. However, if the aim of the analysis is to identify a specific disease history of a single patient, the data needs to be available at the highest granular level, which means looking at individual patient-level data. Finding the right level of granularity requires storing data at various granular levels. A multidimensional data model supports this, including data storage at an atomic level, where individual patient-level

information is obtained for analysis, and at an aggregated level, where data is summarised

and aggregated based on the needs of a specific type of analysis. Further, data arranged at various levels also support drill-down and roll-up activities, where the aim is to conduct in-depth analysis. To exemplify this, Parmento et al. (2005) discuss their need to analyse the rehabilitation data at various levels such as a summarised report of each episode of care for each patient and progress information of each patient between treatments. They generalised that such analysis at a multiple level is imperative in healthcare, usually when the aim is to analyse healthcare outcomes.

Data is arranged at various granularity levels within the fact tables (Ponniah 2001; Stolba 2007). Depending on user needs, this means that sometimes the data is summarised or aggregated to make its availability easy and efficient for users. For example, an organisation manager who is managing many units within her/his organisation might require the average wellbeing/ill-being (WIB) score of all service users within each unit or across all units and she/he might want to see the WIB score based on the gender of service users. DCM data therefore need to be summarised or aggregated to answer such questions. Sometimes the data within the fact tables is arranged at the highest granularity level so that the information can be retrieved at a basic/atomic level. For example, information about each service user's BCC or ME, or related PDs or PEs, can be extracted with reference to a particular mapping, care setting and time period. Likewise, information about a particular care setting having a specific number of PDs or PEs can be extracted with reference to the specific time period. The data within the warehouse can, therefore, be arranged in a variety of fashions for easy and

efficient retrieval. However, the data warehouse needs to be designed with the specific DCM data requirements in

mind. Such requirements can be elicited from the users, from existing systems and from their documentation. However, the literature emphasises user involvement in terms of identifying their data needs, arguing that potential data users are familiar with their analytical needs and the levels of detail they require for a specific analysis (Browning and Mundy 2001).

The previous study that I undertook (Khalid 2010) proposed a data model for a DCM data warehouse. The effectiveness of this data model was demonstrated by running some experimental queries on limited DCM data. The proposed data model consisted of two anticipated facts from the DCM data, such as 'Fact DCM' and 'Fact WIB'. These facts tables were created to represent data at both the highest and lowest granularity levels. Fact DCM (Figure 4) was created at the highest granular level, where the fact about each service user having BCCs and MEs, PDs and PEs, can be extracted according to the individual mapping sessions in a particular unit and in a particular area. A range of queries were written using this table to extract data at an atomic level.

The following exemplary queries also show the multidimensional view of DCM data:

- What are the five most common BCCs for maps conducted during 2007, 2008 and 2009?
- How many mapping sessions were done in UK Trust Homes in 1999 & 2000?
- What is the % of +3's MEs scored on a group level across all units in maps from 1 Jan 2008 - 30th June 2008?

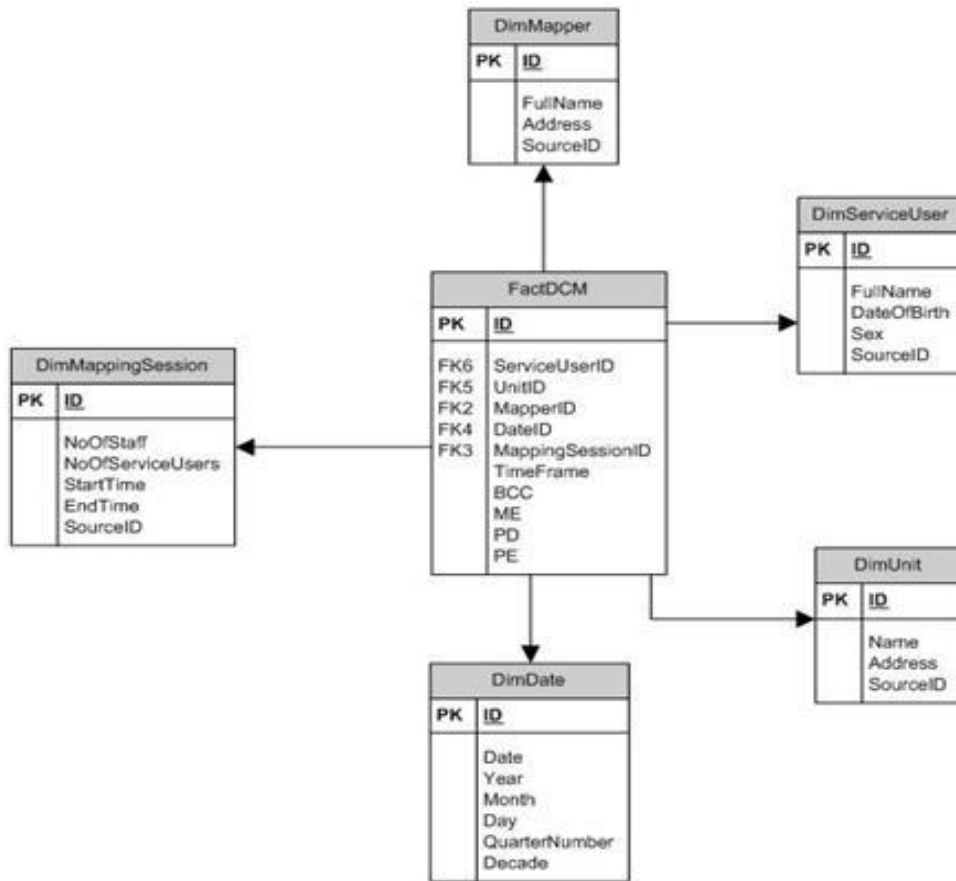


Figure 4: A fact table (data arranged at highest granularity level) and various dimension tables (taken from Khalid 2010: 73)

Fact tables can be created showing aggregated data around specific facts to support queries based on summary numbers (Shahbaz UI Haq 2016). Therefore, to demonstrate the data aggregation, a fact table, Fact WIB, was created to calculate the well- or ill-being (WIB) average measure in service users (Figure 5).

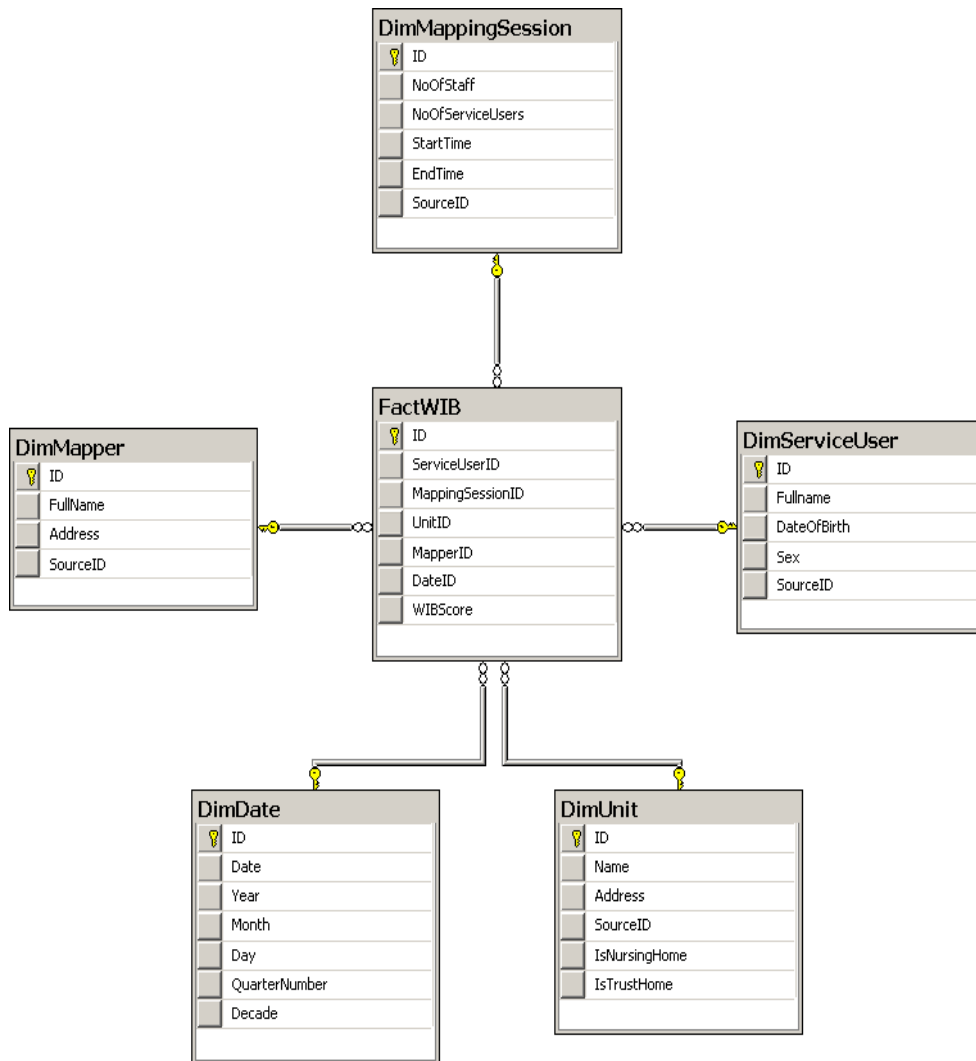


Figure 5: A fact table (data arranged at lowest granularity level) and various dimension tables taken from Khalid 2010: 79)

The following exemplary queries were written to demonstrate the data retrieval at a lowest granularity level.

- What is the average group WIB score for maps conducted during 1999?
- What are average WIB Scores of all female service users in Trust Homes and Care homes ?

While my previous study proposed a multidimensional data model for a DCM data warehouse that show data arranged in both high and low granular forms (as shown above), the users' views regarding their potential secondary uses

were unknown. This information can reflect the user-required dimensional aspects and the granularity of the data, which needs to be considered for designing a user-acceptable DCM data warehouse.

Data within the warehouse can also be stored at various visibility levels depending on user requirements and on the use of the data within the warehouse. For example, a data warehouse designed to store public data for secondary uses, such as research, is usually required to store anonymised or pseudo-anonymised data. While in the previous study that I undertook (Khalid 2010), a concept of the warehouse was proposed for managing DCM data for secondary uses, however, the data-visibility levels were not known as these depend on user requirements.

3.5.3. Data dissemination component

Data warehouses provide a data-sharing platform as well as a data repository that stores historic and integrated data in a multidimensional format (Ponniiah 2001). A data warehouse either provides access to the raw dataset or generates analytical reports from the data (Browning and Mundy 2001). Different analytical applications could be implemented in the warehouse to provide different users with the opportunity to extract data for a number of purposes, such as reporting using an online analytical processing tool (OLAP), decision-making using decision-support systems (DSS) and data-mining⁷ applications (Chaudary and Dayal 1997). The OLAP tools provide users with the facility to drill down to the atomic level or to roll up to

⁷ Data mining (DM) is the process of extracting the most valuable knowledge from a large amount of data.

the aggregated data for summaries (Ponniah 2001). Users are able to analyse the data from

different angles. For example, it is possible to find out information such as the WIB score of each service user in a particular unit in a specific time period, or the details of all mapping sessions undertaken by a particular mapper in a particular year etc. A DCM data warehouse could enable users at different authority levels to customise the process of information retrieval using a multidimensional view of data on OLAP tools.

A data warehouse also provides an ideal platform for retrieving data for data mining. For data mining, the data needs to be clean, of good quality and available in granular form so that various trends and patterns can be extracted at any level of data granularity from different aspects. A data warehouse provides an integrated environment with historical data stored at different granularity levels and abstraction levels for data-mining techniques. The literature offers a plethora of examples where data warehouses are used for providing historic and integrated data for data-mining purposes (Zhou et al. 2010; Yang and Chen 2015; Yang et al. 2015). In my previous study, I also provided an example of using DCM data from the warehouse for data-mining purposes by using a small amount of simulated data for demonstration purposes (Khalid 2010). It is important to capture the user requirements in order to design a data warehouse that can disseminate the required data to various users. There is a lack of knowledge about what type of data the potential DCM data-warehouse users would require and for what purposes within the warehouse.

In summary, my previous study demonstrated the technical feasibility of collecting, integrating and storing DCM data for secondary purposes. It also showed the workability of the data-warehouse structure for providing data for

secondary uses such as data mining (Khalid 2010). However, the major limitation of this study was the lack of potential users' involvement in it. Browning and Mundy (2001) assert that users' information requirements reflect the type of data they need and their analytical needs (e.g. the levels and dimensions on which they want to explore the data), which are significant for designing data models for the creation of various fact and dimension tables and data attributes. The data models designed using users' information requirements can be developed into systems that are not just technically successful but user-acceptable as well (List et al. 2002). Therefore, a study was required to gather user requirements for a DCM data warehouse.

Further, various functions of a data warehouse have different technical and social implications. While the technical aspects define data formats, granularity and visibility levels, the social aspects are related to data quality and security issues for a data warehouse. A data warehouse designed for quality improvement or decision-making within an organisation has different implications in terms of data quality and security than does a warehouse designed for storing and disseminating data for general research purposes. While the previous study (Khalid 2010) proposed the technical concept, it was limited in discussing the data quality and security issues related to warehousing DCM data. The next section explores these issues in detail.

3.6. Data quality and security challenges

So far, I have explained the concept of data warehousing and demonstrated that it provides a technical platform for linkage and usage of healthcare data

for storage and analytical purposes. A data warehouse stores data that is taken from various sources for the purpose of secondary use, which means

that data is usually used away from the time, place and purpose of its original collection (primary use) (Gardyn 1997). Whilst, as discussed earlier, secondary use of healthcare data is enabling exciting opportunities for improving healthcare, the literature signifies a number of issues, challenges and requirements for secondary use. The most commonly reported challenges are related to data quality and data security. With the aim of examining both of these issues in relation to healthcare data and DCM data in detail in this chapter, I will first discuss the data quality challenges and requirements. Data quality is a complex concept and is defined using various dimensions, as will be explained later. I will therefore first explore how the concept of data quality is defined and what approaches are used to identify and define various quality dimensions in relation to information systems, particularly data warehouses. Based on this concept of data quality, I will then examine various studies to illustrate the challenges associated with data quality within healthcare data warehousing. I will also highlight the literature that signifies the importance of user views for identifying issues that can further inform the data quality aspects within a data warehouse.

3.6.1. Data quality

Data quality is considered a multidimensional concept (Wang and Strong 1996; Weiskopf et al. 2013). Wang and Strong (1996: 6) define a data quality dimension “as a set of data attributes that represent a single aspect or construct of data quality”. Quality of data can be assessed through its various dimensions. In the context of information systems, a number of studies reported using various theoretical and practical methods of exploring and developing data quality dimensions, frameworks and

classifications. For example, Shanks and Darke (1998) used a theoretical perspective to identify and define data quality dimensions. They assert that within an information system, data quality could be influenced at four levels. The first is the *Syntactic level*, concerned with the structure of the data, related to *consistency* in data and usually relevant to data warehouses, as data is collected and stored taken from various sources and consistency is noted as a major data quality issue (Gardyn 1997). Second is the *Semantic level*, concerned with the meaning of data and having goals of *completeness* and *accuracy*. Third is the *Empirical level*, which concerns the usage of data and has goals of *usability* and *usefulness* (Kahn et al. 1997). Fourthly and finally, the *Social level* is concerned with social level activities, which concerns how data is interpreted and understood within various organisations.

The differences in interpretations and understanding can have an impact on data quality. According to Shanks and Darke (1998), the social level concerns are important within a data warehouse context when data is exchanged between data providers with various social and cultural differences. Shanks and Darke's work was extended by Shanks and Corbitt (1999) where they give greater focus to the social and cultural aspects of data quality. They also argue that meaning of data is a social construction rather than an objective reality and therefore, social and cultural aspects significantly influence how the data is created and interpreted. As a data warehouse integrates data that might be taken from various countries, where the data is created and interpreted based on their own social and cultural norms, social level data quality issues can emerge. In the context of DCM, it

is used in various countries across many continents. While a data warehouse can facilitate the technical integration of

internationally collected DCM data for richer and deeper analysis, the data might reflect each country's social and cultural norms, which can create data quality issues. Data representation within systems may be different in different countries. For example, American systems use a different format of writing dates from that of many other countries. This can influence how the 'date of mapping', as an important part of DCM data, is presented in a system which will source the data warehouse. Such data quality issues are common in systems sharing data across countries (Kossarova et al. 2015). To deal with these types of issues, as mentioned previously, the DCM international database was suggested as a common system for collecting data from mappers and organisations from various countries (Khalid 2010). This system provides a common interface and application for all types of users to input their data in a standard format.

While the above-mentioned issue has been dealt with by proposing the DCM international database to collect data from various sources in a standard format (Khalid 2010), the literature also suggests some other solutions to deal with social level data quality issues. For example, Shanks and Corbitt (1999: 792) propose three strategies for improving social level data quality and suggest considering these strategies while developing a data warehouse: viewpoint analysis, conflict analysis and cultural immersion. A viewpoint analysis is the process of "identifying, understanding and representing different stakeholders' viewpoints". Conflict analysis is a technique to "encourage groups of stakeholders with conflicting interpretations of data to discuss their differences and develop mutual

understanding of each other's position". Cultural immersion "involves becoming part of a different culture and

requires considerable time and effort". In order to improve data quality from a social perspective, Shanks and Corbitt suggest that it is necessary for the analyst to develop a deeper understanding of another point of view and understand different interpretations of data and origins of bias. Therefore, user involvement to support analysts in identifying their views and perceptions about the use of data is considered significant for a data warehouse (Teixeira et al. 2012).

While Shanks and Darke's (1998) and Shanks and Corbitt's (1999) quality dimensions were driven by a theory and provided a wider perspective on data quality, and Shanks and Corbitt's (1999) quality dimensions encourage understanding of social quality dimensions, both of their approaches lack consideration of data users' views regarding their perception of the quality of data.

Wang and Strong (1996) employed an empirical approach and reported on the identification and definition of data quality dimensions that have been identified based on data users' views. They used a two-survey method to collect empirical data from experienced data users, who used data to make decisions in their professional and academic life, recruited from US-based industries and universities. They explored the experienced users' views regarding their perception of data quality based on their subjective assessment of the characteristics that could fit for their own tasks of using data. Based on these data users' views, Wang and Strong (1996: 6) conceptualised various aspects of data quality and defined it as "fit for use by data consumers". They identified data quality dimensions based on attributes

of *accuracy, timeliness, reliability, currency, completeness, relevance, accessibility, interpretability* and

precision. They organised the identified dimensions into four categories called intrinsic, contextual, accessibility and representational. The intrinsic category contains dimensions that define the quality of data content regardless of any context within which it will be used and the system within which it will be stored. This means that any data obtained from any information system for any purpose should be not just accurate (accuracy) and objective (objectivity) but it should be believable (believability) and come from reputable sources (reputation).

The contextual category contains dimensions that define the quality of data within a specific context. For example, the primary use of data holds different criteria of quality than the secondary use (Gardyn 2009). Wang and Strong (1996) assert that the contextual data quality dimensions can be set based on the specific use of the data. Therefore, they maintain that data should be relevant to the task/purpose of using data (relevancy), timely (timeliness), and should be of an appropriate amount to complete the specific purpose (appropriate amount of data). The accessibility and representational categories include dimensions that define quality in relation to how easy and secure it is to access systems (security access) and data (usability) as well as how easy it is to understand (ease of understanding) and interpret (interpretability) data.

Wang and Strong's framework provides a comprehensive list of important data quality dimensions and has therefore been used for identifying, defining, assessing and comparing quality issues within various fields, including healthcare and data warehouses. For example, using Wang and Strong's (1996) data quality framework, Giannoccaro and Shanks (1999) carried out a

study into understanding the relationship between different stakeholder types in the data warehouse environment and categories of data quality dimensions. They recruited data warehouse stakeholders to identify user's task specific data quality dimensions. They defined the stakeholders as, users involved in creating data (data producers), using data (data users), maintaining data (data custodians) and managing data (data managers) within a data warehouse environment. Using Wang and Strong's framework, they proposed relationship instances between stakeholder groups and related data quality dimensions. They used a case study of a large transport company's data warehouse to examine the validity of the relationships. Based on this case study, Giannoccaro and Shanks (1999) also explained the quality requirements and concerns from the stakeholder groups. According to their study findings, data producers were mostly concerned with accuracy and believability of the data within the warehouse and data users were concerned with accessing relevant data that was consistent and timely to fulfil their needs of data use for a specific task. Overall, Giannoccaro and Shanks (1999) concluded that all data warehouse stakeholders hold different views regarding the quality of data, based on their roles within the warehouse development.

Kumar and Thareja (2013) concurred with Giannoccaro and Shanks that stakeholders involved in a data warehouse project could have various quality issues and might have different views on quality of data. They recommend that a warehouse development project should consider all stakeholders' quality issues and problems and understand what data quality aspects are important to each. In Wang and Strong's (1996) study, in order to define the

quality dimensions and demonstrate the applicability of the data quality framework,

the data users were chosen from an environment where a data warehouse was being used. This might be the reason why the data users' views were mostly related to the accessibility, usability, consistency and completeness aspects of data quality, rather than to trustworthiness or believability of data. The issues of trustworthiness or believability can emerge when data users are not part of the existing data warehouse environment and do not hold any stake in the warehouse project (Schaefer et al. 2011). Believability of data are important issues where data is collected from various sources, outside the organisational control and without checks regarding who is providing the data and how the data was collected, managed and used before its storage within the warehouse. This data quality dimension is of concern to warehouses which store archived data or provide data for research purposes (UK Biobank 2012). In this context, studies have used the term 'provenance of data', which means providing information about what, when, why, where and who collected the data before its storage within the warehouse (Simmhan et al. 2005). Based on this information, the user can assess the quality of data. For example, the UK Biobank data resource provides provenance information to the prospective users about the origin of the data stored within the resource and the methods of data collection, so that data quality can be judged by the data user in relation to the particular research question being addressed (UK Biobank 2012).

The above literature highlights that while various methods have been used for defining data quality, including user-driven approaches, there is, as yet, no consensus on what could be seen as a set of complete data quality dimensions or agreed definitions of these (Wand and Wang 1996; Weiskopf

and Weng 2012). The literature, however, suggests that the identification and definition

of data quality dimensions depends on various data related aspects. These include the context within which the data is used, the intrinsic or inherent nature of the data, the function of the system within which the data is managed for various uses and social aspects of data. While all these data related aspects are related to a data warehouse, the lack of agreement on data quality dimensions and their association with various data related aspects can be seen within both data and warehouse related literature in healthcare, as reviewed next.

3.6.2. Data quality requirements in data warehouses

Data quality plays a significant role and is a key factor in the success of data warehousing (Giannoccaro and Shanks 1999; Kimball and Ross 2002; Verma et al. 2014). Good quality data ensures users' trust in the data warehouse system, making it more usable and acceptable (Kumar and Thareja 2013). Poor quality data within an organisation can have significant social and economic impacts (Wang and Strong 1996; Redman 1998). Therefore, data quality is identified as a major issue in data warehouse literature. It is usually associated with technical processes of the system, the intrinsic nature of the data (e.g. healthcare data) and secondary use of data (Botsis et al. 2010; Weiskopf et al. 2013).

In addition to data quality challenges, the literature also highlights the organisational challenges, which are related to the preparedness of organisations to implement data warehouses that require resources, training and continuous maintenance (del Hoyo-Barbolla and Lees 2002). The aim of the current study was to focus on data related requirements, issues and challenges within a data warehouse. Therefore, this section will only examine

the data quality challenges that can emerge at a technical or system level, in relation to the nature of healthcare data and the context within which it is used. While doing so, I will also identify the potential quality issues with DCM data and how some of these were solved in my previous study and what yet need to be explored.

In relation to technical processes, data quality issues can emerge at various stages of warehousing. The most reported aspects are related to business analysis (Mohammed and Talab 2014) and data acquisition and integration within the warehouse (Singh and Singh 2010). The first aspect is related to business analysis, which is the identification of a business purpose, its problems and solutions (Hass et al. 2007). In the context of data warehousing, business analysis refers to identifying user requirements in order to understand the purpose, scope, problems, issues and solutions related to its design (Mohammed and Talab 2014). According to Gray (2004), a data warehouse within healthcare does not achieve its objectives unless its scope is determined. Failure to identify clear quality requirements to set the scope of a data warehouse is identified in the literature as one of the major challenges (Mohammed and Talab 2014). In order to deal with this quality problem, it is therefore important to set the scope for a DCM data warehouse based on user-driven requirements.

The second aspect is related to the quality issues that can emerge at a data source level during the data acquisition process. One of the reasons for these quality issues is incomplete, inconsistent and inaccurate data within the source systems. Another reason is related to the variations in the type,

amount and format of the data within data sources. Both of these problems can create data

availability and incompatibility issues (Wang and Strong 1996). Data availability issues are associated with the system's inability to collect, store and provide the required data within the warehouse. Incompatibility of data within a warehouse is associated with data that is not relevant, consistent, accurate or complete in regard to the task at hand (Wang and Strong 1996). In order to deal with data availability and data incompatibility issues, the need has been emphasised for identifying good quality data sources for obtaining data for a data warehouse, hence suggesting dealing with data quality issues at data-source level (Singh and Singh 2010).

In order to deal with the DCM data quality issues at the source level, within the two-step DCM data warehousing structure, the international database was proposed as the main data source for the data warehouse. This database system was proposed to be designed to deal with data quality issues at both data structure and presentation levels. For example, the database structure does not allow the entry of any wrong or incomplete combination of the codes within the system. While some of the issues have been considered within the DCM international database, the availability and incompatibility issues are assessed according to users' data requirements within the warehouse. This therefore requires identifying users' views in terms of their potential requirements for a data warehouse. These requirements then need assessing in terms of what data needs to go into the warehouse and what sources can provide the required data.

The third main aspect is that of quality issues at a data integration level. A data warehouse stores data that is collected from various sources. These

sources usually store data in diverse formats and locations. Consolidation of diverse

and fragmented data into a unified view within a data warehouse to facilitate secondary analysis can create quality issues at a data integration level (Mohammed and Talab 2014). Therefore, it is emphasised that use should be made of effective tools and solutions to deal with data quality issues during the data integration process within the warehouse (Takecian et al. 2013). A DCM data warehouse could potentially be collecting data taken from various sources, which will require taking quality measures at integration levels.

While there are a number of studies reporting the technical processes or system related data quality issues within a data warehouse (Takecian et al. 2013; Mohammed and Talab 2014; Verma et al. 2014), the quality issues related to the intrinsic nature of healthcare data are also highlighted in the literature. By nature, healthcare data is considered complex, heterogeneous and fragmented (Mohammed and Talab 2014): complex as it has varying definitions and medical standards; heterogeneous, as it is stored and captured in various formats (qualitative and quantitative data) and also stored in diverse formats (paper-based and electronic); and fragmented, as healthcare data is stored physically in various locations by various care providers.

Based on these characteristics, a number of studies report data quality issues and challenges while using healthcare data for secondary uses and regarding its warehousing for this purpose (Botsis et al. 2010). Botsis et al. (2010) reported data quality issues during their study of survival analysis of pancreatic cancer patients. They used a clinical data warehouse, which stored Electronic Health Records (EHR) data taken from integrated

healthcare information systems and extracted data using an extensive three-step process. They reported data incompleteness as a major quality issue, followed

by inaccuracy and inconsistency. They associate data quality issues with poor documentation, information fragmentation and a lack of contextual and detailed information in structured disease diagnosis. While acknowledging that there are variations in the semantic representation of data within various sources of EHR data, they also proposed a number of solutions for improving the quality of data. These include the use of standard content or standard common data elements within EHR and clinical registries for specific diseases with a pre-defined data format in order to integrate with EHR data.

Similarly, Ancker et al. (2011) reported data quality issues while analysing data availability for secondary uses from a web-based project management system of EHR implementation within New York County, US. While they were not specifically dealing with health data, they acknowledged that their identified quality issues were similar to the clinical data quality issues for secondary uses as identified in other studies, including Botsis et al. (2010). For example, incompleteness, inaccuracy and inconsistency were the main identified issues. They assert that these data quality issues emerge when data is collected solely for internal or primary uses, where data is ensured to be of good quality for local use. They further make a point that if potential secondary use of data is pre-determined, primary data could also be collected and stored to meet the additional criteria of this secondary use. Further, flexibility in systems allows capture of the same data in different formats, variability in documentation and recording of data and variability in standardisation of data. In order to make data available for secondary uses, they recommend carefully documenting the details of the data, including contextual data, ensuring consistent data definitions, the

promotion of uniform standardisation and training for those who collect and enter data into systems.

DCM data is rich but complex in nature. For example, DCM contains four types of codes, meaning that careful consideration is required when choosing the right code for the right situation. During mapping, a number of rules need to be considered while assigning the relevant code to a specific situation. Further, there are some specific code combinations that are not permitted within DCM coding rules which need to be considered to avoid data quality issues. It could therefore be argued that DCM data is intrinsically complex and that data quality issues can emerge while collecting data during observations as well as recording it in a data management system. While mappers are considered responsible for conducting mapping as effectively as possible (Bradford Dementia Group 2014), the data recording system should have an inbuilt functionality to deal with potential quality issues regarding DCM data. As discussed earlier, the DCM international database, the proposed main data source for a DCM data warehouse, deals with these issues by providing a DCM specific interface and a database structure, which will allow only the right codes in the right combinations to be entered by the mappers.

In the context of healthcare, Sandra and Garmon (2007) reported issues related to the secondary analysis of data. These are as follows: difficulty in locating required data, incompatibility of data for primary and secondary research objectives, and data quality. Sandra and Garmon (2007) recommend that, in order to assess the completeness and accuracy of data for secondary uses, the following information should be available alongside

the data set: supporting documentation such as a codebook, summary reports, research

proposals and published studies. Such additional information, alongside the original data set, is called metadata which means data about data (Deelman et al. 2010). A number of real life projects, including UK Biobank, provide detailed metadata alongside data content to ensure the data users have access to all the information that can inform them about the quality of data content stored within the system. In the context of DCM data and its secondary use, there is a lack of any information about metadata requirements and the type of data that needs to be stored as metadata.

In the context of a data warehouse, quality issues are also associated with the secondary use of data, (Gardyn 1997). The literature provides a number of studies, which identify and explain data quality related issues within healthcare, particularly when the aim is to use data for research purposes. Weiskopf and Weng (2012) conducted a review of studies which reported data quality issues and challenges while using EHR data for secondary analysis, particularly for research purposes. During their review, they focused on data quality dimensions and assessment methodologies that were reported within peer-reviewed literature and found 95 studies meeting their criteria. They identified that within the context of secondary use of healthcare data, the most common reported quality issues are as follows, *completeness, correctness, concordance, plausibility* and *currency*.

Completeness of data assures that correct details of a patient are presented within her/his healthcare record. Weiskopf and Weng (2012) assert that while completeness is the most commonly assessed dimension within the reviewed studies, the authors of these studies use diverse ways of assessing

it, which reflected their various perceptions and understandings of the term.

For

example, some attributed the completeness of data to the presence or not of various expected or required data elements (Pearson et al. 1996). On the other hand, some attributed the completeness of data to its availability for the task at hand. For example, if a researcher establishes that the EHR data is complete enough for a specific purpose, the data is considered complete (Linder et al. 2009; Botsis et al. 2010).

However, Weiskopf and Weng (2012) add that many of their reviewed studies were using different methods of assessing the quality of data. For example, the studies were using another source of data as the gold standard to assess EHR data completeness (Noel et al. 2010), looking at agreement between elements from the same source (Linder et al. 2009) or were looking at agreement between the paper records and the EHR (Scobie et al. 1995). Based on their review, Weiskopf and Weng (2012) summarise that completeness of EHR data for secondary uses is assessed on various quality attributes such as, data availability, missing data, validity and sensitivity of data. As a result of this finding, Weiskopf et al. (2013) conducted a study to demonstrate that the concept of completeness in EHR is contextual, which means that there exist multiple interpretations of what is complete data within EHR depending on data needs and on its specific use, which will dictate how completeness is conceptualised. They further asserted that multiple interpretations of EHR completeness could lead data users to find different data sets and this could lead to different results. Therefore, they warned researchers and clinicians, who wish to re-use EHR data, to be mindful of what definition they use for completeness and to be transparent

while presenting their findings about what constitutes completeness of EHR data.

In Weiskopf and Weng's review, the second most assessed quality dimension was the correctness of data. Correctness of data ensures that whatever is recorded about a patient is true. Concordance of data was the third most assessed quality dimension. This checks that a patient's data is consistent between various sources or across various sources of EHR. Plausibility of data means that a data attribute within an EHR measures what it is supposed to measure. For plausibility, the assessment includes the degree to which the available data is trustworthy. Alongside trustworthiness, the terms *validity* and *integrity* were also used to explain the plausibility of EHR data. Trustworthy data is an important requirement in studies reporting the issue of data quality within a warehouse (Buneman et al. 2001; Hartig and Zhao 2009). Currency of the data ensures that the patient record is up-to-date and timely.

The DCM data warehouse will be storing data for secondary uses and therefore it is important to explore the related issues. However, currently there is a lack of any evidence about quality issues with the secondary use of DCM data. Further, various functions of the warehouse can have different data quality criteria and therefore different quality dimensions. For example, while data for benchmarking should be comparable, accurate and suitable, for research purposes the data should be complete, accurate, consistent and relevant (Weiskopf et al. 2012). Therefore, by drawing from the above literature, in the context of a DCM data warehouse, it is suggested that there is a need to first establish the use and then to identify the related data quality issues.

In summary, the technical processes (such as data integration) and the intrinsic nature and specific use of the data collectively influence data quality

within healthcare data warehouses. Further, the identification and definition of various data quality dimensions vary within the literature and are mostly determined by the nature and use of data. The literature suggests that there is a need to understand the nature of data and the need to establish how users perceive the quality of data in their tasks to identify the relevant dimensions or requirements of data quality. These quality dimensions are effective in solving the data quality problems (Tejay et al. 2006). The present study will be identifying the main uses of a data warehouse for DCM data, based on which the relevant data quality issues can be identified.

Next, I will explore data security within data warehousing and the type of requirements that can emerge for such a purpose.

3.6.3. Data security requirements in data warehouses

Healthcare data is considered sensitive, as it includes patients' personal and medical details, including information about their lifestyle choices, which if combined can put patients at the risk of privacy threats (Lamas et al. 2015). Dealing with healthcare data therefore requires additional security measures (Faria and Cordeiro 2014). While primary use of healthcare data is concerned with patients' own care and treatment, where practitioners or care providers are obliged to protect patients' privacy, secondary use of the same data, particularly for research purposes, can raise additional data-protection and privacy concerns that elevate further ethical, legal and social requirements (Wiesenauer et al. 2012; Lamas et al. 2015). These requirements need to be identified and met to ensure data security within healthcare data warehouses (Wiesenauer et al. 2012). The ethical, legal and social requirements are discussed in detail later in this section.

Data security ensures that patients' healthcare data is handled with confidence, integrity and privacy (Kaplan 2014). Within the warehouses, data security is achieved by implementing technical solutions that adhere to data protection and privacy laws and also meet users' data-access requirements (Aleem et al. 2015). A number of individuals/organisations are responsible for ensuring data security within the warehouses. For example, data providers (e.g. health and social-care organisations) ensure that patients' data is legally secure so that it can be shared with others. Data controllers/data custodians/data processors are those organisations or individuals who collect data from the data providers and ensure its lawful collection and processing within the warehouse and its further dissemination to potential data users. Data users are those who receive data from the warehouse, for example, researchers. Data users ensure that data is used only for the purpose(s) for which it was acquired from the data warehouse by meeting specific ethical requirements. All these individuals/organisations together ensure that healthcare data is collected, processed, disseminated and used within appropriate ethical and legal grounds.

The literature indicated a number of challenges that must be met in order to achieve data security within healthcare data warehouses. Lamas et al. (2015) argue that current ethical and legal frameworks deal with data that is collected for primary uses, which is concerned with privacy and security issues related to patients' data that is collected, processed and used for their own care treatment. However, as a data warehouse stores data taken from various sources Lamas et al. (2015) criticise the incompatibility of existing

ethical and legal frameworks for dealing with issues raised during warehousing data for

secondary uses, particularly for research purposes. Emphasising the need for specific ethical and legal considerations for warehousing the healthcare data for secondary uses, Lamas et al. (2015) identify the following main areas that need to be considered:

- Patients' rights (information and consent)
- Care providers' rights (agreement on data sharing)
- Access issues (right of access)
- Optimisation of data confidentiality (data-sharing)
- Solidarity and common good (shared infrastructure)
- Transparency
- Trust

Further, within the warehouses, like any other information system, secure data access is also a major security requirement. For example, the system must provide access only to authorised people and deny any unauthorised access for using or modifying the data. It must also provide the right data to the right users at the right time and keep a record of activities performed by its users (Aleem et al. 2015). While these basic security requirements do not just apply to the warehouses but to any other electronic information system, a warehouse is more prone to data security threats as it stores data taken from various sources and can be more attractive to hackers (Stolba et al. 2006). To deal with these requirements, the need for a robust security infrastructure, encryption technologies and security governance process within the warehouse is emphasised (Aleem et al. 2015). Meeting these requirements means ensuring that healthcare data is made available in a trustworthy way (De Moor et al. 2015).

Any technical solution for securing healthcare data within data warehouses is only useful if the data being secured meets ethical, legal, social and technical requirements. While technical requirements ensure that warehouses are designed and developed to keep healthcare data access secure within the warehouse, Kaplan (2014) asserts that the ethical, legal and social requirements are related mainly to ensuring the privacy of patients through protecting their healthcare data in accordance with data-protection and privacy laws. Next, I will examine the ethical, legal and social requirements in detail with the specific aim of explaining how the areas of consent, transparency and anonymisation are dealt with in healthcare data warehouses.

3.6.3.1. Ethical requirements

The ethical requirements ensure that patients' privacy is respected, which means that their healthcare data, particularly their personal data, is handled according to their own wishes. The patients' wishes regarding secondary uses of their healthcare data is ensured through adherence to a consent procedure. According to the EU Data Protection Directive 95/46/CE, Article 2(h) (EU Directive 1995), consent is "any freely given specific and informed indication of his wishes by which the data subject signifies his agreement to personal data related to him being processed". Personal data is referred to as "any information relating to an identified or identifiable natural person (data subject⁸); an identifiable person is one who can be identified, directly or indirectly" (Article 2(a) (EU Directive 1995). Data processing refers to any operations carried out on the data, including recording information, storage,

⁸ The person whose personal data is collected, held or processed.

alteration of records and usage and disclosure (European Medical Information Framework 2013).

It is required by law to seek a patient's consent about secondary uses of their healthcare data, particularly their personal data. However, if a patient's healthcare data is effectively de-identified⁹, the user is exempted from obtaining the patient's consent regarding potential use of their data for secondary purposes (Information Commissioner's Office 2015). A number of studies emphasise the importance and the processes of consent management within healthcare data warehouses.

One example is a European data-sharing project, the Electronic Health Record for Clinical Research (EHR4CR), which uses data warehouse technology to integrate and disseminate various Electronic Health Record (EHR) data for secondary research purposes (De Moor et al. 2015). Emphasising the importance of securing patients' privacy, the EHR4CR maintains that only patient-consented data (where the patient has given informed consent about the reuse of their healthcare data) or de-identified and aggregated¹⁰ data is collected from the data providers' locations. Similarly, UK Biobank, a major data-integration and data-sharing initiative for clinical and non-clinical research within the UK, ensures that only consented patient data or de-identified data is disseminated for research purposes (UK

⁹ The process of rendering data anonymisation and pseudonymisation, where anonymisation is the 'process of removing all elements allowing the identification of an individual person (i.e., of rendering data anonymous)' and pseudonymisation is the 'process of removing all elements allowing the identification of an individual person, except the key(s) allowing linking the data to the person. Such key(s) shall be generated randomly' (Chester 2011).

¹⁰ Data of several patients that have been combined to show general trends and values (Code of Practice 2014).

Biobank 2012). While it is necessary to gain patients' consent for the secondary use of their

data, a DCM data warehouse might face additional consent-related issues. For example, people with dementia may lack the capacity to give informed consent, or their capacity may change over time. This requires establishing procedures that ensure an effective process of consent for the data warehouse. However, if a DCM data warehouse stores anonymised data, there is no ethical obligation to gain consent from people with dementia regarding their secondary use of DCM data. While formal consent is not required, it is suggested that patients should be aware of the potential future uses of their data (Information Commissioner's Office 2015). The visible tagging of DCM data within the warehouse as identified or anonymised depends on user requirements for accessing data within the warehouse and should be part of the design process.

Regarding secondary use of healthcare data, alongside the need to obtain consent, being transparent in the use of the data for various purposes is also a pressing ethical issue (Lamas et al. 2015). Transparency asserts that patients are informed accurately and in a timely way about the secondary use of their healthcare data. On ethical and moral grounds, the data providers are required to be transparent to the patients about any further use of their data (both identifiable and de-identified data) beyond their individual healthcare provision. For this purpose, the care providers need to ensure that the patients are given accurate and timely information about how, with whom and for what purposes their data will be shared (Health and Social Care Information Centre 2013). Negligence in this regard can raise human-rights issues as well as public anger (Presser et al. 2015).

A recent example comes from the UK's largest data-integration project, care.data. Care.data aims to collect and integrate patients' health and social care data from various care-providing organisations across England to support secondary use such as clinical and commercial research. The Health and Social Care Information Centre (HSCIC), also called a 'safe haven' for patients' healthcare data, is responsible for data collection from the care providers. The care providers are, however, responsible for ensuring that patients are provided with information about their opt-out option (a form of consent chosen for this project). During their first attempt at data collection in 2012, HSCIC failed to manage the appropriate and informed consent procedure (Presser et al. 2015). While the chosen consent method in itself was a major issue¹¹, it was also poorly communicated to the patients. For example, the process was criticised for using an inappropriate method of communication with patients, a lack of management in providing a complete or one-stop-shop information point to patients and providing vague and incomplete information about the consent and withdrawal process (Presser et al. 2015). Due to these reasons, such a major data-sharing project became unacceptable to the public (patients) and therefore failed in its first attempt at collecting data.

Appropriate consent and transparency is also ensured by those who are in charge of collecting, processing and disseminating data from the warehouses for secondary uses. For example, in the case of a research data

¹¹ Patients were coerced to allow the sharing of their healthcare data by using the opt-out consent method, which is also identified as being inflexible. For example, if a patient was unable to register for the opt-out option and their data was given to the 'safe haven', he/she could not opt out retrospectively. Similarly, if parents did not opt out on behalf of their children, once they are grown up and able to make their own decisions, those children cannot opt out (Presser et al. 2015).

warehouse, UK Biobank (2012) acts as data controllers (data collectors) and data processors and is responsible for ensuring that data for the purposes of secondary uses is disseminated in a secure environment where a patient's personal data is dealt with according to data-protection and privacy laws. While informing patients and gaining their consent is comparatively easy within prospective studies or data (as data providers are aware of the specific use of data and can inform patients about it), the same process becomes challenging when retrospective data is used for secondary purposes (Lamas et al. 2015). While establishing this as a major issue in warehousing retrospective healthcare data, Elger et al. (2010) assert that such an issue can emerge in two situations. The first situation is when retrospective data is taken from studies where patients consented only to a specific secondary use of data, and a second situation is when patients' routinely collected data does not cover their consent for any potential secondary use of data. To remedy the first situation, UK Biobank (2012) explicitly obtains consent for using data for any 'general research' purposes. This type of consent allows the data controllers to reuse patients' data for any future unplanned research purposes. However, in the second situation, it is the care providers' responsibility to ensure that appropriate consent is sought from patients regarding secondary use of their healthcare records. While meeting ethical requirements is imperative for securing data within the warehouse, it is important that these requirements are met in accordance with data-protection and privacy rules, thereby leading to legal requirements, as discussed next.

3.6.3.2. Legal requirements

The legal requirements ensure that patients' healthcare data is handled in accordance with data-protection and privacy legislation. The Data Protection Act 1998 (DPA) (Department of Health 2011) was introduced in the UK in response to the EU Data Protection Directive 1995 (EU Directive 1995) which, in its Article 1(1), states that EU member-states shall protect fundamental rights and freedoms of national persons and their right to privacy when processing their personal data. In line with this directive, in regard to personal information, one of the fundamental principles of the DPA 1998 is to use 'the minimum personal data to satisfy a purpose and to strip out information relating to a data subject that is not necessary for the particular processing being undertaken'.

For the purpose of secondary use of healthcare data, the recommendation therefore is to, wherever possible, use anonymised and pseudonymised healthcare data, for which the NHS has specific rules to follow (Chester 2011), which ensure that a patient's identity is either erased (anonymisation) or hidden (pseudonymisation). Effectively anonymised healthcare data used for any kind of further processing is exempt from the DPA 1998 legislation. However, if the use of identifiable data is necessary, the patient's consent is required. According to Section 60 of the Health and Social Care Act 2001 (The National Archives 2001), the Secretary of State for Health can permit the use of patients' healthcare data, including their identifiable information, without their consent in England and Wales for certain reasons. For example, consent is not required for secondary use of data, such as where medical research

conducted for the wider public and the patient's interests, and where gaining consent is nearly impracticable.

The data-protection challenge escalates when data warehouses are required to store international data. Elger et al. (2010) point to the lack of conformity in the implementation of data-protection and privacy legislation among various countries, specifically within the EU, where the EU Data Protection Directive was meant to provide harmonisation. As national legislation and its implementation in terms of privacy and data protection vary across EU countries (Bahr and Schlunder 2015), Elger et al. (2010) argue that combining data from various institutions and countries for the purpose of secondary use, such as research, has become complex and challenging, as a large amount of varied legislation has to be taken into account.

In order to deal with this issue, a major initiative was taken in 2012, during a European summit, when various stakeholders (delegates representing national governments, academics, patient groups, researchers, industry experts and the European Commission) met to discuss the need for a united framework to allow the trustworthy reuse of healthcare data within the EU. The result of this summit was agreement on developing a system, which ensures that the healthcare data is 'fully' regulated and the patients are 'fully informed' (Geissbuhler et al. 2013). To enable the development of such a system, a new proposal for a regulation of the European parliament and the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data is in process to replace the European Directive 95/46/EC. Lamas and colleagues (2015) compare the in-force EU Directive

and the proposed new regulation and maintain that the new proposal is aimed at allowing the secondary use of healthcare data for research purposes, which will be binding for all EU countries (Lamas et al. 2015). However, Lamas and colleagues (2015) criticise that the new regulation will still be unable to clarify some of the privacy issues, which can emerge when healthcare data is shared across nations (Lamas et al. 2015). For example, the new regulation will yet be unable to define a common set of guidelines for de-identification of healthcare data for research purposes.

There is a growing trend of supporting the secondary use of healthcare data for medical research purposes. One of these examples is Innovative Medical Initiatives (IMI), Europe's biggest healthcare data-sharing platform for research purposes. The IMI-funded projects require a common set of data-protection and privacy rules and guidelines that could be used for research purposes (De Moor et al. 2015). Considering the lack of existing specific legal frameworks, a 'Code of Practice' on the secondary use of medical data for research purposes has been developed by the IMI, which addresses a number of practical issues related to consent for prospective data collection, and dissemination of new research findings (Bahr and Schlunder 2015).

However, data security becomes more complex when countries outside the EU may also be involved in data sharing, some of which may have variable local, legal and ethical contexts under which the data was originally collected and stored. In the context of a future DCM data warehouse, the complexity of combining data from various countries can escalate, as DCM is used across the globe. This includes EU countries, as well as other countries that are outside EU laws, such as China, Australia and the US. Acquiring DCM data

from these countries within a future DCM data warehouse would require meeting the non-EU countries' local legal and ethical regulations on data security.

Compliance with the laws and regulations depends on the existence of security measures taken within the data warehouses. For this purpose, studies have suggested various technical methods and conceptual frameworks for dealing with data security. For example, Stolba and colleagues (2006) propose a conceptual model of a federated data warehouse for a health insurance company, where the aim is to integrate data taken from various data sources to allow data mining. Due to the high confidentiality of healthcare data and various privacy policies of care-providing organisations, they propose a three-phase data security process. This process includes depersonalisation (anonymisation), pseudonymisation and federation. This means that the data taken from various data sources are first anonymised and pseudonymised based on user requirements and then integrated into a data warehouse. Such types of federated systems are called virtual systems or federated data warehouses (Stolba 2007). Once the data is conceptually integrated in such as system, a role-based multi-level security mechanism is applied to the data before it is disseminated for data mining. This mechanism ensures that only authorised and appropriate users have access to the data. Whilst Stolba et al. (2006) and many others (Zhang et al. 2005; Lo Iacono 2007) focus on proposing secure data architecture for a data warehouse, Elger and colleagues (2010) also emphasise the need to anonymise free-text data within the warehouse, which is mostly part of a

patient's healthcare data. Uzner and colleagues (2007) provide a number of technical methods for anonymising

free-text data. This suggests that it is feasible to build technical solutions into a warehouse that can help to meet the highest security levels required for warehousing healthcare data. Such a technical solution might be required for DCM data within a data warehouse as DCM produces both quantitative and qualitative data.

3.6.3.3. Social requirements

Social requirements ensure that a patient's healthcare data is shared with others in accordance with the ethical and legal aspects of dealing with data. For this purpose, both the data collector¹² and the data provider¹³ are obliged to comply with data-protection and privacy regulations when sharing patients' data with others. For example, it is the data provider's responsibility to ensure that their patients' personal data does not leave the premises without the patients' appropriate consent or until the data is fully de-identified. Data providers and data collectors are also legally required to develop data-use and data-sharing agreements, which should also be compliant with the rules and regulations surrounding the preservation of patients' privacy and the protection of their data (Information commissioner's Office 2016). Further, data controllers are also responsible for providing a shared infrastructure for ethical and legal data sharing between data providers and data users.

In summary, because of the sensitive and confidential nature of healthcare data, the secondary use of this data from the warehouses needs to meet ethical, legal, social and technical requirements. These requirements, however, first need to be identified. The technical requirements are mostly

¹² Individuals and organisations who collect data within a data warehouse.

¹³ Individuals and organisations who are responsible for sharing their data with others.

focused on secure data access within the warehouse, where technical solutions need to be used to provide authorised access for the right user to the right data at the right time. The ethical, legal and social requirements set out appropriate solutions for securing patients' privacy and their data by complying with the relevant legislation. For this purpose, data can be de-identified within the warehouse and only aggregated de-identified information can be shared. De-identified data does not constitute 'personal data' and therefore is legally exempt from adhering to data-protection legislation. Further, appropriate data-sharing infrastructure needs to be in place to meet the ethical and legal requirements of healthcare data security. In the context of DCM data, consent issues will need to be managed by the data collector and given the many national and international individuals and organisations involved with DCM data collection and the complexities around consent in people with dementia this will be a challenging area to negotiate.

3.7. Summary of the chapter

This chapter has introduced the concept of data warehousing and its structural components. It commenced with outlining various definitions of a data warehouse and argued its uses for data integration, storage and dissemination for various secondary purposes. The chapter then underscored the significant differences between a data warehouse and a traditional database. It went on to discuss the role of data warehousing within healthcare, arguing that the warehousing of healthcare data is growing as awareness grows of the value of data for its historic and integrated use for a number of analytical purposes. The chapter also featured various

functions/uses of a data warehouse within healthcare. These included data mining, decision-making, research,

benchmarking and quality-improvement purposes across a range of healthcare areas.

The chapter continued by providing a critical review of my previous study, the only study to date that has proposed the concept of a data warehouse for DCM data to support its potential secondary use. The details of each component of the conceptual system were provided. In doing so, the chapter critically analysed the work to date and emphasised the need to involve users to identify their requirements that could inform various aspects of the warehouse design. This critical analysis led to highlighting the limitations of the previous study, where the need for a user-driven approach for designing a data warehouse was highlighted.

The chapter then focused on the challenges of warehousing healthcare data. It discussed in detail the two main reported challenges, data quality and security. A detailed insight was provided to explain the concept of data quality and the issues in defining the concept. However, it was argued that, while there is no consensus on the definition, the quality of data could be determined by understanding the context within which the data is used (e.g., secondary use of data within a data warehouse), the inherent nature of the data and the specific technical functions of the system within which the data will be stored. This chapter also emphasised the significance of users' views in terms of identifying data quality issues within a data warehouse. It further argued that trust and believability issues can potentially emerge in relation to the integrated data within the warehouse, which requires provenance information to be collected alongside the original data set within the

warehouse for users' data-quality assessment purposes. The chapter then went on examining various

studies to exemplify the data quality challenges that could emerge at a technical/system level in relation to the inherent nature of healthcare data and its use for secondary purposes.

The chapter then discussed data security as a challenge in warehousing healthcare data. It argued the importance of ethical, legal and social requirements for a data warehouse, which stores historic and integrated data for secondary uses, particularly for research purposes. The chapter examined in detail how issues of consent, transparency and anonymisation can potentially occur and how they can be dealt with in a data warehouse.

Overall, the chapter tried to argue that, while my previous study designed a data model for a DCM data warehouse, there is yet need to explore users' perspective regarding their potential use of data within the warehouse. Such information is important to design and develop the future system. A study is therefore required to explore the potential uses of DCM data within a future data warehouse and associated challenges and issues as requirements. For this purpose, users need to be involved to identify requirements. The rationale of why a user-driven approach is required for a DCM data warehouse is discussed in Chapter 4.

4. Requirement Analysis for a Data Warehouse

“The hard part of building systems is not building them; it’s knowing what to build; it’s in acquiring the necessary knowledge”

Armour (2000: 17)

4.1. Introduction

The aims of this chapter are to provide a rationale for a user-driven approach for gathering requirements for a future DCM data warehouse, and to argue the need for an appropriate methodology for this purpose. The chapter first explores the concept of requirements within the context of a data warehouse. It then highlights the significance of the role of the analyst who acts as an interpreter. The chapter further explores various steps and methods used for identifying requirements for general information systems. It examines a user-driven approach and continues by exploring why it is important to identify requirements from potential users’ perspective, particularly within the context of a data warehouse. It also examines various potential challenges that can emerge from a user-driven approach. The chapter then reviews research literature that reports the requirement analysis process for a data warehouse and demonstrates the use of various methodologies underpinning a user-driven approach. The chapter concludes by emphasising the need for a specific methodology for requirement analysis for a future DCM data warehouse.

4.2. Requirement analysis

This study aims to identify requirements for a future DCM data warehouse employing a user-driven approach. Within the context of a data warehouse,

requirements are considered as the information that defines what needs to be

provided by the system to meet the expectation of the user (Abai et al. 2013). Requirement analysis is the process of gathering such information from various sources using a choice of methods and then transforming it into the specifications to inform a useable design for the system (Maguire and Nigel 2002). In a data-warehouse development life cycle¹⁴, requirement analysis is the most important phase (Paim and de Castro 2003), as it determines the success of the system (Maguire and Nigel 2002; Paim and de Castro 2003; Nasiri et al. 2015). Yet, due to a number of challenges (as will be explained later), some either skip the requirement analysis phase (Paim and de Castro 2003) or do not pay sufficient attention to conducting the process effectively (Golfarelli 2010). Kimball et al. (2008) assert that requirement analysis influences almost every decision in a data warehouse. For example, the right requirements can lead to the right design and development of the warehouse. Giorgini and colleagues (2008) therefore warn that the failure of requirement analysis in terms of identifying poor and incomplete requirements can lead to unstable data warehouse design. The significance of conducting the requirement analysis process effectively, therefore, cannot be too highly emphasised for the success of a data warehouse.

The process of requirement analysis commences with identifying sources that can provide the relevant information. One of the main sources is the potential system users, as will be explained later in this chapter why this is

¹⁴ A data warehouse has a standard development cycle that involves stages as outlined: a pre-design/requirement analysis phase (feasibility or scoping and requirement gathering and specification), a design phase (database and application), a development phase, an implementation phase, a testing phase and a maintenance phase.

the case within a data warehouse. Only a sample of intended or potential users is approached for their requirements, as it is not feasible to gather requirements from all anticipated users. The chosen sample should be willing to take part, have enough knowledge of the new system or the type of data that needs warehousing, be aware of their requirements or feasibly be available to comment on the future system's functionality (Niès and Pelayo 2010).

Further, the requirement analysis continues with identifying and choosing the appropriate methods that are suitable for gathering specific information and assessing its quality as requirements for a specific system and ends with specifying them as proposed aspects of the new system (Kimball et al. 2008). The individual who identifies, analyses and presents these requirements in a form understandable by the system developer, is called an analyst (Dalal and Yadav 1992). In this section, I will first examine various types of requirements and the role of the analyst and then investigate the main aspects of a requirement analysis process by emphasising the existing methods and their usability in various situations. Further, I will examine the user-driven approach, as one of the preferred approaches for requirement analysis for a data warehouse and look at its significance and the challenges of gathering user requirements within the warehouse.

First, I begin by describing the terminologies used for different types of requirements and the role of the analyst in a requirement analysis process.

4.3. Types of requirements and terminology

As previously stated 'requirements' provide the information that describes the nature of a new system. In this context, the term 'system requirements' is used

to explain what is desired for the system to work effectively (Abai et al. 2013) according to the user expectations. This information includes goals of the system, business processes, data needs, usability and technical-design constraints and the behaviour of users (Byrd et al. 1992; Dalal and Yadav 1992; Pitts and Browne 2007). The requirements of a system can be defined as 'technical', 'business' (Gosain and Singh 2008) or 'social' (Lamas et al. 2015) depending on the aspects of the system they refer to. In the more technical literature, the terms 'functional' and 'non-functional' requirement are used in place of 'technical' and 'social' requirements. In general, functional requirements are concerned with the features that explain a system's technical functions (El Mohajir and Latrache 2012), for example, a system's ability to record, delete and update specific information. The non-functional requirements refer to the requirements that explain the behaviour of the system, for example, how fast, reliable, secure and efficient a system needs to be (El Mohajir and Latrache 2012). In the context of a data warehouse, 'functional requirements' are referred to identifying the type of data that needs to go into the warehouse and 'non-functional requirements' are referred to the information that explains how the data for the warehouse can be collected and used (El Mohajir and Latrache 2012).

The word 'requirements' is also associated with the sources that information has come from. For example, if the information about the new system is collected from users, the term 'user requirements' is employed. Browne and colleagues (2002) argue that despite the type of sources these are collected from, all requirements are gathered for developing systems that should be user-accepted. This is the reason why many studies employ 'user

requirements' as an umbrella term to refer to various types of gathered information for a system (Golfarelli 2010; Abai et al. 2013). While this study employs a user-driven approach (as will be explained in detail later) for gathering requirements for a DCM data warehouse, I will be using a range of terms for 'requirements' based on the type of information that these refer to. For example, I will use 'data requirements' when users express their data-related requirements and 'system requirements' when the aim is to explain the requirements related to a specific system or to aspects of a system.

The role of the analyst in identifying user requirements is important and is discussed next.

4.4. Role of the analyst

It has been argued that requirement gathering from users is a 'communication-rich' process (Zowghi and Coulin 2005) and a social activity (Amber et al. 2011). This means that the analyst conducts meetings with the user(s) to ask them about their needs, expectations and requirements for the new system, the system that is going to be developed or to replace existing solutions. During these meetings, the analyst plays an important role as an investigator/interviewer (Gallivan and Keil 2003). In order to make the meetings successful, Nies and Pelayo (2010) assert that the analyst needs to possess the knowledge, skills and personality suitable for dealing with specific types of users. In this regard, the literature highlights that, within the healthcare domain, one of the major challenges is to understand the specific medical vocabulary and terminologies of users (Niès and Pelayo 2010). Therefore, it is suggested that analysts have specific training to understand

users' specific terminology (Gallivan and Keil 2003) or that they involve experts from the field

who can translate or interpret user needs correctly into specific system requirements (Niès and Pelayo 2010).

Further, Darke and Shanks (1995) assert that user requirements usually emerge in a 'natural language' rather than in the language that describe the technical features of the system or portray the system requirements. Pace (2004) further adds that user requirements could emerge as their concerns and complaints about the existing systems as well as about their needs and expectations with the new system. This information needs to be translated into the system requirements to ensure that the developer could understand the user needs and develop a system that fulfilled user requirements. The role of an analyst is considered significant in these cases for understanding and translating user requirements into system requirements (Urquhart 2000). On this basis, it has also been argued that requirements are an analyst's interpretations of a users' needs, expectations and experiences (Gallivan and Keil 2003; Pace 2004). The process of interpreting user requirements into system requirements requires methods that endorse and acknowledge the interpretive process. The method that acknowledges the interpretative process and which is used within this study is explained in detail in Chapter 5.

Nies and Pelayo (2010) explain that users normally express their requirements in a superficial manner. For example, a user might express a requirement such as 'the system is not working' and then explain a few related issues. This means, as Seyff and Graf (2010) also advise, that the need is to understand the contextual information within which the user requirements emerge. The contextual information provides a means for

translating user needs into relevant requirements (Power and Moynihan 2003). In order to understand the

context within which the requirements emerge, the analyst needs to understand the entire work system. Further, Sutcliffe and colleagues (2006) maintain that cultural, emotional, social and material aspects can also influence user requirements and therefore, to elicit and interpret these requirements, it is important to understand these influential contexts. In this context, Nies and Pelayo (2010) assert that an analyst plays an important role in understanding and analysing this contextual information either by asking various related questions or by observing users during their work.

So far, it has been argued that users express their requirements in a non-technical language, which reflects their views, perceptions, expectations, concerns and needs regarding the new system, which need to be interpreted into requirements that can support further designing and development of a system. Further, it has also been argued that an analyst plays an important role in gathering, identifying and interpreting the user views, perceptions and needs into system requirements. In this study, I will take the role of an analyst for identifying requirements for a DCM data warehouse. My role as a researcher and its impact on various aspects, particularly on data collection and analysis, is discussed in Chapter 5.

Next, I will explain the three main aspects of the requirement analysis process – the gathering requirement and its analysis and presentation – by emphasising the types of methods used within each aspect.

4.5. Methods for gathering requirements

Requirement gathering is the first important aspect of a requirement analysis process (Abai et al. 2013). Different techniques are used for requirement

gathering depending on four main factors: the kind of information that is required, the availability of resources, the type of system that needs designing, and the organisational context within which the new system will be introduced. These techniques and methods are discussed next by highlighting their effectiveness in relation to the four factors.

Both technical (objective) and non-technical (subjective) techniques are used for gathering requirements (Galal-Edeen and Paul 1999). However, as the success of a data warehouse depends on user acceptance, subjective techniques are preferred for requirements gathering (Teixeira et al. 2012). Subjective techniques involve obtaining the users' views regarding their needs, expectations and requirements for the new system (Schaefer et al. 2011). In order to gather users' subjective requirements, Zowghi & Coulin (2005) argue that the methods for gathering requirements should not be derived from technical areas of information system research where objective and automated methods are used for requirement gathering. As a result of this belief, requirements gathering techniques are derived from disciplines of social sciences, organisational theory, knowledge engineering and practical experience (Zowghi & Coulin 2005). While considering subjective techniques for requirement gathering, Amber and colleagues (2011) broadly divide various methods into four main categories: conversational, observational, analytic and synthetic. Conversational methods can be interviews (structured and semi-structured), focus groups, Joint Application Development (JAD) (where the system designer, the system developer and the users all get together to identify the system's requirements through communications) and

storytelling (when users explain their experiences and requirements in a story

form). Conversational methods encourage communication between the analyst and the user.

Observational methods involve observing systems themselves and also users while they interact with systems (Amber et al. 2011). Analytic methods include documentation analysis/content analysis, laddering (a technique to find users' sub-conscious motives related to the new system), card sorting (a method where users organise topics into categories that are important and make sense to them) and repertory grid (a way of identifying how a user constructs her or his experiences). Within these methods, the main sources for requirement gathering are documents (where important information about the new system can be retrieved) and experts within the field (who can explain their views, experiences and motives). Synthetic methods, on the other hand, are based on a systematic combination of conversational, observational and analytic methods. Some examples are prototyping (when a method, either paper-based or a small electronic prototype, is introduced to the users to seek their opinion and identify further requirements), passive and interactive storyboarding and scenarios (Zhang 2007).

The methods within each category have their strengths and limitations and the choice of these methods is carefully made by analysing the situation within which the system is designed and implemented (Amber et al. 2011). For example, observation methods for requirement elicitation are considered time consuming and therefore are not usually preferable in situations where there is a tight schedule in which to complete the project in a specific period of time (Amber et al. 2011). Further, Martin and colleagues (2012) report on

the unsuitability of using a focus group or the JAD method due to the nature of

users' requirements. For example, in their study identifying user requirements for designing medical devices, their aim was to let users openly express their concerns regarding the new system. They assert that a focus group (where many users get together in a group conversation) could impede some users expressing their natural concerns with the new system. However, on the other hand, with good facilitation, a focus group method is considered a platform where different expression of ideas are encouraged by listening to each other, thus generating more ideas (Kitzinger 1995).

Similarly, the storytelling, prototyping, JAD, passive and interactive storyboarding and scenario (Mavin and Maiden 2003) methods are only useful when users are, to some extent, familiar with their requirements or have some vision of a new system. The most common and preferred method for subjective requirements gathering techniques are interviews (Davis et al. 2006). This method can also work in situations where users might not be familiar with their requirements, as it provides an opportunity for interacting with the users to identify, probe, verify and explore their experiences, needs and expectations, which can further be translated/interpreted as requirements for the new system. For this purpose, Zowghi and Coulin (2005) suggest designing semi-structured interviews rather than unstructured interviews, as the inquirer is then usually in control and can take the conversation in the right direction.

In summary, a review of the literature suggests that there are a number of methods and techniques available for requirements gathering, depending on the situational context (Abai et al. 2013), which are related to the required

information, users, systems and organisational contexts and to the available resources for the analyst. This suggests that there is no specific method for

requirements gathering for information systems. On the contrary, the choice of methods depend on a range of situations. This highlights the importance of assessing the specific situation for requirement gathering for a data warehouse and emphasises the need of a methodology that provides flexibility in making decisions about what, when and how to gather requirements.

4.6. Methods for analysing and presenting requirements

The requirements gathering process generates a large amount of information (user views) that needs to be managed and analysed in order to identify the coherent and mutually agreed requirements to be met for achieving an acceptable and usable system. There is a dearth of research on how user views can be systematically managed, analysed and interpreted to reach requirements that are meaningful, coherent and mutually agreed at a conceptual level. Most of the collected information from users is qualitative in nature (conversations). Browne and Ramesh (2002) warn that user conversations could produce a large amount of data, which could be overwhelming for the analyst to manage and to present in a coherent form. Browne and Ramesh (2002) argue that categorising data reduces human cognitive demands by making it more manageable and therefore suggest categorising large amounts of data into easy and manageable chunks or themes. The use of thematic approaches can facilitate such process. Eleveld and colleagues (2003) illustrated the use of such approaches within their requirement analysis process for a virtual coastal and marine data warehouse, where user views were gathered using scenarios and then categorised into broad themes based on the types of user groups and their

functions. However, there is limited research available in the field, suggesting the implementation

of systematic methods for managing and analysing user views (Lindgaard et al. 2006).

Once the requirements are gathered, managed and analysed, there are different methods of presenting these in a technical form (system design) and in a language that is understandable in the technical world (Kujala et al. 2001) so that user needs can be developed and implemented in user-acceptable systems. For this purpose, the requirements are documented in a plain language¹⁵ as well as in a diagrammatic form (a data model)¹⁶. A multidimensional structure is used for data modelling within the warehouses, as it can show multiple relationships between multiple data points, which can support complex analysis and data representation (Kimball and Ross 2002).

In summary, the requirement analysis process includes three main activities: requirements gathering, their analysis and presentation. The choice is available to use a number of methods (as discussed above) for each activity that can either be employed individually or in a synthesised form. However, before making any decision, it is important to consider some important situational factors such as the type of system, the available resources, the nature of requirements and the type of requirement sources. A detailed review of the literature reporting various methodologies for requirement analysis for a data warehouse is presented later in this Chapter. While reviewing these methodologies, the situational factors for a DCM data warehouse are identified

¹⁵ A list of requirements is prepared in a plain language, which means using only simple sentences and not any kind of diagrammatic notations.

¹⁶ A data model is a diagrammatic representation of entities (e.g. mapping participant, mapper, mapping session) and their relationships.

and discussed to highlight the inappropriateness of the existing methodologies.

Within data warehouses, the requirement analysis methods are categorised into two major approaches and are discussed next.

4.7. Requirement analysis for a data warehouse

A wrongly designed data warehouse does not serve the purposes it is meant for and eventually is not accepted and used by the users (Lindgaard et al. 2006; Teixeira et al. 2012). Therefore, effective and appropriate approaches are required for designing the data warehouse. These approaches are determined by the method according to which the requirements are obtained and analysed (Weiskopf et al. 2013). Two main approaches are used for requirement analysis within warehouses; data-driven and user-driven approaches. These are based on the sources from which the requirements are obtained and analysed.

4.7.1. Data-driven approach for requirement analysis

When requirements are obtained from existing systems through analysis and determination of the features that could be part of the new system, the process of requirement analysis in the data warehouse is called a data-driven approach (Inmon 1996). Inmon (2012a) asserts that this approach is usually suitable when the aim is to replace the old system with a new system that is robust and efficient. The success of this approach depends on two main aspects: the effectiveness of the existing system in doing what a new system is supposed to do; and understanding the requirements to make the

new system more robust and efficient (List et al. 2002). Within both aspects,
the analyst uses

technical or objective methods to understand, analyse and modify the functionality of the existing system, with the modifications usually technology-driven by automated methods rather than user-driven (Song et al. 2007). Abai et al. (2013) assert that, while a data warehouse built using the data-driven approach of requirement analysis might be technically workable, it may fail in terms of lack of user satisfaction and acceptability standards.

Further, the data-driven approach is not suitable when the aim is to introduce a new system within an organisation or to obtain a system's future requirements (information about the functions of the system that is not currently available), as only users can perceive the future requirements, not the existing technical system. This approach is argued as being reliable and successful in areas where the existing primary data management systems are perfect and the new system (the data warehouse) could be successfully developed by replicating existing data models (Moody and Kortink 2000).

As was mentioned in Chapter 1, my previous study employed a data-driven approach for requirement analysis, where existing primary DCM data management systems were analysed to gather requirements for a future data warehouse (Khalid et al. 2010). While the previous study used a data-driven approach to demonstrate the technicality of the approach, the proposed data warehouse design cannot be guaranteed for user acceptance. Two reasons support this supposition: first is that the existing systems are not effective. The Excel programme has limitations in terms of effective data management (Khalid 2009) and the arc|hive DCM database, although developed to resolve existing issues, is still in the early stages of its usage.

Further, both of the existing DCM data-management systems were developed to support primary

use of DCM data rather than secondary use. The secondary use of DCM data might highlight additional data needs and relevant issues. Second reason is that DCM data warehouse is a new system and its future requirements can only be perceived by the potential users rather than the existing systems.

Based on these reasons, therefore, the reliability and effectiveness of the data-driven DCM data-warehouse model cannot be guaranteed particularly in terms of user acceptability. However, the users' views for their potential secondary use of DCM data can contribute towards identifying requirement that can support designing a future data warehouse, which is potentially user-acceptable. A user-driven approach is explained next with a rationale of its need for within requirement analysis for data warehouses more generally and specifically for a DCM data warehouse.

4.7.2. User-driven approach for requirement analysis

When an analyst obtains requirements from the users, the approach to requirement analysis within a data warehouse is known as a user-driven approach (Golfarelli 2010). This approach is suitable when the aim is: to introduce a new system into the organisation; to understand organisational structure; to understand issues with the existing system from the users' perspective; and to explore users' needs and expectations for a future system (Golfarelli 2010; Teixeira et al. 2012). Some studies also use the term 'goal-driven' alongside user-driven approaches that involve understanding the organisation's intentions regarding the business process, or in other words, determining the services that the organisation is providing to users (List et al. 2002). This organisational knowledge can be retrieved

from the organisational policies and system documents or from the stakeholders, i.e. those people at

the decision-making and managerial level, those who will use the system at the end and those who will be affected by the system within the organisation. However, people are the main users of a system and of the data stored within it and, therefore, they play an important role in establishing the requirements (Scandurra et al. 2008; Martikainen et al. 2014). A system can be technically flawless, but if it is not acceptable to users, it fails and therefore the literature has emphasised gathering user requirements for designing user-acceptable and thus successful systems (Kappleman 1994; Raab 1998; Teixeira et al. 2012).

While both data-driven and user-driven approaches have their unique characteristics in regard to designing data warehouses for specific reasons, combining both approaches (this is called a mix-driven approach) is also encouraged (Golfarelli 2010; List et al. 2002). According to Golfarelli (2010), a mix-driven approach can provide a well-balanced approach that reflects both previous systems and also the user views. Currently, a data model for a DCM data warehouse, designed using a data-driven approach, exists and this study aims to identify user-driven requirements that will contribute to designing a system that has collected requirements using a mix-driven approach.

4.8. Significance of a user-driven approach

User needs and their requirements are considered important for information-system designs. ISO 9241-210 (2010) recommends involving users so as to understand their needs and requirements for the system with which they will be interacting in the future. While criticising the lack of any clear definition of

user involvement, Kujala and colleagues (2001) argue that user involvement is a broad term that describes a direct contact with users, which can be

achieved through various approaches. They categorise user involvement as: informative, where users are providers of information, or objects for observation; consultative, where users are consulted about specific ideas or issues related to the system; and participative, where users are actively involved in designing the system with the analyst and the designer. In the design of any information system, user involvement is encouraged for designing and developing user-acceptable systems. Kujala et al. (2001) argue that existing systems and documents (data-driven approaches) might provide only the information that can be used to develop a technical workable system and that these artefacts do not describe the context within which the system will be used. In order to understand the context, including issues and concerns, the user views are important (Kujala et al. 2001). Within the context of a data warehouse, user involvement is considered significant because of the following main reasons.

4.8.1. Identifying Information requirements

Winter and Strauch (2003) emphasise understanding the user's 'information requirements' when designing data warehouses. This involves recognising the kind of questions that the user will be asking from the system, the kind of data required to answer these questions, the format that such data needs to take in order to answer the questions efficiently and the sources that will provide that data (data sources).

4.8.2. Recognising social issues

Data warehouses collect and store data taken from various sources. While there are technical issues associated with what and how the technology is used to extract, integrate and store data effectively and efficiently within the

data warehouse, Lamas and colleagues (2015) identify social and ethical issues such as patients' rights, solidarity and common good, transparency and trust when it comes to using data for secondary purposes. Further, Ancker et al. (2011) and Law (2005) both argue that these issues cannot be ignored when the aim is to populate the data warehouse with healthcare data for secondary use.

While data sharing, privacy and security are important issues, users' trust of both the data and the system is also reported to be significant for acceptability of the future system (Wu and Chen 2005; Kassim et al. 2012). Lankton and colleagues (2014) assert that users will not use the system if they do not trust the data. To be a user-acceptable and successful system, therefore, a data warehouse should not only be technically workable but it should also be socially trustworthy and acceptable by its users. Involving users can identify the social issues that need to be considered in order to design a trustworthy and acceptable system.

4.8.3. Usability requirements

Furthermore, users are involved for identifying the usability aspects of a system, i.e., the information about how often they will want to use the system, what type of interface they would like and how they would like to access the data (Kassim et al. 2012). This suggests involving users in requirement analysis. While user requirements are essential for any system's success, the actual gathering, management and presentation of user requirements is a complex, demanding and challenging task and therefore requires careful consideration (Niès and Pelayo 2010). I will next discuss these challenges and the suggested solutions identified in the literature.

4.9. Challenges of a user-driven approach

While a user-driven approach is considered significant for a data warehouse success, the literature also reports a number of challenges that emerge when employing a user-driven approach for designing any information system. These challenges and their existing solutions are discussed, with the specific aim of explaining their significance for the process of requirement analysis for a DCM data warehouse.

4.9.1. Identifying the right users for the right requirements

Abai and colleagues (2013) assert that only the right users can lead to the right requirements and, as assured by Nies and Pelayo (2010), only the right requirements ensure that the right decisions will be taken for system design and development. However, identifying the right users is reported to be one of the major challenges when the user-driven approach is chosen for requirement analysis (Niès and Pelayo 2010). The perception of the right user varies in relation to the various systems. In the context of user-interactive systems, where usability is the main feature, Seyff and Graf (2010) urge identifying the individuals who will be directly interacting with the system and who can explain precisely what they require from the system for good usability. However, in the context of a data warehouse, where the main aim of requirement analysis is to identify the type of data and its potential uses, Kujala et al. (2001) suggest identifying the users who can in turn identify the type of data and associated needs. This also means that one criterion for user identification is that they must possess sufficient knowledge about the data and their potential needs. For example, users should be familiar with the data that can provide them with

specific information from the data warehouse and with both its existing and potential uses.

Within this study, the aim is to gather requirements for a future DCM data warehouse from the user's perspective. As a DCM data warehouse is a conceptual proposal, where potential users yet need to be identified for requirement analysis, it is important to recruit the individuals who have sufficient knowledge and understanding of the data that will potentially be stored within the warehouse. Considering this, mappers, who have DCM training at any level and have some mapping experience, were considered to be the potential users of the warehouse.

Another motive of approaching mappers as potential data warehouse users could arguably be that their views have always been at the heart of any new intervention in DCM, such as the developments in DCM as a method/tool have been based on the suggestions of experienced DCM users (Innes and Surr 2003). In 2001, at the University of Bradford, a think tank (Brooker and Rogers 2001), containing mappers from various backgrounds such as Bradford Dementia Group, DCM trainers, practitioners and researchers, was established to share experience and their use of DCM data to establish how DCM can be used at its best and how it can be modified.

Similarly, Brooker and Surr (2006) also report a study of an initial validation of DCM 8 (current version) over DCM 7 (previous version), where they used views and suggestions of national and international mappers (experts working groups) to examine various aspects of DCM with the aim of

refinement and development of the new version. Considering a data warehouse as a new

development within DCM to support the use of DCM data for secondary uses, it could be argued that mappers can be potentially in a best position to inquire about how they perceive the potential use of DCM data for secondary uses.

4.9.2. Users' understanding and knowledge of the new system

A number of other issues can emerge while using a user-driven approach. Christel and Kang (1992) list such issues that could be faced during the requirement analysis process. The first issue is related to users' lack of understanding of the new system, a problem, which transpires when they are not completely sure of what they require or need from the system. The second issue is related to users' knowledge of specific technical requirements for the system, which could mean users expressing needs and expectations that may conflict with what can actually be translated into requirements of the specific system. For example, users can request or require something which is not technically possible or which would create an unstable, poor quality or unsecure system. Sandurra and colleagues (2008) link this issue with a lack of computer knowledge and literacy among healthcare professionals. They further assert that, in such cases, as users do not know their technical requirements, they therefore cannot communicate them effectively. These issues could emerge for DCM since potential users work in health and social care and so may not be aware of technical requirements or possibilities of a data warehouse for secondary uses. In these cases, the users have freedom to express their views in a non-technical language, which are further interpreted by the analyst into the system requirements.

According to Christel and Kang (1992), the third issue emerges when the same user changes her/his requirements over time or when different users have

conflicting requirements at the same time. Nies and Pelayo (2010: 79) maintain that users' conflicting requirements emerge when "users don't know what they want". They use examples from their own study in which healthcare professionals provided conflicting requirements for designing a patient's treatment information display within a hospital information system. Some users required a detailed view of the patient's treatment information on a single screen, while some required a synthesised view of the same information. To deal with these issues, Nies and Pelayo (2010) suggest using human-factor methods, such as activity analysis, which means observing users while they perform various tasks to understand what could help them to gain a better view of the patient's treatment details. In these cases, the analyst could use inferences and her/his own imagination and judgment to envision user needs that support developing a system beneficial for all. Browne and Ramesh (2002) suggest that the analyst's imagination and judgment come from her/his prior knowledge and experience.

While Christel and Kang (1992) only focus on the users, Browne and Ramesh (2002) associate user-requirement gathering challenges with both the user's and the analyst's different cognitive and behavioural characteristics. They list four challenges, the first being cognitive bias, such as 'ease of recall' and 'overconfidence', which could influence users' responses. For example, users could either forget to recall important information or could become overconfident in explaining their requirements, which might result in incomplete information being presented. On the other hand, because of cognitive bias, the analyst might adjourn the requirement gathering process pre-maturely, thinking that she/he had all the information.

The second challenge is that of satisficing, which develops as a person adapts to a certain environment or situation, learns to use short cuts to perform tasks and then describes their requirements (Browne and Ramesh 2002). For an analyst, satisficing is not spending enough time gathering and then understanding user requirements. Browne and Ramesh (2002) associate satisficing with a kind of behaviour that reflects humans' natural habit of trying to terminate knowledge gathering with very little understanding. The third challenge is that of faulty reasoning, when the user might express incomplete or incorrect requirements based on their incomplete mental model of the new system (Browne and Ramesh 2002). Brown and Ramesh further explain that, in these cases, users make illogical inferences, as they do not fully understand what the new system could be. The fourth challenge is automaticity, which refers to a human's task-performance habit, with time, becoming part of their automatic routine, which sometimes results in users unintentionally failing to explain routine tasks (Browne and Ramesh 2002). These tasks might be important for the analyst to know the detailed picture.

As a DCM data warehouse is a novel concept for managing DCM data for secondary use, it can be assumed that potential users might not be familiar with specific requirements for the data warehouse. Therefore, the analyst needs to adopt methods, which encourage the users to express as much information as possible that can establish the data-warehouse requirements. The methods chosen for this study were designed to elicit in-depth responses of the users and are presented in Chapter 6. These explain how

the mappers were encouraged to explain their needs, issues and concerns regarding secondary use of DCM data.

4.9.3. Communication Issues between the analyst and the users

Another challenge of the user-driven approach is communication between analyst and user (Gallivan and Keil 2003). According to Browne and Ramesh (2002), communication issues usually emerge when user and analyst have different backgrounds, different languages (the user having a domain-specific language and the analyst a technical language) and a different understanding of the terminologies specific to each other's domain. In this context, the analyst is required to understand the user's language rather than the user communicating in a technical language.

As argued above, mappers could be considered as potential data warehouse users and thus could be considered for requirement analysis within this study. Therefore, there is a risk that such communication issues may emerge, as mappers have specific training, knowledge and understanding of the terminology used for DCM data and likely have limited technical and data warehousing knowledge. Chapter 6 illuminates how this issue was dealt with within this study.

4.9.4. User-requirements expression

The literature has highlighted that users express requirements differently (Wang 2014), which introduces additional issues for the analyst in terms of understanding and interpreting a range of diverse requirements. For example, users can express requirements directly or indirectly. According to Pace (2004), direct requirements are those where users are familiar with what they want from the new system. The direct requirements usually emerge when the new system replaces the existing system and the users

(who were using the existing system) are familiar with the issues and problems that need rectifying

in the new system. Indirect requirements are those where users are not familiar with their requirements regarding the new system (Pace 2004). Pace (2004) argues that, when users are not familiar with what they want from the new system, their requirements could be translated or interpreted from their experiences of using their existing systems. It could therefore be suggested that, in the context of a data warehouse, users can be asked about what they want from the data rather than from the new system. Kujala et al. (2001) verify this when asserting that, for a data warehouse, the focus should be more on learning about how users anticipate the use of data from the warehouse rather than on technical aspects of the system. This suggests that, if users are familiar with the data that need to be stored within the warehouse, they can indirectly express requirements for the new system. As potential users might not be familiar with the DCM data warehouse and their specific requirements, it is important to explore how they would perceive the potential uses of DCM data within the warehouse. This might require focusing on both directly and indirectly expressed requirements by the users, which requires appropriate and exploratory methods.

In summary, a number of issues can arise while gathering requirements from users. These are usually related to identifying the right users who can provide the right requirements, users' lack of knowledge and understanding of the new system, communication issues between analyst and user and, finally, the diverse ways in which users express their requirements.

4.10. User-driven approach within a data warehouse

The literature reports a plethora of methodologies that explain user involvement during the requirement analysis process for a data warehouse.

In

this section, I examine some of these studies with the aim of arguing why their described methodologies were not chosen for requirement analysis of a DCM data warehouse. While doing so, the need to employ a suitable methodology for identifying user requirements for a DCM data warehouse is also emphasised.

Paim and de Castro's (2003) methodology divides the requirement analysis process into three phases in which users are involved in identifying organisational and decisional goals. These goals are further extended to design data models for the warehouse. As the users identify the organisational goals and decisions, this methodology will only work effectively within an organisational structure when the users are visible, both at organisational and decision-making levels. Gam and Salinesi (2006) call their methodology 'CADWA'. Using this methodology, they recommend beginning requirement analysis by involving users in establishing the main goals of the organisation for their data use and further involving users in elaborating the identified goals in more detail. Further data models are designed based on the details of organisational goals. Similarly, Mazon et al. (2005) also proposed a methodology where they identify the organisational goals and divide these into three types, strategic, decision and information goals. Starting from the top-level management, strategic goals are identified. Further, these goals are elaborated and developed into decision goals by first involving decision-making users¹⁷ and finally information users¹⁸. The

¹⁷ Those individuals who are at the top level (in terms of decision-making) of an organisation and who require reports of summary and aggregated data.

¹⁸ Those individuals who interact with the data to generate reports for further decision-making.

decision goals are explained in terms of specific information that would facilitate the achievement

of strategic goals. Like CADWA, this methodology also seems workable within an organisational context where the organisational structure, from strategic to decision to information level, is clear and established and users are also visible for requirement elicitation.

Similarly, a 'triple-driven' methodology by Guo et al. (2006) combines analysis of the existing systems (data-driven approach), the user needs and also the organisational goals (user-driven approach) in parallel fashion for identifying the requirements for designing a model for the data warehouse. Again, this methodology would work in a situation where users and data sources are visible and where users wait to be asked their requirements for the data warehouse. Kaldeich and Se (2004) also reported on their methodology in which they combined both user- and data-driven approaches within the requirement analysis process. They commenced by analysing the existing data models from the source system, identified the business process and produced a model they called the AS-IS model, which showed the current business process. They involved end-users and, based on their requirements, they produced a TO-BE model that showed the user-required business process. They then produced an IPD (interview process model) by integrating the AS-IN and TO-BE business models. The IPD consists of requirements that are further validated by involving the users from senior management to ensure that these requirements will achieve the business goals. This methodology also demonstrates a user-driven approach within an organisational context, where users at various management levels are identified for requirement elicitation.

Schiefer et al (2002) propose the easy REMOTE^{DWH} methodology, which considers requirements from various stakeholders' perspectives, according to several levels of abstractions, starting from understanding business needs through to user needs and then going further into functional and non-functional system requirements. Romero and Abello (2010) present an AMDO (automating multidimensional design from ontologies) methodology where they re-engineered the multidimensional concepts from the existing data sources using an automated process and then further validated these requirements from the user perspective, thus used a mix-driven approach. They call this approach, however, a user-driven approach rather than a mixed method as users are the main contributors in identifying the requirements. Both the easy REMOTE^{DWH} and AMDO methodologies and their workability were demonstrated in an organisational context. The applicability of these methodologies within a non-organisational context is not known.

Usually, the data warehouses are designed and developed within an organisational context with the aim of bringing an organisation's segregated data into an integrated format to facilitate secondary analysis such as the establishment of profit margins, cost effectiveness and marketing (Lyman et al. 2008). Therefore, the above examples of a user-driven approach for designing a data warehouse highlight the importance of the organisational context within which the traditional data warehouse system is situated. This is also the reason why the requirement analysis process is mostly aimed at understanding the organisational context, the processes and goals and the

decisions made by those who hold some kind of stake in a data-warehouse development project. Therefore, the emphasis is on identifying requirements

that reflect organisational needs (Schaefer et al. 2011). Within the organisational structure, the identification of users and their requirements is comparatively a less challenging task as the users are usually familiar with the basic requirements that the new system needs to meet. However, as Schaefer et al. (2011) argue, there are some situations when data warehouse development is initiated based on the potential significance of the data for secondary use rather than on the organisation's need to use data for secondary purposes. They give an example from a project where a data warehouse for educational data was developed.

International large-scale education-assessment studies provide statistical data about student performance that could provide a rich source of educational data for policy makers and researchers to understand the institutional factors that influence student-learning outcomes (Schaefer et al. 2011). When data is increasing in volume, it needs to be managed (integrated storage) to give meaningful outcomes and also made accessible to the users for their secondary analysis; a data warehouse can be one of the workable approaches for managing data as it is an established approach for dealing with data integration, complex analysis and reporting (Lyman et al. 2008).

In such cases, the development process commences with understanding the data needs of users who might not be part of a single organisation and therefore might identify diverse data needs that do not reflect a single organisation's goals. As mentioned by Schaefer et al. (2011), in such cases, the organisational view is usually missing. Therefore, the analysis of the

organisational environment is not possible and the requirements are usually elicited from individuals who are not stakeholders within the project and who

might be using the data for their specific and individual purposes. Yet it is important to understand their information requirements so as to identify the type of data and its arrangement within the data warehouse. Schaefer et al. (2011) assert that the situations where data warehouses are designed and developed within a non-organisational context are also emerging within healthcare, particularly for providing data for medical research. The literature presents a plethora of methodologies (as reviewed above) when data warehouses are designed and developed within an organisational context. However, there is a dearth of methodologies which specifically support the requirement analysis process for data warehouses within a non-organisational context, particularly focusing on the above challenging situations.

Further, the existing methodologies (as mentioned above) for requirement analysis within the context of a data warehouse do not report on how the collected data could be analysed. For example, Mazon et al (2000) collected qualitative data (user views) and identified three types of goals, strategic, decision and information goals. However, they did not explain how they identified these goals from users' interviews and thus detailed analysis techniques are not part of the proposed methodology. Similarly, Giorgini et al. (2005) proposed an approach called GRAnD (Goal-oriented Requirement Analysis for Data warehouses), which is a combination of goal-driven and data-driven approaches. They divided the requirement analysis process into organisational and decisional modelling and mixed design framework. They started by understanding organisational objectives through interviewing

stakeholders and identifying facts and dimensions from their objectives. This is followed by interviewing decision-makers to obtain their business goals.

Facts, dimensions, and measurements were extracted from these goals and then both were mapped with the data source and then enhanced further. However, how the interview data was reduced to identify various data attributes was not reported as part of the methodology. There is limited research on how user views can be systematically managed, analysed and interpreted to reach requirements that are meaningful, coherent and mutually agreed at a conceptual level. There needs to be a flexible and exploratory methodology that focuses on aspects of the identification of stakeholders (users), on the elicitation of requirements from those users who might not be familiar with their requirements and on reducing user-conversational data in meaningful categories (requirements).

In summary, the literature on existing methodologies focuses on the aspects, which do not reflect the existing situation of DCM data. For example, the above reviewed studies' focus is on gathering requirements within an organisational context, which means that users are familiar with their requirements and these are visible as part of the requirement analysis process. Further, the reported methodologies capture users' views using various methods, for example, interviews, group discussions, workshops etc. However, very little is known how a large amount of data (user views) is processed from a raw form to become meaningful information (requirements).

4.11. Summary of the chapter

This chapter began by establishing the concept of requirements within a data warehouse and argued that, when users are involved, the requirements can emerge as their views, perceptions, concerns and expectations regarding the

new system. As argued further, in this case, an analyst plays a significant role

in translating users' views, perceptions and concerns into requirements that need to be considered for designing and developing the future systems. The chapter then continued exploring various methods currently used for requirement analysis. The need for an effective set of methods was highlighted and suggested that it requires a careful assessment of the situational context within which the new system will sit.

The chapter then examined the two main approaches for designing a data warehouse, data-driven and user-driven. A detailed and comparative view of both approaches was provided, with a specific aim of justifying the need for a user-driven approach for a DCM data warehouse. This was followed by a description of the significance of a user-driven approach for a data warehouse during the requirement analysis process. It argued the need for involving users so as to determine information needs that can reflect the type of data that should be stored in the warehouse to meet the users' analytical needs. It further highlighted the possibility of identifying the social issues and the usability requirements while involving users.

The chapter then went on to identify the challenges of involving users in requirement analysis and outlined a number of challenges. When highlighting the challenge of identifying the right user for the right requirements, it argued the need to find those individuals who are data users and are therefore familiar with the data that will go into the warehouse. Based on this, the chapter therefore assumed that mappers could potentially be the potential DCM data-warehouse users. The chapter also discussed some of the challenges that could emerge when users have little or no understanding of

the new system. It therefore argued, that the analyst needs to encourage users to express as

much information as possible to establish the data-warehouse requirements and that this requires employing suitable methods. The chapter continued by discussing the communication issues as a challenge, which can emerge between the analyst and users due to differences in knowledge, skills and specific terminology. It was also highlighted that another challenge could be that of recognising how users express their requirements which might emerge directly expressed by them or which might need extrapolating from their conversations.

Briefly, this chapter has explored the significance of a user-driven approach for a data warehouse. Based on this, the importance of a user-driven approach for a future DCM data warehouse has been argued. While a user-driven approach is encouraged within the context of a data warehouse, it is important to identify the right methodology for applying a user-driven approach. While considering an effective methodology for requirement analysis for a future DCM data warehouse, it has also been argued that various situational factors need to be considered such as the need for identification of potential users and eliciting requirements from users with limited technical knowledge and understanding of the new system. Considering these situational factors for DCM, Chapter 5 provides a detailed view of the type of methodology chosen for employing a user-driven approach for requirement gathering and analysis within this study.

4.12. Aim of the study

The aim of this study was to explore requirements for the secondary use of DCM data within a data warehouse using a user-driven approach. The research question is as follows:

“What are the requirements for the potential secondary use of DCM data from a data warehouse from its potential users’ perspective?”

4.13. Objectives of the study

In chapters 2, 3 and 4, the literature was reviewed for three purposes. Chapter 2 explored existing and potential uses of DCM data by reviewing the relevant literature. This exploration contributes towards understanding the potential of DCM data for secondary use, which requires the need for an effective IT system for such purposes, where user requirements need to be identified.

Chapter 3 examined my previous study where a data warehouse approach was proposed for DCM data to facilitate its potential secondary use. It highlighted the aspects that could potentially be unfolded by a user-driven approach. The main aim of involving users is to identify their information requirements which can provide insights about the data that needs to go into the warehouse.

Chapter 4 then provided a detailed view of a user-driven approach for requirement analysis and its significance for a data warehouse in general and for a future DCM data warehouse in particular. This chapter argued that, when users are not familiar with their requirements, the requirements could

emerge as their perceptions, concerns and expectations from a new system.

This

requires a specific methodology to elicit, explain and interpret user views into requirements for a future data warehouse.

The data in chapters 2, 3 and 4, combined led to the formulation of the following study objectives:

- Identifying potential use/function for a future DCM data warehouse
- Identifying data that needs to go into the warehouse to meet the data warehouse's potential use
- Identifying the concerns and issues (if any) related to the secondary use of DCM data for the identified potential use

In order to meet these objectives, an appropriate methodology and suitable methods of data collection and analysis were required. These will be addressed in chapters 5 and 6.

5. Philosophical and Methodological Considerations

“Never ask the end user ‘What do you want in your data warehouse?’ That puts them in the position of designing the system. That is your job. Besides, there is only one right answer to this question: ‘Everything’. Instead, ask questions that help you learn what the end user does, and then translate/interpret this into what needs to go into the system”.

(Mundy et al. 2006: 12)

5.1. Introduction

Denzin and Lincoln (1998) suggest four main steps for defining a research process. The first step is choosing a research paradigm that reflects the researcher’s philosophical assumptions in understanding the nature of the world. The second step is selecting a research approach consistent with the researcher’s chosen philosophical beliefs. The third step is choosing a research methodology that guides the process of data collection and analysis. The fourth step involves the data-collection and analysis methods that are underpinned by the chosen research methodology, approach and philosophy. Below is a diagrammatic representation (Figure 6) of the four steps of the research-design process.

The aim of this study was to identify requirements for secondary use of DCM data using a user-driven approach. To achieve this aim, within this study, these aforementioned four steps were followed to design the research process. This chapter focuses on the first three steps, which are related to the philosophical and methodological considerations employed within this study. It first outlines my philosophical position and critiques its appropriateness for the analysis process required within this study. It then

outlines the use of grounded theory as a methodological framework,
underpinning a qualitative approach, to guide

the practical steps taken to design the data-collection and analysis process. The pertinence of grounded theory as an effective research methodology for identifying potential users' requirements is also justified in detail. The methods of data collection and analysis, which are the fourth step of the research process, are then presented in Chapter 6.

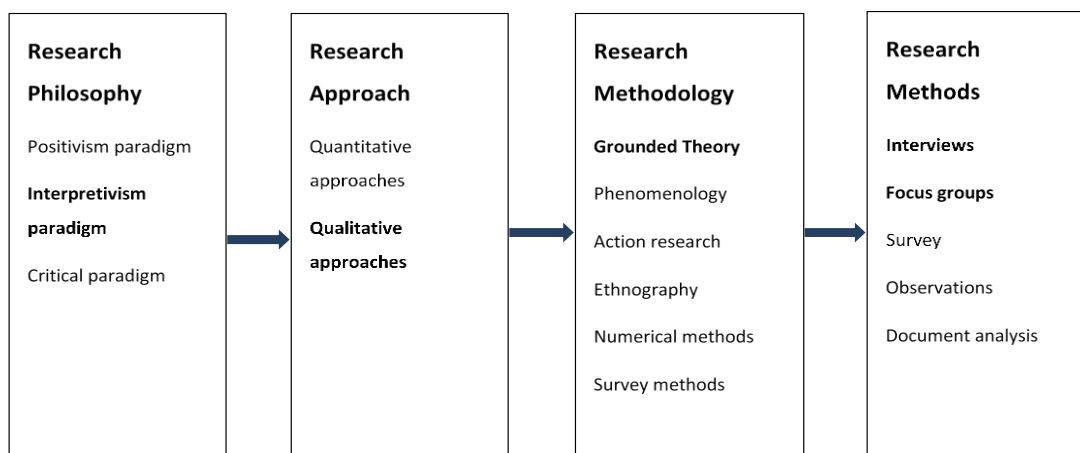


Figure 6: Research process stages

5.2. Research philosophy

Every research study, regardless of the kind of problem under inquiry, is underpinned by some kind of philosophical assumptions (Myers 1997). These assumptions guide the research towards valid and appropriate approaches to solve a particular problem and are related to beliefs that guide the study design, the investigations and presentation of the findings (Guba and Lincoln 1994; Elsheikh 2011). These assumptions can be classified as both ontological and epistemological. Ontological assumptions are related to the beliefs that one holds about the 'phenomenon' or the 'meaning of the object' under study. This also describes the researcher's views about the

nature of reality, which see reality either as a single unbiased truth (objective), or as a

set of multiple realities that are constructed socially (subjective). The epistemological assumptions refer to one's concept of knowledge; what knowledge is and how it can be obtained (Hirschheim 1992). Orlikowski and Baroudi (1991) maintain that such assumptions also concern the criteria of validating, constructing and evaluating knowledge. The researchers usually take one of two main positions to reflect their ontological and epistemological assumptions, positivism and interpretivism.

Interpretivism concerns those researchers who believe in multiple realities (Krauss 2005) and believe that these are constructed socially through understanding the interaction between social actors within a particular phenomenon¹⁹ (Orlikowski and Baroudi 1991). They also accept that knowledge about the phenomenon or the object under study can be presented or delivered based on interpreting meanings of the actions those individuals had within their social world. Charmaz (2006) asserts that these interpretations are consistent explanations of the phenomenon under study rather than law-like generalisations. Positivism, however, concerns those researchers who believe in one true reality and who conduct the research process to test that reality as established hypotheses or as a theory about a social phenomenon (a deductive approach²⁰) (Creswell 1994). Positivists accept the social reality as objective, testable controlled and independent of external factors (including the researcher) and believe that it can be quantified using scientific research approaches, i.e. quantitative methods.

¹⁹ 'Phenomenon' refers to events, situations, objects, or social settings (Denzin and Lincoln 1998).

²⁰ A deductive approach is concerned with commencing a research study with established hypotheses and theories (Creswell 1994).

The main aim of this study was to understand how users interpret their current use of DCM data or have expectations for its future use for various purposes, both in individual and organisational terms. This research study, therefore, intended to explore individuals' and their organisations' needs, expectations and concerns of using DCM data as regards the requirements for a data warehouse, rather than quantifying what they already use in terms of their satisfaction levels with the existing systems. Therefore, my ontological and epistemological perspective of understanding and acquiring knowledge within this study is underpinned by interpretivism, which holds the assumption that DCM data use and its management is not objective but subjective knowledge. Therefore, users' views related to their subjective perspectives and their shared meanings of the potential use of DCM data will inform the requirements of the future data warehouse. In addition to this, this thesis will also adopt the assumption that the phenomenon can only be understood through the meanings that users who are part of that phenomenon assign to them. My view of looking at requirement analysis is in line with what is proposed as a preferred philosophical position to designing systems within information systems research, which is where a data warehouse, as a technical concept, sits.

The literature highlights that, like any other research study, information system research is also underpinned by some philosophical paradigms informing the researcher's philosophical beliefs or views. These views are related to the role of the information system within an organisation, which can be considered technical (e.g. the information system acts as a technical

tool providing specific well-defined functions), social (e.g. the information system as a social tool

involving social processes that are dynamic in nature) or socio-technical (e.g. both a technical and a social tool). It is widely acknowledged that the views of seeing an information system have shifted from pure technical to social and socio-technical (Ghaffarian 2011). The major reason for this shift was the understanding that an information system deals with people, data and technology, with all three aspects important for its success (Iivari and Hirschheim 1996). The change in understanding an information system has also influenced the perception of the system requirements.

Iivari and Hirschheim (1996) argue that the system requirements can be: objective (an objective reality that will be part of the system as a measured feature); subjective (requirements that are subjective to each user, based on their role and cognitive abilities); and inter-subjective (the synthesised set of common or homogeneous requirements gathered from various users). Taking a technical view of the information system within an organisation directs the analyst (the researcher) towards using objective methods of requirement analysis for designing the system (Iivari and Hirschheim 1996). These methods are underpinned by the positivist approach, whereby the researcher aims to acquire the functional descriptions of the systems objectively from the existing systems, a data-driven approach for designing a data warehouse, as discussed in Chapter 4. However, the social and socio-technical view of the information system encourages the researchers to explore the social side of the system by involving users and understanding their subjective perspectives on the system use, which considers requirement analysis as a social process (Jantunen and Gause 2014;

Kroeger et al. 2014). As argued in Chapter 4, the users' subjective perceptions of the system, which might emerge from their

needs, complaints, expectations, concerns and issues, are interpreted by the researcher as the system requirements, thus suggesting the suitability of the use of interpretivist approaches for requirement analysis. Myers (1997) adds that the need to understand information systems from an organisational and individual context, rather than merely a technical perspective, has also shifted the trend towards using interpretivist approaches.

Another reason for adopting the interpretivist approach in this study was its flexibility in exploring new areas. Ruddy (2010) suggests that the interpretivist paradigm is valuable in facilitating research into new perspectives on already known areas. Strauss and Corbin (1998) also assert that the interpretivist approach is useful for exploring a new area to reveal the processes and the phenomena when there is little known already (Strauss and Corbin 1998). Given that there was a dearth of literature and, therefore, knowledge on secondary use of DCM data generally, and from users' perspectives specifically, the interpretivist approach seemed appropriate to explore the requirements for the secondary use of DCM data, an area that is yet unexplored.

Within this study, a qualitative research approach, underpinned by interpretivism, is applied in order to understand the use of DCM data by potential data warehouse users, as presented next.

5.3. Qualitative research approach

According to Creswell (2007), researchers apply those research approaches, underpinned by interpretivism, that are appropriate to exploring the nature of a problem, issue or phenomenon; in this case, a qualitative research

approach. A qualitative approach is appropriate to use when the aim is to understand the situations, events, process and experiences of people and the context within which these take place (Huberman and Miles 2002), rather than to quantify objects, events and their occurrence in terms of numbers. Kaplan and Maxwell (1994) argue that, when data are quantified, the goal of understanding people and the phenomenon in the social and cultural context within which they live is lost.

A qualitative approach to research can incorporate either deductive or inductive methods of inquiry. A deductive inquiry commences with pre-established assumptions and theories, where qualitative approaches are used to validate such existing assumptions and theories (Creswell 1994). However, inductive inquiry commences with very few or no assumptions, where the qualitative approach is applied to generate new hypotheses or theories rather than validating them from the collected data (Strauss and Corbin 1998). The qualitative research approach with the inductive inquiry method is used when the intention is to understand social events, situations and needs, which are not described in existing theories or hypotheses in relation to a specific area, through the meanings that people bring to them (Myers 1997; Klein and Myers 1999). As there is a lack of knowledge of the requirements of secondary use of DCM data from the user's perspective, in this study, a qualitative approach incorporating an inductive inquiry method will be used to explore and identify the new hypothesis or theory (a set of requirements) from the collected study data.

As mentioned above, since information systems are perceived as a combination of technology, organisations and people, an increasing number

of studies are focusing on understanding the context, experiences and narratives of people/users involved either in designing and developing, or using, systems. This has initiated the use of qualitative research approaches within information systems (Hughes and Jones 2012). It was also argued in Chapter 4 that requirement analysis is considered a social process, where communications take place between users and the analyst to elicit users' requirements (Zowghi and Coulin 2005), thus suggesting the use of qualitative research approaches for this purpose. The aim of this study was to explore the needs, expectations, issues and experiences of potential users regarding their secondary use of DCM data and the context within which these requirements emerge. Therefore, qualitative approaches were considered suitable and effective for such a purpose.

5.4. Grounded theory techniques as a methodological framework

According to Denzin & Lincoln (1998), a methodology provides a set of frameworks or practical guidelines to conduct inquiry through a data-collection and analysis process. There are several methodologies, underpinned by the qualitative research approach, that are used for both deductive and inductive inquiry purposes. The most common methodologies are ethnography, phenomenology, narrative approaches, case studies and grounded theory. Creswell (2007) contends that the application of these methodologies depends on the nature of the inquiry, the situation under which the inquiry takes place and the type of research questions that need answering. Within this study, the techniques of grounded theory are used as a methodological framework for inductive inquiry underpinned by qualitative

methods of data collection and analysis. It is beyond the scope of this chapter to describe other methodologies

or to justify why these were not used for this particular study. However, the discussion will justify why the techniques of grounded theory were appropriate to and thus were adopted to inform various methodological aspects of this study. Therefore, the next section will explain the grounded theory methodology and my rationale for using it as a methodological framework for this study.

5.5. Grounded theory

The theoretical origin of grounded theory is linked to symbolic interactionism. Blumer (1969) proposes that symbolic interactionism is concerned with the meaning that individuals and groups give to objects, events and situations, which is influenced by how they interact with these in their environment. These interactions therefore shape their interpretations about the objects and events and their situations. With this in mind, Cummings and Turner (2009: 231) define grounded theory as “a methodology that aims to understand how individuals and groups interact, act and engage in response to phenomenon (i.e. objects, events and situations) they experience or encounter in their everyday lives”. It is within this context that theory is developed, which provides descriptions or theoretical analyses of a specific phenomenon. Strauss and Corbin (1998) maintain that grounded theory provides a methodological framework with practical guidelines for building and developing theories, or an in-depth theoretical analysis from the empirical evidence.

Kerlinger and Lee (2000: 9) define the word ‘theory’ as “a set of interrelated constructs (concepts), definitions and propositions that present a systematic view of phenomenon by specifying relations among variables, with the

purpose of explaining and predicting phenomenon”. Within the context of this study, the phenomenon is defined as the set of requirements for a data warehouse that are shared among users and which emerge from their needs, expectations and experiences of using DCM data. Grounded theory is, therefore, an applicable methodology for this study, since it facilitates reaching such phenomenon in a systematic manner.

According to Charmaz (2006: 507), the term ‘grounded theory’ refers to both a “method of inquiry” and “product of inquiry”, representing, therefore, a process and an outcome. As a process, grounded theory encompasses a set of techniques or principles to guide the research. However, as an outcome, grounded theory is concerned with the development of theory, which is achieved through using a systematic process of data collection and analysis. Monks (2006) argues that the outcome is entirely dependent on the process, since without a set of principles that guides the research process in a systematic fashion, the emergence of theory is not possible, thus highlighting the significance of grounded theory as a process of reaching a theory. A number of empirical and theoretical studies on requirement analysis report the use of grounded theory as a process to reach theories or in-depth explanations (user requirements in the context of this study). The primary aim of these studies is to explore an area where there was a lack of previous research and published literature (Urquhart 2000; Chakraborty and Dehlinger 2009; Jantunen and Gause 2014; Wang 2014). While using grounded theory as a process, many of these studies present their study outcomes (user requirements) as systems’ data models (Chakraborty and Dehlinger 2009;

Halaweh 2012) and as theoretical models that could explain the set of general

system requirements²¹ (Crook and Kumar 1998; Pace 2004; Kroeger et al. 2014; Wang 2014). While suggesting that it is best to remain open-minded about what could be the outcome in terms of a descriptive theory, or merely theoretical explanations of some parts of a phenomenon and how it could be presented, Strauss and Corbin (1998) endorse the idea of using grounded theory as a process to reach explanations that are grounded within the collected data. My primary intention was to use grounded theory as a process of reaching potential users' needs and expectations of their secondary uses of DCM data, which could then lead to a theoretical/conceptual explanation of what constitutes the requirements for the secondary use of DCM data within a data warehouse.

Strauss and Corbin (1998) argue that, through a systematic process of inquiry, the purpose of using grounded theory is to develop a theory or categories that are linked on a theoretical basis within the data that reflect the 'reality' of participants' social phenomenon. Kelly (2010) adds that such a theory provides an in-depth description of the interpretations of participants' experiences within their social settings. These careful interpretations are referred to as 'theoretical associations' of participants' realities rather than as an exact picture (Guba and Lincoln 1994; Schwandt 1994; Charmaz 2006) and are achieved through a systematic and rigorous process of data collection and analysis. As argued in Chapter 4, requirement analysis is considered as an interpretive process where users' needs, expectations and experiences are

²¹ The requirements are not only related to the information requirements that can be modelled into a system's data models, but also provide descriptive information that need to be considered at various stages of the system's development, implementation and adoption and can depict the system's overall success.

interpreted by the analyst (researcher) within the context of their emergence. It was also argued that requirements represent realities that are a product of one's own ability to see the current or future uses of a system. Strauss and Corbin (1998) assert that grounded theory, as a systematic methodology, enables the researcher to bring multiple realities together and explain them at a conceptual level. In this study it is the intention to interpret, based on my ability and knowledge as a researcher, as discussed in 'The role of the researcher' later in this chapter, multiple realities (requirements) as a combined truth (shared set of requirements), described at a conceptual level. Chakraborty and Dehlinger's (2009) study of identifying requirements for a university support system justifies the applicability of grounded theory for interpreting user requirements as categories which provide in-depth descriptions of requirements that reflect similarity in users' views, needs and expectations at a conceptual level. Based on the above, and several other, examples of using grounded theory for such purposes (Galal-Edeen 2005; Teixeira et al. 2012), it seems promising that, within this study, grounded theory can support a pursuit of mutually agreed requirements for secondary use of DCM data.

Furthermore, Charmaz (2006) asserts that grounded theory can provide an understanding of individual and organisational perspectives that reflects events and activities within a specific social context. Therefore, the techniques of grounded theory, as analytical tools, for understanding the organisational and social phenomenon from the individual's perspective, will allow the issues related to designing a new system for individuals and organisations, which can

be political, ethical or organisational, to emerge automatically through conversations with the potential users.

The literature highlights several versions of grounded theory, as it has gone through various modifications and developments since its conception by Glaser and Strauss (1966). The current study claims to employ a modified version of grounded theory. First, I will describe the process of grounded theory before moving on to argue the case of why its modified version was used and what constitutes 'modified' grounded theory.

5.6. Grounded theory process

Using grounded theory, various methods (i.e. interviews, observations, focus groups, content analyses) can be used for collecting data. The first interviewee, or the data source, is identified through preliminary review of the literature. The concepts that emerge from this literature are devised as a set of initial questions to start and later to focus the interview process. Grounded theory is conducted through an *'iterative and simultaneous process'* of data collection and analysis. This means that the researcher collects the first set of data (either through an interview, focus group or other means) and starts analysing it through coding the data. The coding involves giving meaningful names (labels) to the pieces of text from the data: this stage of coding is called open coding (Strauss and Corbin, 1998). Strauss (1987) calls these labels indicators, which represent the concept (building block of theory) within the data. These can be called codes (emerged concepts from the data), categories (merged concepts based on the similarities and differences) and sub-categories (concepts related to a category through

various properties and dimensions). The codes and categories emerge from the data and further

develop by establishing links or relationships between categories and subcategories using properties (characteristics or attributes of a category) (Gray 2009) and dimensions (the scope of a property at various extents) (Strauss and Corbin 1998). This stage of coding is called axial coding (Strauss and Corbin 1998). Examples, with reference to the study data, are given in Chapter 6, which describe the process of category emergence and its relationship with other categories and sub-categories.

Researchers make sense of the data through analysis activity, establish a conceptual vision of what is happening in the data and formulate some concepts that are obvious within it. Keeping these concepts in mind, the researcher locates further participants who can help to provide data on specific themes/categories. This is called '*Theoretical Sampling*' (Strauss and Corbin 1998).

As the data collection and analysis continues, the sampling becomes more in-depth and eventually comes to a halt (saturation point). Data is said to be saturated when no new information emerges from 'new' data collected during subsequent data collection in terms of generation of new categories. However, during the theory-generation process, the aim is usually to saturate the data theoretically; this is called '*Theoretical Saturation*'. According to Glal-Edeen (2005), the data is theoretically saturated when the same concepts emerge from the data repeatedly.

During the analysis process, data are constantly compared with other data and with the emerged categories for differences and similarities and

eventually classified into other categories. This process is called '*Constant Comparison*'

which is a fundamental analytical tool within grounded theory for making the analytical process rigorous and systematic (Glaser and Strauss 1999; Strauss and Corbin 1998). During the constant comparison process, several memos are also written as the researcher's reflexive notes on the analysis process. The memos provide a trail of evidence for decision-making during data analysis for further sampling, saturation and concept development during the analysis process. Writing memos throughout the whole process of data collection and analysis is a fundamental part of grounded theory. According to Glaser (1992), the researcher who omits the process of writing memos is not using grounded theory. Researchers need to analyse theoretically the whole process step-by-step through writing memos, which help them to theorise the codes and their relationship during write-up. Therefore, it is imperative to write memos whenever a new idea emerges during analysis or when a relationship is seen between categories or concepts that lie within the data.

'Asking questions' is another important analytical technique of grounded theory (Strauss and Corbin 1998). During the constant comparison process, various kinds of questions are asked regularly to refine and develop the concepts that have emerged from the data. The questions are usually asked by the person who is conducting the analysis at various stages, based on when they emerge (Strauss and Corbin 1998), in order to enhance the understanding of the concepts (Strauss 1987). These are called 'Sensitising Questions' which originate when the researchers become sensitive (perceptive) to data based on their knowledge of the area under study and

their background knowledge. They ask questions to understand the concepts in depth. The technique of asking questions from data is stimulated by the

concept of '*Theoretical Sensitivity*'. This refers to the personal quality of the researcher and their 'sensitivity' (Galal-Edeen 2005) in giving an appropriate meaning to the data in order to generate and develop the categories through theoretical understanding of the data. Sensitivity can come through, or be enhanced by reading the preliminary literature, the researcher's background or their professional knowledge (Charmaz 2006; Strauss and Corbin 1998).

The modifications within the grounded theory and my rationale for using a modified version are presented next.

5.7. Developments or modifications within grounded theory

This section will justify my use of modified grounded theory within this study. However, prior to that, I will describe the origin of grounded theory. Then I will analyse the literature that encourages modifications in the use of grounded theory by describing how various authors perceive these modifications. I will also discuss how the modified use within grounded theory has encouraged many information system researchers to use it as a methodology for understanding people's experiences, needs and expectations of system development and use (Urquhart 2000). By drawing on the presented knowledge, I will also provide my view of modified grounded theory and of how it is used in this study. The origin of grounded theory and its basic reasons are presented next.

In 1966 two sociologists, Barney Glaser and Anselm Strauss, developed grounded theory as a research methodology (now called the 'classic version') during their joint research project on dying hospital patients. They

published their methodology in their book 'Awareness of dying: the discovery of

grounded theory' (Glaser and Strauss 1966). Their key purposes for developing grounded theory were as follows:

- To ensure the rigour and quality within qualitative research.
- To show the research community that sometimes theories need to be developed before they are validated, particularly when there is a lack of any relevant literature in the field, thus introducing the inductive-research approach. This is in contrast to what happens in traditional research where existing theories or hypotheses are tested and validated.

While remaining consistent with their original purposes of developing grounded theory, Glaser and Strauss continued developing the methodology further, albeit in separate ways. For example, Strauss, in partnership with Juliet Corbin (Corbin and Strauss 2007), introduced a model for coding, relating and developing various aspects of data during analysis. Strauss and Corbin (1998) believed in 'developing' a theory in a systematic manner and provided a set of instructions on how to do this. While such guidance is acknowledged as extremely helpful for novice grounded theory users when dealing with research data (Urquhart 2000), Glaser criticises this as a 'forcing act'²² rather than an 'emergence'²³ of theory from the data (Glaser 1992). Glaser strongly advocates the 'emergence or discovery' of theory from data rather than the 'developing' of theory. This means that, while analysing the

²² According to Glaser, Strauss and Corbin's procedures force data into pre-conceived categories, which is apparent when they talk about identifying an event, about the contextual and conditional information to explain that event and about the strategies and consequences of those events.

²³ Glaser perceives grounded theory as a methodology of discovery, where categories are allowed to emerge from the data without there being any pre-conceived container to put them in, of the type seen in Strauss and Corbin's (1998) coding model.

data, a researcher does not bring any pre-conceived ideas, not even a set of instructions or any pre-determined models on how to code data. Such debates between Glaser and Strauss were the beginning of introducing the concept of modifications within grounded theory.

Glaser (1992), as one of the primary developers of grounded theory, views any further developments as changes to the original method, i.e. the 'classic' version of grounded theory. However, his co-developer, Strauss (1978), invited researchers from various philosophical backgrounds and disciplines to use the methodology and to modify it according to their area of inquiry. Strauss's suggestion encouraged others to perceive grounded theory differently (Seale 1999; Clarke 2003; Charmaz 2006; Bryant and Charmaz 2007). For example, Charmaz (2006), along with Bryant (2002) and Clarke (2005), perceives grounded theory as a constructivist process and asserts that a theory should not be 'developed' or allowed to 'emerge' (as asserted in Strauss and Glaser's views of grounded theory) but 'constructed' from the data. Charmaz (2006) asserts that, being consistent with grounded theory's philosophical origin in symbolic interactionism (as argued earlier), a theory is the product of those interactions that happen between the participant and the researcher. She also emphasises the role of the researcher being as significant as the role of the participants in constructions of reality that depict the phenomenon under study, thus emphasising the importance of the researcher's role. How the role of the researcher is viewed within this study will be discussed later in this chapter.

From a methodological perspective, the modifications are also seen in the practical use of various components of grounded theory, which are the

constant comparison of data with emerging theory or hypothesis, the simultaneous act of data collection and analysis, theoretical sampling, theoretical sensitivity and memo writing (these are described earlier in Section 5.6). Charmaz (2006) refers to these fundamental components as 'grounded theory guidelines'. While these guidelines sit within both classic and modified versions of grounded theory, their application is varied across various individuals and disciplines, as encouraged by Clarke (2003) and Charmaz (2006: 19). They both argue that, if the aim of research is to explore a phenomenon, various researchers from any discipline can use the grounded theory guidelines to achieve this. Charmaz also encourages researchers to "adopt and adapt" these guidelines according to the needs of their research. Strauss and Corbin (1998) propose using these components as procedural suggestions for conducting data collection and analysis, rather than as strict rules. These grounded theory components will be referred to as 'grounded theory guidelines' within this and the following chapters.

Denzin and Lincoln (1998: 164) relate the modification of grounded theory to the researcher's personality, skills and abilities. They argue that grounded theory's actual use in practice has varied with the specifics of the area under study, the purpose and focus of the research, the contingencies faced during the project and perhaps also the temperament and particular gifts and weaknesses of the researcher. They suggest that a researcher can adapt the use of methodology according to their own thought processes. They, however, insist that the adaptation in methodology does not refer to a change in basic grounded theory guidelines that can impact on the rigour of

the method, but to introducing one's own social and intellectual ideas into the analytical process.

They are referring here to the impact of the researcher's intellectual and background knowledge on the analytical process, called theoretical sensitivity within grounded theory, which can be achieved via interaction with the literature. How theoretical sensitivity is achieved in relation to this study is described in the literature review chapters 2, 3 and 4. However, the use of literature to achieve this is explained later in this chapter.

The inception, and most of the early uses, of grounded theory occurred within the social sciences field, but the literature indicates that its use is also becoming popular in many other disciplines, particularly within information system research, usually focusing on requirement analysis process, which is considered a social process (Kroeger et al. 2014). Urquhart (2002) argues that it is the flexibility and freedom of the grounded theory guidelines that have encouraged many to use them within the information system discipline. Based on her own experience of using a modified version of grounded theory within her study of understating the client-analyst communication process during requirement analysis, Urquhart (2002) recommends the use of grounded theory within information system research. Whilst many studies justify the use of grounded theory within information systems, where the aim is to understand users' experiences, needs and expectations about a system, Mavetera and Kroeze, (2009) caution the researchers to use grounded theory guidelines carefully according to their area of interest or discipline. This means being sensitive to modification of its use. It also means careful application, based on the underlying philosophical and theoretical demands of a particular discipline. A number of studies have reported a successful use of grounded theory for

requirement analysis, based on its philosophical and theoretical similarities, which are explained later in this chapter in Section 5.12.

The above literature suggests that there is not a definitive answer to what constitutes 'modified grounded theory', as many scholars agree that there can be various ways of conducting research using grounded theory. These can be dogmatic (Strauss and Corbin 1990; Strauss and Corbin 1998) or flexible (Clarke 2003; Gasson 2003; Charmaz 2006). Birks and Mills (2015) observe that studies use the term 'modified grounded theory' when they use the guidelines to identify categories and themes within the data for the purpose of description and exploration of a phenomenon only and are not claiming to generate theories. Based on Birks and Mills observation of what constitutes modified grounded theory, in this study I have adopted this approach, where I considered the use of grounded theory guidelines for conducting the data-collection and analysis process only. I perceive modified grounded theory as giving the researcher a set of guidelines and the freedom to use them as they find works best within their discipline, given the type of research data and the context such as limited sources of data, or lack of resources to collect more data for theoretical sampling.

The application of the grounded theory guidelines within this study, though, has been considered based on the situation within which the research was conducted, the type of data collected and the challenges faced during data collection and analysis. These components, their application and the challenges I faced and finally overcame are explained in Chapter 6. Alongside the grounded theory guidelines for data collection and analysis,

there are some important considerations, which are important to take into account while

conducting any version of grounded theory. These considerations are presented next, along with a critique of literature on how others have considered these and an assessment of how I have taken these views into account within this study.

5.8. Substantive area

The particular area of study or inquiry is called the substantive area. It is argued that researchers adopting grounded theory methodology do not begin the project with pre-conceived ideas/assumptions about the substantive area of interest where the developed theory will be applied (Strauss and Corbin 1998; Glaser and Strauss 1999). Rather, they start with a general area of study and systematically allow concepts to emerge from the empirical data in the form of a theory (Strauss and Corbin 1998). There is a difference of opinion between grounded theorists about how much knowledge of a substantive area a researcher should have prior to commencing the research project. The diverse views in this regard are presented in 'Use of the literature within grounded theory' in Section 5.10 of this chapter.

The researcher can possess knowledge of the substantive area through her/his professional background (e.g., when a software designer/analyst uses grounded theory for requirement analysis) and involvement in studies within the same area (e.g., if someone has been involved in a project related to requirement analysis or a data warehouse). However, it is imperative that the researcher recognises her/his substantive area. For this purpose, Mavetera and Kroeze (2009) suggest that researchers search for, and be clear about, their substantive area of study. They further warn researchers

who are not familiar with their substantive area of study that they can find themselves in

the chaos of 'multiple substantive areas' and end up not justifying any of these correctly and, therefore, not being able to claim to be using grounded theory.

The main aim of my research study was "requirement identification for a future DCM data warehouse" which reflects two important facets, DCM and a data warehouse. It is important to stress here that neither the data warehouse nor DCM is the substantive area, rather it is the combination of both. Put simply, this deals with identifying the requirements for the secondary use of DCM data that emerges from mappers' current and potential use of DCM, which can potentially contribute to designing the data structure/models for the DCM data warehouse.

Prior to this study in 2009, I gained knowledge and understanding of DCM, and of how it works, during a three-day DCM basic user training course I attended, delivered by the University of Bradford. During the same period, I gained knowledge of the data-warehousing concept through researching the general use of data warehouses and their technical role in managing data for secondary purposes. Further, as part of my MPhil study (Khalid 2010), where I proposed a data-management framework for DCM data, I also attended a data warehouse-specific system-design course, where I gained the essential skills to identify the user requirements and translate them into the warehouse data models.

The knowledge of DCM and data warehouses does not contradict the knowledge that I need to acquire through this study, namely, users'

perceptions, views and needs as requirements for a DCM data warehouse.

On the other hand, my understanding of data warehouses and DCM has

enhanced my theoretical sensitivity, which according to Strauss and Corbin (1998), can hugely improve the researcher's skill of interpretation and analysis of study data in terms of identifying and synthesising the findings to achieve the study aims. For example, within this study, my knowledge of DCM and data warehouses facilitated me in interpreting the user needs, expectations and concerns as requirements that will influence the warehousing of DCM data.

5.9. Use of research question

It is common in any traditional research study to commence with a clear and specific research question which reflects the researcher's specific area of study. In contrast, those who intend to use grounded theory are specifically recommended to commence with a general topic or area of interest (Glaser and Strauss 1968). Glaser (1992) argues that a specific research question restricts the natural emergence of theory and, thus goes against the inductive nature of grounded theory. Strauss and Corbin (1998), however, suggest commencing with some preliminary hypothesis or theory that will guide the researchers and define the scope of the study. They argue that, without any such hypothesis, the researcher could have many aspects to consider during a single project, which might jeopardise the quality of inquiry.

However, they caution researchers to develop a wide question, which is as general as it can be, to let the exploration happen. Charmaz (2006) also asserts that the research question in grounded theory should be free (as much as possible) from the researchers' pre-assumptions about the general area of the study. In agreement with Strauss and Corbin (1998), Charmaz

(2006) suggests that the researchers should commence their study with a general

research question at the outset. Marvetera and Koereze (2009) consider such action important in particular cases where the researchers are familiar with the area under study and possess some background knowledge.

Within this study, I devised a research question at the outset, which was general, presented the wider study area and reflected the study aim. The devised research question was as follows: “What are the requirements for the potential secondary use of DCM data from a data warehouse from its potential users’ perspective?”. This question reflects my intention of identifying the potential users, as well as the uses, of the warehouse. Further, the aim was to construct a question that is as general as possible and allows the inductive exploration of the user needs and requirements. However, the research question can become more specific and targeted, or to use Strauss and Corbin’s (1998) word, ‘developed’ as the study goes through the data-collection and analysis process. Within this study, the above research question remained the same during the data-collection and analysis process. However, a set of study objectives was also identified (as presented in Chapter 4) to set the scope of the study.

5.10. Use of the literature within grounded theory

There are various discussions, or schools of thought, on using literature prior to the research process when adopting grounded theory within a study. Glaser is against reading literature that gives any information or knowledge on the research area. Mavetera and Kroeze (2009) challenge this, as they stress that it is impossible for the researcher not to have any knowledge of the area when she or he has expertise in, or has been related to, the area for some time.

They add that there is always something new and unexplored that needs a different perspective within the same area.

Strauss and Corbin (1998), however, encourage researchers to use the literature (not in a substantive area) at the beginning of the study in order to support formulation of the study design and general research questions. They argue that reading some related literature can provide directions and dimensions, which help the researcher to think from various angles and explore the problems in-depth by increasing analytical understanding (Strauss and Corbin 1998). Further, this helps in enhancing the researcher's theoretical sensitivity and rationalising the purpose of research (Gasson 2003; Mavetera and Kroeze 2009). Strauss and Corbin's advice on conducting some review of the 'not substantive area', however, can only be taken if a researcher defines her/his substantive area of study at the outset (Mavetera and Kroeze 2009), as argued above in Section 5.8.

Dey (1999) also warns researchers not to enter the field in ignorance. Researchers need to have some knowledge about the area under study, but they should not allow their pre-conceived ideas to be brought into the study as these can contaminate the emergence of the original theory or theoretical concepts from the empirical data (Glaser 1968). However, putting pre-conceived ideas to one side is not easy. It requires significant effort from the researcher to be objective while conducting the grounded theory process. Charmaz (2006) and Urquhart (2000) suggest that, in order to achieve some objectivity, the researcher should delay the 'relevant literature' review as much as possible. How I managed to put aside my pre-conceived ideas and assumptions within this study is explained in this chapter in Section 5.11.

In the context of this study, there was no literature available on the substantive area (identification and exploration of the requirements for the secondary use of DCM data from mappers' perspective) and therefore this study commenced without establishing any preconceived ideas borrowed from the existing literature. Further, within this study, review of relevant literature related to DCM data management was not possible due to the following two reasons. First, as this study is the first of its type within the area of DCM data management, there was limited literature available to review. Therefore, a traditional way of reviewing the research area using relevant literature at the beginning of the study was not possible. Second, the aim was to let the user requirements emerge from the empirical data rather than from the literature, thus bringing in study participants' views as requirements at the forefront.

During this study, the literature was reviewed at various stages. The first review occurred during the establishment of the general research aim, while understanding the existing work within the field and establishing a rationale for the study, including valid aims and objectives. This review forms part of what is presented in chapters 2, 3 and 4. During this period, the literature was also reviewed to identify and justify the effective philosophical and methodological considerations taken for this study, as presented in Chapter 5. The second review was performed during the analysis process in order to saturate and validate emerged categories (Charmaz 2006; Strauss and Corbin 1998). Such reviews contextualised the study findings and enhanced the discussion. These literature reviews are presented in chapters 7, 8 and 9. Some of the literature review conducted during data analysis formed also

part of chapters 2, 3 and 4 to define and explain the concepts emerged within the study data.

5.11. The role of the researcher

The researcher plays a significant role in the whole research process, particularly during data collection and analysis. Denzin and Lincoln (1998) refer to the researcher's role as an interpreter who tries to understand the natural phenomenon under study and, therefore, holds some responsibility for being part of the process by bringing in her or his own perspectives. During the use of grounded theory, such perspectives are considered significant in shaping researchers' interpretations. In this context, Charmaz (2006) states that the researcher and the study participants together construct the reality, which holds shared meaning and shared interpretation of the real world under study.

In this study, the role of the researcher was of great importance. The researcher worked as an 'analyst' and possessed technical and background knowledge of the data warehouse and its design methodologies, as well as the way it worked. Patton (1987) argues that the subjectivity of the researcher, which reflects her/his background and discipline knowledge, cannot be entirely ignored in the research process. However, the research suggests that the impact of the researcher's subjectivity can be minimised in both data collection and analysis stages (Guba and Lincoln 1994; Strauss and Corbin 1998; Patton 2002).

Within this study, to reduce this impact, during the data-collection process, open-ended questions were asked and plenty of time was given to the participants to answer the questions during the interview process. My background knowledge enabled me to be 'sensitive' to the concepts that emerged from the data and, therefore, to modify the interview guide and the

prompting questions in the semi-structured interviews to guide participants' in the right direction, to ask the right questions and to probe the interviewees to get the required information. In this process, the researcher's existing knowledge is used as a tool to facilitate in guiding the participant. This process, in relation to this study, is presented in Chapter 6.

The grounded theory literature acknowledges that the researcher brings her/his subjective perceptions to data collection and analysis, which can result in the rigour of the data-collection and analysis process and the reliability of the findings being questioned (Gasson 2003). Therefore, grounded-theorists suggest that researchers be reflective of the processes. The researcher's reflections about the data-collection and analysis process is captured in memos. Further, it is also advised to describe the reflexivity²⁴ while writing each stage of the data collection and analysis, which means describing why and how the decisions were taken by the researcher during theoretical sampling and making constant comparisons. This shows the rigour in the process and acts as an indicator to the reader with which to judge the quality of the study findings. Within this thesis, my reflections on the process are provided while explaining sampling and the data-collection and analysis process in Chapter 6. Further, wherever possible, I also provide a rationale of the decisions taken while presenting the findings in chapters 7, 8 and 9.

²⁴ Reflexivity is the process of questioning your own beliefs, attitude and presumptions during the research process. According to Malterud (2001: 484) reflexivity is "an attitude of attending systematically to the context of knowledge construction, especially to the effect of the researcher, at every step of the research process".

5.12. Philosophical and methodological similarities between grounded theory and requirement analysis

The literature indicates a number of studies that use grounded theory within information system research (Urquhart 2000; Kaplan and Maxwell 2005; Mavetera and Kroeze 2009). Further, the studies also specifically report their use of grounded theory for requirement analysis (Chakarborty and Dehlinger 2009; Pidgeon et al. 1991; Halewa 2012). One of their argument is that there exists philosophical and methodological similarities between grounded theory and requirement analysis processes. For example, Chakarborty and Dehlinger (2009) use grounded theory methodology to identify system requirements embedded within an enterprise architecture. They argue that system-requirement elicitation and grounded theory share similar methodological steps and that grounded theory can facilitate a systematic process of reducing a large amount of qualitative data into system requirements. Pidgeon and colleagues (1991) describe how grounded theory and requirement analysis share not only methodological similarities but a philosophical background as well. Based on their study of the use of grounded theory for knowledge engineering, they maintain that both grounded theory and knowledge engineering deal with qualitative data, both trying to understand behaviours through contexts. They also argue that human behaviour is linked to the context within which it appears and grounded theory allows understanding of this contextual knowledge in order to interpret the users' needs, experiences and expectations (Pidgeon et al. 1991).

As mentioned in Chapter 4, the aim of requirement analysis is to obtain requirements and then develop these in terms of their accuracy, completeness

and relevance and this is done through a cyclic process of data (requirement) collection and analysis; this is the typical process in grounded theory, as described above in Section 5.6. Further, the practical and research evidence shows that various methods of data collection are used in requirement analysis, such as interviews, observations, focus groups and document analysis, as mentioned in Chapter 4. This is a standard process in grounded theory. Furthermore, recognising potential users to develop the requirements further is an essential step towards achieving the detailed set of requirements, a process called theoretical sampling in grounded theory. Chapter 4 also stresses the role of the researcher as an analyst who uses her/his knowledge in understanding user needs and interprets them into future system requirements. In grounded theory, the researcher uses her/his theoretical sensitivity to develop an understanding of the emerged findings and presents this in a form that can best reflect the situation or phenomenon under study.

The above comparison suggest that requirement analysis and grounded theory share philosophical and methodological similarities. However, grounded theory provides a rigorous and systematic process of analysing data through coding, which produces an in-depth version of findings (requirements) rather than a 'surface-level' description (Kelly 2010). Corbin and Strauss (Corbin and Strauss 2008) claim that grounded theory guidelines can provide a theoretical framework that contributes towards exploration of a core concept/issue/phenomenon within the data, one that is relevant across various groups of participants and can be explained in detail through sub-categories. These sub-categories provide an explanation in

terms of the context within which that phenomenon emerges, the conditions that influence that

phenomenon and the actions and strategies adopted, or suggested, for dealing with the phenomenon.

5.13. Summary of the chapter

This chapter has described the three main steps followed to design the research process. It commenced by explaining the philosophical position adopted within the study, which argued the use of the interpretivist approach to identify and explore the users' requirements for the secondary use of DCM data as their own interpretations of the potential data-warehouse system. It then outlined the use of grounded theory, underpinned by a qualitative approach. A detailed explanation of grounded theory was then presented, followed by a discussion of its modifications. The chapter outlined a number of perceptions for modifications in grounded theory and argued that there is not a definitive answer to what can be called 'modified grounded theory'. Based on this, it further argued that a grounded theory methodology is modified when the researcher uses the main guidelines of data collection and analysis according to her/his own discipline and research data and the context within which the research takes place.

The chapter then discussed various important aspects of grounded theory, which are important to consider for an effective use of the methodology. The discussion included the significance of the identification of a substantive area. It further highlighted how I recognised that, within this study, both DCM and a data warehouse together form the substantive area, which described the DCM mappers' perspectives of secondary use of DCM data within the warehouse. The chapter then examined the use of a research question in

grounded theory studies, where the significance of a broad and wide question or topic area was

highlighted. Further, it also highlighted the use of a research question and its development within this study.

This chapter then discussed the existing arguments for the use of the literature within grounded theory studies and highlighted the reasons and timings when specific literature was reviewed within the present study. Finally, the chapter discussed my role as the researcher in the process of designing the study and generating and interpreting the study data. It discussed how a researcher's subjectivity, which should be acknowledged within a qualitative study, could be minimised in both data collection and analysis. It further highlighted the actions taken to minimise the researcher's impact on various aspects of data collection and analysis within this study. The chapter then closed by comparing grounded theory and requirement analysis based on the philosophical and methodological similarities the two processes hold in identifying the requirements as users' perceptions, views and expectations of a new system.

In the next chapter, the thesis demonstrates the practicality of the use of grounded theory guidelines for data collection and analysis.

6. Using Modified Grounded Theory; Data Collection and Analysis

"A researcher's background and position will affect what they choose to investigate, the angle of investigation, the methods judged most adequate for this purpose, the findings considered most appropriate, and the framing and communication of conclusions" (Malterud 2001: 483-484).

6.1. Introduction

This chapter presents the fourth step of the research process involving data collection and analysis methods. As has been discussed in Chapter 5, the guidelines of grounded theory were used as a framework for data collection and analysis in this study. This chapter demonstrates the practicality of the grounded theory guidelines in detail, using insights from the study data.

As mentioned in Chapter 5, using grounded theory means that data collection and analysis are conducted simultaneously; the researcher conducts some initial data collection for analysis purposes to guide further data collection and analysis. It was also mentioned that, during this process, theoretical sampling takes place, which guides what, when and who to recruit for further data collection, thus weaving the sampling process into the data collection and analysis. For the purpose of clarity however, within this chapter, all three processes (e.g. sampling, data collection and analysis) are presented under separate headings.

The chapter begins by examining the sampling techniques most often used in grounded theory studies. Through a consideration of these sampling techniques, the process of recognising, identifying and recruiting participants for this study is presented. I then examine the data-collection methods used

within grounded theory studies, particularly within information-system studies, and present my rationale for using my chosen methods of data collection. Next, the empirical process of data collection within this study is presented.

Additionally, this chapter describes data analysis methods underpinned by the grounded theory guidelines and their applications within this study. In order to show transparency and rigour, each analysis step is justified and rationalised with examples taken from the study data. The chapter ends by explaining the steps taken to maintain the trustworthiness and credibility of the data collection, analysis and findings presentation process.

6.2. Sampling; recognising and identifying participants

Grounded theory advocates the use of theoretical sampling. As already defined in Chapter 5 (Section 5.6), theoretical sampling involves identifying data sources based on the concepts that emerged from the study data. Strauss and Corbin (1998: 201) maintain that such data sources could be “places, people, or events” that could provide information to maximise the chances of exploring variations in the emerged concepts within the data, thus providing flexibility about what and who to approach for the required information. Halaweh (2012) asserts that such flexibility of theoretical sampling works extremely well for requirement analysis as the researcher has the freedom to identify and then explore the requirements from various sources, including potential users, documents and existing systems.

Theoretical sampling can only be conducted when one has collected and analysed an initial set of data. The question, however, remains: ‘Where to

start to get the initial data?' Coyne (1997: 625) argues that, using grounded theory,

“the researcher must have some idea of where to sample, not necessarily what to sample for, or where it will lead”. In this case, Breckenridge and Jones (2009) suggest that theoretical sampling should involve an element of purposeful selection of data sources to commence an initial idea. A number of grounded theory studies therefore commence with ‘purposive sampling’, defined as the selection of participants who are considered to have shared knowledge or experience of the particular phenomenon under study, which has been identified by the researcher as a potential area for study (Sandelowski 1995).

Strauss and Corbin (1998) suggest some considerations that could help and guide the researcher to commence purposive sampling for initial data collection. They maintain that, based on their area of interest, the researchers need to know the site or group to study, the types of data that need to be collected, the length of time that an area should be studied for and the number of sites she/he needs to access. They caution, however, that all these considerations should be taken based on the researcher’s preliminary understanding of the area of interest, the accessibility and availability of resources and researcher’s time and energy (Strauss and Corbin 1998).

Within the context of this study, I used purposeful sampling conducted in two phases. The first phase was informed by what was argued in Chapter 4 (Section 4.9.1), that DCM mappers trained at any level and had some experience of mapping could be the potential data warehouse users as they would be familiar with the DCM data. This formed the first criterion for choosing the study participants.

Strauss and Corbin (1998) maintain that theoretical sampling evolves during the process of data collection and analysis. In line with theoretical sampling, the second phase commenced following analysis of data from the first few interviews (the analysis process is explained in Section 6.10 of this Chapter). During the analysis, based on their existing uses of DCM, three mapper roles emerged. These were as follows: DCM researchers, being those who were using DCM for research purposes; DCM practitioners, being those who were using DCM for practice development; and DCM trainers, being those who were using DCM for training purposes. The second selection criterion, therefore, was to recruit mappers from the above three categories to provide information about various potential secondary uses of DCM data and associated issues and concerns. There was some overlapping of roles, however, as some practitioners and researchers were DCM trainers as well. Similarly, some practitioners were also using DCM for research purposes. Therefore, at any given time, a mapper could be a DCM practitioner, researcher and trainer.

As the process of data collection and analysis progressed, I also adopted snowballing sampling, considered one of the suitable sampling approaches within grounded theory (Glaser 1978). For this purpose, I asked each participant to recommend a number of further people who would potentially be interested in sharing their views regarding their current and potential use of DCM data. Where the interviewee identified other potential participants, I asked if they could introduce me to them so that I could ask them to participate. However, the final selection of these suggested participants was based on the previously discussed criteria, which required the participant to

be a mapper, using DCM either for research, practice development or training purposes. The

snowballing sampling approach did work in many cases where participants were able to identify potential participants within their own organisations or those they thought would have been able to comment about their uses of DCM data.

Whilst the identification of study participants was based on specific criteria, my recruitment strategy was conducted purely on a convenience basis, referred to as convenience sampling (Patton 2012). This means that, as a UK-based individual it was convenient for me to recruit UK-based mappers. While this decision worked for DCM practitioners and trainers, the UK-based researchers using DCM limited the pool of potential participants. Therefore, I made a decision to recruit any researcher from any geographical location, who was interested in taking part in this study because, as asserted and encouraged by Strauss and Corbin (1998), convenience sampling is usually the most practical way of collecting data.

Requirement analysis is conducted with a sample of intended or potential users, as it is not feasible to gather requirements from all anticipated users (Maguire and Nigel 2002; Nies and Pelayo 2010). Therefore, not all organisations and individuals who use DCM or who may be potential data-warehouse users could be included in the study because of time and interview number limitations. As designing and developing the data warehouse is an on-going process (Davis et al. 2006), the requirements gathered at this stage will explore the secondary uses of DCM data, which can potentially be used to design the first prototype of a data warehouse. Data warehouses are usually encouraged to be designed in such a way that

they can be easily modified by adding new requirements taken from other organisations and users in the

future (Davis et al. 2006). The requirements gathered from the current sample organisations and individuals will ensure that the basic structure of the data warehouse can be enhanced gradually by adding more organisations and potential participants' requirements within its design at a later stage, if required.

Initially, the practitioners from a range of UK-based organisations involved in dementia-care provision – monitoring, regulation and research – were contacted to take part in the study. These organisations and individuals were identified through existing networks established with the School of Dementia Studies, at the University of Bradford, as well as through my supervisors' and my own knowledge of key individuals or organisations within the field, gained by networking with mappers during DCM training and at conferences and meetings attended during completion of my previous studies. I also used published research as a way of identifying potential participants. In addition, practitioners and researchers were also contacted via an existing database of trained mappers who had agreed for their contact details to be passed on to others, held by the School of Dementia Studies at the University of Bradford.

6.2.1. Recruiting

Potential participants were sent personalised emails inviting them to take part in the study. This introduced the study, its aims and objectives and my connection with the University. An information sheet (see appendices 3 and 4) was also attached with the email in order to provide a detailed overview of the study, its purpose, procedures, potential benefits and risks and data-confidentiality issues. While providing an overview of the study, an

introduction to a data warehouse concept for DCM data was provided in a language that

would be understandable to the layperson. The information sheet also described the role of participants within the study and the potential benefits or disadvantages for them or their organisations of taking part. They were also reassured about the security and anonymity of their data. Potential participants were informed that replying to the email would indicate their informal agreement to participate in the study and told that a formal consent form (see appendices 1 and 2) would then be sent to them to be completed and returned via email or post before the commencement of data collection. Those who replied and agreed to take part in the study were contacted again by email or telephone to arrange a time for interview (face to face, or via Skype or phone) at a location, date and time convenient to them.

However, those who did not reply within a given timescale (two weeks) were approached again by email. If they replied and agreed to participate, they were contacted again for the interview; if not, they were not approached again. Different information sheets were prepared for the various types of participants (researchers, practitioners and trainers) and written in a language appropriate to them (see appendices 3 and 4).

6.2.2. Number of participants

Like any other qualitative research, the aim of grounded theory is not to 'generalise' but to present a specific area of study from a specific group of people (Strauss and Corbin, 1998). Polit and colleagues (2001) assert that, in this case, there could be any number of study participants, who might not be representative of the larger population under study but might provide enough information to explain the area under study. While using grounded theory, Strauss and Corbin (1998: 214) assert that the concern should be

“representativeness of concepts and how concepts vary dimensionally”. This means that the number and the representation of participants depend on the nature of the concepts emerging from the data. Strauss and Corbin (1998) assert that even a single interview can provide multiple examples to explain the concepts. However, they suggest that, if subsequent interviews do not add anything new, they still add validity to the concepts already identified, leading eventually to data saturation. In total, 29 participants were recruited to take part in this study to identify and saturate the emerged concepts and categories within the study data. The study participants’ details are provided in appendix 5. The saturation criteria are explained later in this chapter.

6.3. Ethical approval

In order to collect data from mappers, it was a legal requirement to gain ethical approval from a recognised body. Considering that the mappers usually are from health and social-care organisations, most of which work for the UK’s NHS (National Health Service), its guidelines for ethical approval were considered (NHS 2015). However, since September 2011, researchers are not required to obtain ethical approval from an NHS ethics committee should they only wish to recruit NHS staff members, either from social or from health care, as participants in research studies. This exempted me from applying for NHS ethical approval for this study. However, in order to gather data from any human subject, the University’s ethical approval was required and, therefore, obtained from the Humanities, Social and Health Sciences Research Ethics Panel at the University of Bradford at the outset of the study. This was to ensure that all ethical concerns would be taken into consideration in terms of

recognising, approaching and recruiting participants, as well as gaining their consent and collecting data from them.

6.4. Data-collection methods

As discussed in Chapter 4, there are several methods used for requirement elicitation, including one-to-one interviews, group discussions or focus groups, observations and document analysis. It was also highlighted that the researcher chooses the data-collection method based on various factors such as user availability and time, resource availability and the potential system's needs (Abai et al. 2013). Grounded theory as a methodology provides the flexibility to choose any data-collection method that could provide the required information to answer the research question. The grounded theory literature highlights that interviews, particularly semi-structured interviews, are the most widely used method of data collection (Jones and Alony 2011). Interviews enable the researcher to inquire into participants' knowledge, experience and behaviour regarding the phenomenon under study, or to identify the new phenomenon from the study data, while at the same time empowering the researcher to guide the inquiry process in the right direction (Goulding 2002). Given that this study required the exploration and identification of users' needs, understanding and perceptions in relation to their current and potential use of DCM data, interviews were therefore chosen as the primary method of data collection. Further, as argued in Chapter 4, interview techniques, particularly semi-structured interviews, usually work effectively when users might not be familiar with their own requirements, as such techniques can provide an

opportunity to explore, probe and verify their views, experiences and expectations regarding the new system while interacting with them

directly. Given that a data warehouse is a new system/concept for DCM mappers and that there is a possibility that potential users might not be familiar with their requirements, semi-structured interviews were an appropriate method for data collection.

Focus groups were also considered an appropriate data-collection method. Focus groups provide an opportunity to explore a topic from various perspectives and group members might be part of the same community or organisation (Kitzinger 1995). In order to attain mappers' views regarding their organisational use of DCM data, I decided to explore the DCM practitioners' views using a focus-group method. The intention was to conduct a focus group within a single organisation, should more than four participants from the same organisation agree to take part in the study. For this purpose, the information sheet sent to the practitioners included the option for participants to take part in a focus group. However, on the return email of the informal consent, there were no more than three participants who were interested from a single organisation. As the criteria for forming a focus group state that there should be a minimum of four participants (Kitzinger 1995), it was only possible to conduct one-to-one interviews.

6.5. The development of the first interview guide

Whilst studies encourage the use of interviews, particularly semi-structured, for exploring participants' needs and experiences within grounded theory, the emphasis has been on using an interview topic guide rather than a structured questionnaire with a set of pre-identified questions (Gasson 2003). Strauss and Corbin (1998) suggest that, in order to encourage the exploratory

process, it is important to have broad topics to ask the participants about and to help

them to talk about these in detail, with appropriate use of probing questions by the researcher. Within this study, therefore, a general interview guide (appendix 6) was prepared to facilitate the interview process.

The interview guide covered the general topics related to the study participants' current and potential uses and management of DCM data and related issues and concerns. The choice of these topics was based on two suppositions. The first that the study participants should be asked about the data that need to be stored within the warehouse rather than the system that will store the data (Chapter 4). The second was that, as a data warehouse is a new concept/system for DCM, the study participants might not be familiar with their requirements and therefore the focus should be on what they currently know about DCM data and its use and their perceptions (if any) of its potential use for secondary purposes within the warehouse.

The interview questions were tested for their practicality during pilot interviews. These are discussed in detail in Section 6.6. The people interviewed during the piloting phase were asked to comment on the topics and questions within the interview guide. Based on their feedback, I modified my interview guide (see appendices 7 and 8). This resulted in more open-ended and general questions.

As data were collected through various iterations (as will be described in Section 6.7 of this chapter), the interview guide was modified alongside the process of data collection and analysis to add new questions and topics that needed further exploration. For example, the topic of 'data quality', related to

the DCM data, was added in a modified interview guide, during the subsequent iterations of data collection.

6.6. Pilot interviews and development of interviewing skills

Strauss and Corbin (1998) assert that many novice researchers experience a lack of confidence in commencing the grounded theory process, usually due to the lack of any lead regarding where and how to begin. The major aims of piloting are to practise interviewing skills, to formulate and refine the interview guide and questions, to gain confidence during the interview process and to gain some initial experience of conducting data analysis (van Teijlingen and Hundley 2001; Kim 2010).

Within this study, I also conducted pilot interviews with two mappers (a DCM trainer and a DCM researcher). The data collected during these interviews was not used as study data. The main purpose of the pilot interviews was to practise my interviewing skills and style and to test the practicality of the interview questions. During these interviews, I identified the researcher's bias and the interview question bias, as described in Section 6.11.1 of this chapter. The volunteer interviewees were also asked to comment on my style of interviewing. Based on their feedback and my own reflections on the process and on listening back to the interview recordings, I also changed my style of interviewing by concentrating more on listening carefully and probing the participant further, to support them to explain important points in more detail, rather than thinking about what I needed to ask next. This piloting ensured that I felt confident during the actual data-collection process, as explained next.

6.7. Data-collection process

According to the grounded theory guidelines, data collection is conducted in segments or phases to allow analysis and to guide further data collection. Strauss and Corbin (1998) suggest that such a data-collection process enables the researcher to analyse the data in segments to draw various initial hypotheses. Further, during data collection, the researcher examines the emerged hypotheses against the research aims and objectives and against the relevant literature. She/he then compares and contrasts these hypotheses with subsequent data collection and then examines again the new data against previous hypotheses and aims and against the objectives of the study. This cycle of data collection comes to a halt when the researcher takes a pragmatic decision that the emerged hypotheses are making sense in terms of explaining a phenomenon (user requirements) within the study data. This point is called saturation and is explained in the end of this section where I discuss the criteria of saturation established for this study.

Within this study, the interview process was divided into three phases. During the first phase, which commenced soon after ethical approval was gained, 10 semi-structured interviews were conducted with mappers to explore with them their current and potential uses of DCM data. This phase was exploratory and inductive in nature. It commenced with purposive sampling and the data was analysed soon after its collection. The main aim of this phase was to set the scope of the warehouse in terms of who will potentially be using it and for what purposes. The significance of setting the scope of a data warehouse has been argued in Chapter 3. One of the main

reasons for doing this was to streamline the further requirement elicitation process. During analysis, a number of

recurring elements emerged from the interview data and were identified as important concepts to explore further during subsequent interviews. The analysis process is explained in Section 6.10.

During the first phase of data collection and analysis, one of the potential uses of DCM data for secondary purposes that the study participants' identified was associated with related concerns, issues and types of required data. This caused the emergence of concepts, which required further exploration. Gasson (2003: 83) advises the researchers to "carefully note the emergence of insights and explicitly reflect on how these insights are bounding the research problem through selecting some categories and not others". This means that the researcher can choose the categories, which she/he thinks need further exploration to give depth to the area under study, and which are relevant to answering the research question and achieving the aims of the study. In the context of this study, only those categories were chosen for further exploration, which could provide an in-depth understanding of the data requirements related to identified potential use of DCM data.

The second phase of data collection commenced with conducting semi-structured interviews with 14 more participants (e.g. DCM trainers, researchers and practitioners) in order to collect more focused and detailed data. After these interviews, some literature was reviewed, based on concepts emerging from the data, such as data quality as an issue for secondary use of DCM data and lack of effective data-management systems for primary use of DCM data. The aim of this review was to enhance my theoretical sensitivity to enable me to understand users' views, expectations,

concerns and 'directly expressed requirements' in the context of a data warehouse.

Further, during the third phase of the data collection process, more participants (DCM researchers, n=5) were recruited to explain and clarify the requirements that emerged. Both subsequent phases two and three of data collection were more structured and were guided by a focused and theoretical process of data collection following the open and explorative style. Glaser and Strauss (1968) and Strauss and Corbin (1998) call this stage the deductive approach to gathering data. This is where a researcher goes into the field and collects data to develop and validate the 'hypotheses' that emerged from the data collected during the exploratory stage (phase one within this study) which encompasses the first set of interviews. Therefore, the interview guide was modified during phases two and three by adding new topics that facilitated the deductive process of data collection. The new topics were reflective of the themes that emerged during the first phase of data collection. The modified interview guides developed during the second and third data-collection phases. During these phases, the aim was to develop the categories that emerged from the initial set of data but also to keep an open mind to identify new concepts that could either contribute within existing categories or lead to developing new categories (Strauss and Corbin, 1998). The following criteria were followed for data saturation:

- When no new concepts (requirements) were emerging from the interview data;
- When enough examples were gathered to explain each category from various dimensions;

- When it was evident that enough explanation was available to explain the objectives of the study, that is, the requirements for the secondary use of DCM data from the perspectives of its users.

In total, 29 participants were interviewed. The choice of interview mode was based on the participants' preference and convenience. Nineteen of the 29 participants were interviewed through face-to-face meetings. Six interviews were conducted using a video- and audio-conferencing facility (Skype). Four interviews were conducted using only an audio-conferencing facility (telephone). The interviews ranged in length from 45 to 60 minutes and were audio-recorded. Each participant was interviewed only once and the interview was conducted at their work place, if interviewed face-to-face, and at a date and time of their choice.

An informal style was adopted; this was a conversational style with no strict sequential set of questions, thus encouraging a free-flowing conversation rather than an overly formal interview. This encouraged participants to talk about aspects within the main topics that concerned them, rather than the ones presented by the researcher (Gray 2009). However, the questions from the interview guide were used as guidance and were adapted according to the interviewee's responses and the flow of conversation over the period of data collection. This helped to establish a bond of trust with the participants (Elsheikh 2011) and allowed further exploration of topics that were relevant in terms of answering the research question. This informal and conversational style encouraged participants to talk about their current use of DCM data, their needs and expectations of its further use and relevant issues and concerns.

Gray (2009) maintains that, within qualitative semi-structured interviews, the informal conversations may seem irrelevant to the study objectives at first; however, they can provide a useful context for the categories that emerged within the data, supporting knowledge of the requirements and the context within which these conversations emerged. The importance of understanding contexts (human and organisational) during the requirement analysis process has already been described in detail in Chapter 4. In order to understand the contexts of the requirements, it was important to give importance to having entire conversations without interruption.

6.8. Audio-recording and interview-transcription process

All the interviews were audio-recorded using a digital audio voice recorder and then transcribed by myself. During the manual transcription process, each interview tape was played back and forth several times to check the accuracy of the transcribed text. The average audio recording was 50 minutes long. Hansen and Kautz (2005) consider self-transcribing a good practice and a valuable experience for the researcher, as they maintain that such a process can contribute to enhancing and adding in-depth knowledge to the researcher's understanding of the data, thereby facilitating further interpretations. In interpretive studies, this is another way of making the researcher sensitive to the data and of gaining more related interpretations from the data (Strauss and Corbin 1998).

As recommended by Hensen and Kautz (2005), to preserve my understanding of the data during the transcription process, the main points that emerged from the interview conversations were recorded as a summary

within a memo for each interview. The interview summary was based on my understanding of the

data, where I highlighted the main issues emerging from each interview. These summary memos (see an example in appendix 9) enabled me to understand the main concepts and themes within each interview and this further facilitated the focused analysis of the data, as presented in Section 6.10.2.

Each interview transcription was anonymised by changing the details that might lead to identification of a participant. Each participant was given a pseudonym, which was used throughout the presentation and the discussion of findings during the write-up. These pseudonyms included the role of the participant (researcher, practitioner or trainer, while describing their current and potential uses of DCM data) followed by a random number, for example, practitioner 1, researcher 9, trainer 22 etc. Further, in order to refer to the participants during the findings and discussion (e.g. chapters 7, 8 and 9), the joint pronoun 'her/him' is used to ensure as much anonymity as possible. This was considered necessary since there are far fewer male than female mappers particularly in some types of user category e.g. DCM trainers (Personal Communications 2013).

6.9. Interview data-management process

Managing and analysing qualitative data is a time-consuming and strenuous process. However, advances in computer technology have facilitated this process by introducing Computer-Assisted Qualitative Analysis Software (CAQDAS) (Bringer et al. 2006). Johnston (2004) argues that, while researchers can accelerate the analysis process through using software, they might not fully possess the understanding of the technique required for in-depth analysis. Kelle (1995) further maintains that the role of the

researcher as an interpreter of the data can diminish when using software.

However,

Bringer et al. (2006) and Johnston (2004) claim that, if software is used appropriately, the process of data management and analysis is not merely accelerated but can become a learning process for the researcher. For this purpose, Richards et al. (2004) suggest that the researchers look for software that works for their data rather than trying to fit their data to a particular software package. There are a limited number of existing software packages available for qualitative data analysis. QSR Nvivo, ATLAS ti and NUDUS are three such software packages. The QSR Nvivo 9 was the only software provided by the University of Bradford to PhD students for qualitative data analysis, which further limited my choice. However, Bringer et al., (2006) assert that QSR Nvivo can facilitate the analysis of data using grounded theory techniques. Therefore, within this study, I used QSR Nvivo version 9 for managing the interview data.

Soon after each phase of data collection, I uploaded the transcribed interview data into the QSR Nvivo 9 and commenced analysis (the data-analysis process is explained in Section 6.10.2). The interview data collected during each phase was stored in its labelled folder with reference to the specific phase of data. The appropriate labelling of folders enabled me to remain aware of various phases of data collection. As the data-collection and analysis process progressed, the folders increased in number, as new interview data was stored along with relevant literature and identified concepts. While the software was useful for managing the interview data and for generating memos at various stages of data analysis, there was limited scope for comparing the codes and categories that emerged during each

phase of data collection. Therefore, I preferred using manual techniques for data-

comparison purposes. For example, I used an A3 sheet to write down all concepts that emerged during the first phase of interviews with a coloured pen and I then wrote down the concepts that emerged during the second phase of data collection with a different coloured pen. I compared the concepts in order to see similarities and differences and continued this process until all data collection and analysis was completed and the final set of categories had emerged from the data.

6.10. Data analysis

Similar to the data-collection process, which was conducted within three phases, the data-analysis process was also conducted systematically using two types of analysis, preliminary or informal analysis and focused or formal analysis. The preliminary analysis was conducted within each phase while the focused analysis was being conducted during each phase of data collection (Figure 7). Next, I will explain the purposes, and the processes, of conducting both types of data analysis.

6.10.1. Preliminary analysis of interview data

Grounded theory recommends analysing data as soon as it is collected to inform the subsequent data-collection process. Within this study, as mentioned above, the data collection was conducted within three phases with a number of interviews conducted during each phase. For example, the first phase included 10 interviews, the second 14 and the third 5. Following the grounded theory guidelines, formal analysis of interview data was conducted after each phase of data collection. However, during each of these phases, after each interview, I prepared a summary of the interview conversation, in

memo form, as a preliminary analysis before conducting a formal analysis, which included the formal coding process (discussed in Section 6.10.2).

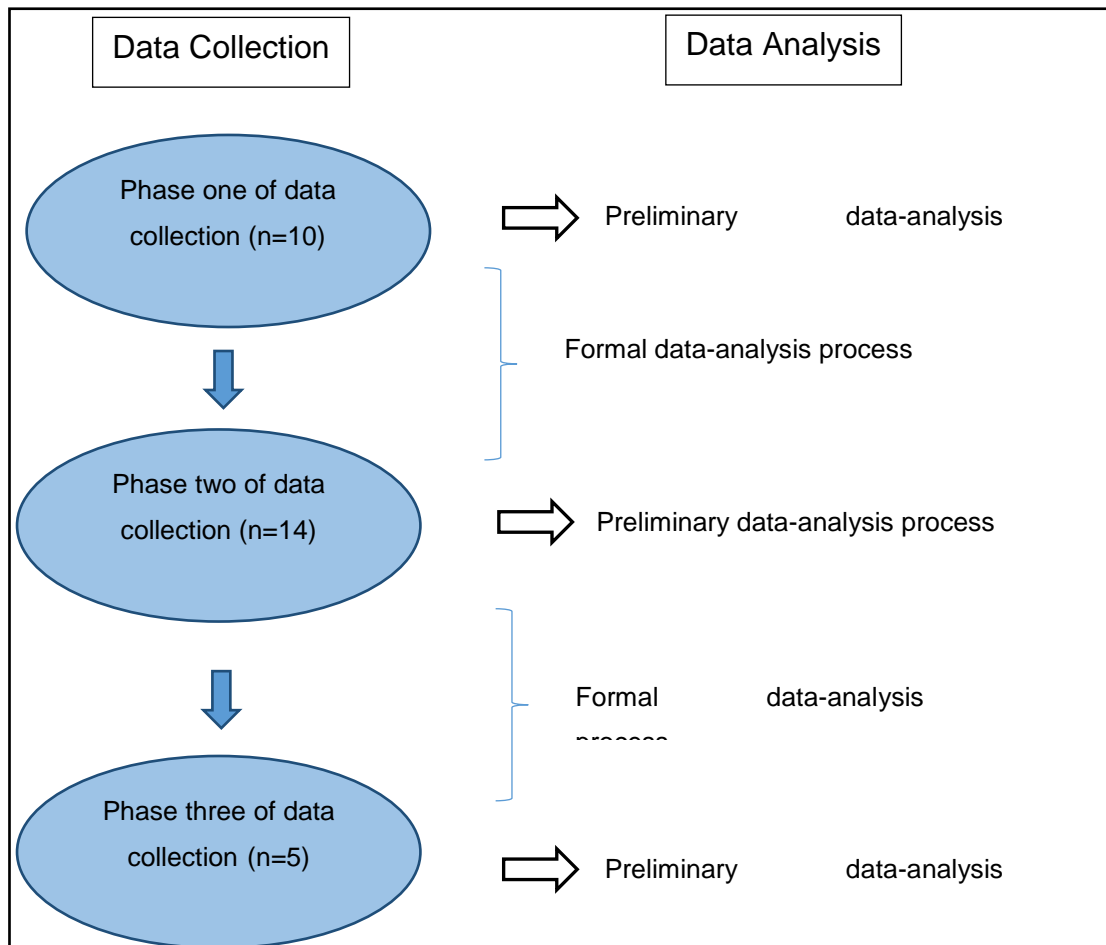


Figure 7: Phases of data-collection and analysis process

The memo included the main points of discussion within the interview, my reflection of what it meant and what I needed to explore further during the subsequent interview, thus becoming part of the process of informal analysis. For example, the memos I wrote after the first interview highlighted the main concepts or themes arising from the data, consisting of issues, needs and expectations as requirements for using DCM data. I wrote down the context from which these concepts emerged and what I wanted to explore further

(see appendix 9 as an example). This informal analysis facilitated me in conducting

the next interview with some initial concepts in mind for further exploration. The subsequent interview might then clarify, validate or bring new themes or concepts to light, which were absent in the previous interview. In this way, the first phase of data collection was completed.

Hansen and Kautz (2005) maintain that informal analysis is important in those cases when data collection is conducted using topics rather than strict and structured questions, as the researcher is open to develop hypotheses during and after each interview (Hansen and Kautz 2005). Using grounded theory, the development of these hypotheses is imperative to continuing the process of data collection, sampling and analysis until saturation is achieved. These hypotheses are developed or refuted as the researcher moves from one interview to another (Strauss and Corbin 1998). When the researcher feels that the emerged hypotheses are developing over the course of a few interviews, the interview guide is modified to add the hypothetical ideas to questions for their further exploration, verification and development, taking multiple views from subsequent interviews.

6.10.2. Formal analysis of interview data

After each data-collection phase, the interview data was transcribed (as mentioned above in Section 6.8) and formal analysis was begun. The formal analysis of the interview data included coding. Coding (naming text) is a core and fundamental element of data analysis used in most qualitative data-analysis approaches, particularly grounded theory. Within grounded theory, coding is a systematic process of highlighting issues, concerns and subjects, which are important to the study participants (Strauss and Corbin 1998). Charmaz (2006) maintains that coding presents a link between an empirical

reality and the researcher's view of that reality. She further explains that coding creates a bridge from the emic form of data interpretation (the participants' views) to the etic form (the participants' views influenced by the researcher's views).

In order to reduce data from emic to etic form, I conducted a systematic process of coding using open coding, axial coding and selective coding techniques (Strauss and Corbin 1998). A coding framework was devised for such purposes, as shown in Figure 8.

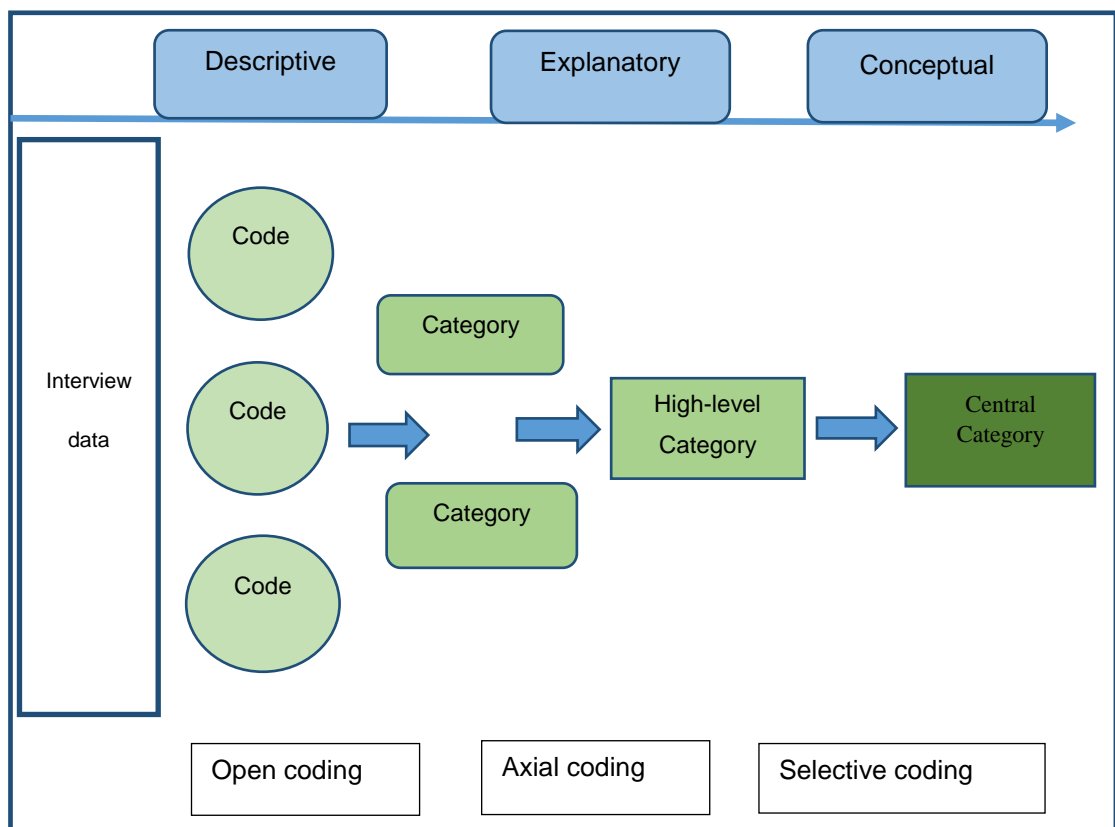


Figure 8: A coding framework.

Next, I present techniques to demonstrate how I reduced large amounts of interview data to a few categories that explained my interpretation of the requirements for the DCM data warehouse that emerged from the study participants' perspectives.

6.10.2.1. Open coding; exploring data

According to Wang (2014: 613), “open coding is a process of breaking down, examining, comparing, conceptualising and categorising data”. Open coding is the initial step of data analysis using grounded theory, which Strauss and Corbin (1998) refer to as examining data at a micro level (i.e., looking at small details). Miles and Huberman (1994) term this process as finding seed categories; these are focus points that guide further data collection and analysis. Charmaz (2006) refers to open coding as initial coding and recommends it at a very early stage to get detailed understanding of the study data.

The open coding began on the interview data collected during the first phase (n=10). For this purpose, I read and re-read interview transcripts as well as individual interview summary memo several times to understand the data in detail. Once I became familiar with the data, I started coding interview transcripts, which meant giving descriptive names or labels (codes) based on segmentations of the data. Segments comprised one or more lines of text that contributed to the discussion of a particular concept, as defined by myself through reading the transcript. In this way, lines could contribute to one or more segments.

As a novice grounded theory user, I found it challenging to code the first few interview transcripts. The main challenge for me was to decide what was important to code. While Lowe (1995), in this regard, suggests developing a topic guide based on the initial research question and coding data that is

relevant under the identified topics, Glaser (1978: 57) recommends the use of three questions to guide open coding:

- What is this data a study of?
- What categories does this incident indicate?
- What is actually happening in the data?

In this study, the initial research aim was to explore the requirements for the secondary use of DCM data from the potential users' perspectives. After several readings of the interview data, I chose to analyse data for three purposes driven by the objectives of the study. The first purpose of the analysis was therefore to set the scope of the warehouse, which means identifying the secondary uses for, and users of, the DCM data. Based on the identified potential uses and users, the second purpose was to identify the information requirements from the potential users. This analysis required focusing on the type of data required by the potential users. The third purpose was to focus on the issues and concerns related to the secondary use of DCM data. Further, I also focused on the new aspects emerging from the data, which could be useful for answering the research question and the study's overall aims and objectives.

As mentioned in Chapter 4, the requirements could emerge from the users' needs, expectations and concerns related to their existing experiences of using a system or their future expectations of a new system. These requirements can either be directly expressed by the users or can emerge from conversations with them, in which they express their concerns, issues, complaints and expectations relating to the system (Pace 2004). To achieve

the third purpose, therefore, the focus was also on identifying users' directly and indirectly expressed requirements within the interview data.

During open coding, another challenge was to decide how to code and deciding what kind of names or labels should be given to the text that captured the main essence of what the participants' needs, expectations and concerns were within the interview data. For this purpose, Charmaz (2006) suggests looking for action words or phrases (verbs) within the data. While the focus should be on identifying the verbs, Charmaz (2006) also suggests looking for the subject (who performed the action) and an object (who was receiving the action) as this helps in providing the context for the emerged action codes. I applied Charmaz's approach to the first few interviews to assess its applicability for my data. As the participants' needs and expectations involved actions they were already doing, or wanted to do in terms of using and managing DCM data, I gave names (codes) to the text that represented action words, for example, 'wanting to analyse', 'integrating data', 'looking at data' etc. However, the events, objects as nouns were also given names (codes), for example, 'mappers' (practitioners), 'mapping data', 'data quality', 'Excel system' etc. These codes give context to the action codes. For example, in Table 3, the main concept is 'experiencing difficulty with existing system', representing the action codes within the data. However, the subject and object codes provide the context for this action. For example, the practitioners (the subjects) were experiencing difficulty with the Excel system (the object) that they were using for integrating data (the action) for analysis purposes.

Table 3: An example of open coding process.

Interview Text	Code	Category
Part of our problem is trying to pull all the data together , this is the work what we are doing by ourselves... to look at data for reporting because the Excel spreadsheets are very difficult to work with . Especially if you are wanting to look at it for service... you have everything from Excel from one map but a patient might be there for a number of different Excel spreadsheets and trying to pull it together is very difficult .	<p>Mappers (DCM practitioners); Mapping data; Excel system.</p> <p>Primary Use_ Trying to integrate various mapping data.</p> <p>Primary Use_ Difficult working with existing system.</p> <p>Primary Use_ Difficult to pull data together.</p>	Primary Use_ Experiencing difficulty with existing system (data integration).

During open coding, a large number of codes emerged, which were descriptive rather than analytical in nature, and most of these were the participants' spoken words and sentences. The next step was to merge the codes into categories, based on conceptual similarities observed by myself as the researcher. The journey from codes to a category is signified as 'conceptualisation' which Strauss and Corbin (1998: 103) refer to as the first step in theory building. This takes the coding from the descriptive level (codes) to the level of explanation (categories). According to Strauss and Corbin (1998: 103), a concept is an "abstract representation of an event, object or action/interaction that a researcher identifies as being significant in the data". This means that the researcher is given the freedom to bring various codes together into categories that explain the data at an abstract

level, based on her/his own understanding of the data and the emerging concepts. These

categories represent not one participant's or one group's views (that is user requirements in this study), but a collective set of views of many participants, which is reduced to explain the conceptual details of a specific area under study (Strauss and Corbin 1998) (see appendix 12 for an example).

The conceptualisation and formation of categories were influenced by the 'constant comparison' method. As mentioned in Chapter 5, constant comparison is the most important grounded theory technique (Strauss and Corbin 1998) and the main indicator of the quality of study data and findings (Gasson 2003). According to this technique, various aspects of the data are compared to formulate differences and similarities within them. This process allows the identification of concepts within the data. Further, the emerged concepts are also compared with new and existing data. Based on this, similar concepts are housed in a category. The process of comparison between data and emerged concepts was conducted constantly until I decided that new data did not identify a new concept and that it did not add to the explanation of the existing categories.

Within this study, during open coding, the code given to each segment was compared to the code given to the other segment of texts from the same interview for similarities and differences. Here the questions arise: what do similarities mean in data? In what sense is the data similar? Is it similar in its meanings, the context through which it has emerged, or the sense it is making to the researcher? Or is it similar because the participants were using the same words while describing the issue/concern? Within this study, the codes were merged based on the similarity of the context within which

they emerged, where the context was defined as being the set of circumstances where many

participants share the same situation. For example, one of the codes – ‘comparing data’ – emerged from the interview data. When it was compared with other codes, such as ‘integrating data’ and ‘manipulating data’, it was noticed that all these codes shared the same context related to the issues with existing data-management systems. The above-mentioned codes were some of the data requirements that the participants mentioned and were associated with the lack of an effective data-management system. In the context of this study, the example of conceptualisation is shown above in Table 3 where the codes were merged into the category ‘Primary use_ experiencing difficulty with existing systems’.

In this way, the whole analysis was carried out by comparing various aspects of data and categories to identify enough details to explain the emerged categories. During the constant comparison process, I wrote several memos as reflective diary entries explaining my rationale of comparison and outcomes. Examples of one of these memos can be seen in appendix 10. This information further helped in remaining consistent while comparing categories based on the identified similarities and differences. During the coding process, memos were written at each stage, for example, during code generation, concept emergence, the identification of similarities within codes, the amalgamation of codes into categories and category generation and development. Within these memos, I recorded my reflections on the analysis process; for example, how I coded, why I gave the specific names to the pieces of text and how I illustrated the underlying concept, based on which I merged various codes into categories, recognising the categories that

needed further development. According to Strauss and Corbin (1998), memos allow the

researcher to write their reflections on the data-analysis process, which ensures the rigour and quality of the process (discussed in Section 6.11).

As the interview data indicated that related issues and concerns were emerging as categories in relation to both primary and secondary use of DCM data, a naming convention was developed for each type of category. For example, in order to distinguish them, each category was given a prefix to indicate either primary or secondary use, as in these examples:

- Primary use_ difficulty in analysing data;
- Primary use_ limitations of existing systems;
- Secondary use_ requiring additional data;
- Secondary use_ estimating the quality of DCM data;

Following the adoption of this naming convention, it was much easier to manage, interpret and analyse categories within a specific context of how DCM data was used and managed.

During analysis, theoretical sampling also guided me in identifying new sources of data to explain and then saturate the concepts and categories that emerged from the existing data. Here, 'new sources' does not only mean conducting more interviews to saturate existing categories, but as suggested by Strauss and Corbin (1998), the new sources could be existing data (collected during phase one or two) that could be re-analysed and re-coded based on new insights that have emerged. Every time a concept emerged, I re-analysed the previously coded interview data to find additional relevant examples. In order to integrate all relevant data as part of the concept, I re-coded them wherever necessary. For example, the concept 'Mapper's role

influencing the quality of DCM data' emerged after a few interviews, when some of the participants (researchers) showed concerns regarding the quality of DCM data for secondary use and mentioned the mappers as the most responsible entity in this regard. Whilst more data was collected to explore how and when mappers can influence the quality of DCM data, the previously coded interviews were also analysed again, specifically to answer these questions. In this way, the coding and re-coding of interview data continued until the whole body of interview data was collected and analysed and decisions were finalised to establish that the emerged categories explain how the study participants' requirements related to the use and management of DCM data.

The next section explains the process of axial coding which links various categories that emerged during open coding, the next step of conceptualisation.

6.10.2.2. Axial coding; linking categories

Axial coding is the second type of coding, where data reduction moves to another level of conceptualisation (Strauss and Corbin 1998). This is driven by the relationships between categories. Axial coding enables the researcher to identify relationships, based on which she/he can link categories at a conceptual level and give them a name that represents the main concept within the merged categories. In grounded theory, this is referred to as conceptual or high-level categorisation of data. During this coding process, the researcher collects more data that could strengthen, justify and explain the relationships between categories.

In order to facilitate the axial coding process, Scott (2008) suggests adopting an interrogative style which he calls a 'conditional relationship guide'. This refers to the process of exploring each category and discovering relationships by asking relational questions, for example, questions about what, when, where and with what consequence, in order to explore each category and its relationship with other categories. Strauss and Corbin (1998) place these questions in a 'paradigmatic model' for further exploration and development of each category. Using this model, they suggest exploring a category by: identifying the *phenomenon* within a category; understanding the *conditions* that escalate this phenomenon; identifying the context within which the *phenomenon* is embedded; recognising the *actions/strategies* that people use to manage or solve the *phenomenon*; and finally postulating the *consequences* that emerge as a result of the actions taken. According to this model, the categories are scrutinised and explored, with the intention of gaining further understanding, until the researcher thinks that the questions pertaining to these are answered, thereby assuming saturation.

Within this study, I used Strauss and Corbin's (1998) model as a framework to explore categories and their relationships. For example, a link could be seen between all the categories that referred to the concept of issues and limitations of existing data-management systems. In order to explore the concept, I collected additional interview data and also returned to existing data to understand contextual reasons as to why these issues and limitations were important concerns for the study participants. Further, I examined the situations and conditions where participants experienced these issues and limitations and explored their perception of how to deal with the identified

limitations and the strategies they perceived could provide them with an effective system to manage their DCM data.

The linking process is intensive and time-consuming and involves moving back and forth between categories and data for comparison purposes and determining the gaps within each category. Within this study, this process continued until all the categories and relevant data were analysed and a high-level category was established to represent the new relationships. For example, to represent the issues and limitations within existing data-management systems, the high-level category was named 'Primary DCM data-management systems: limitations and requirements' (Figure 9).

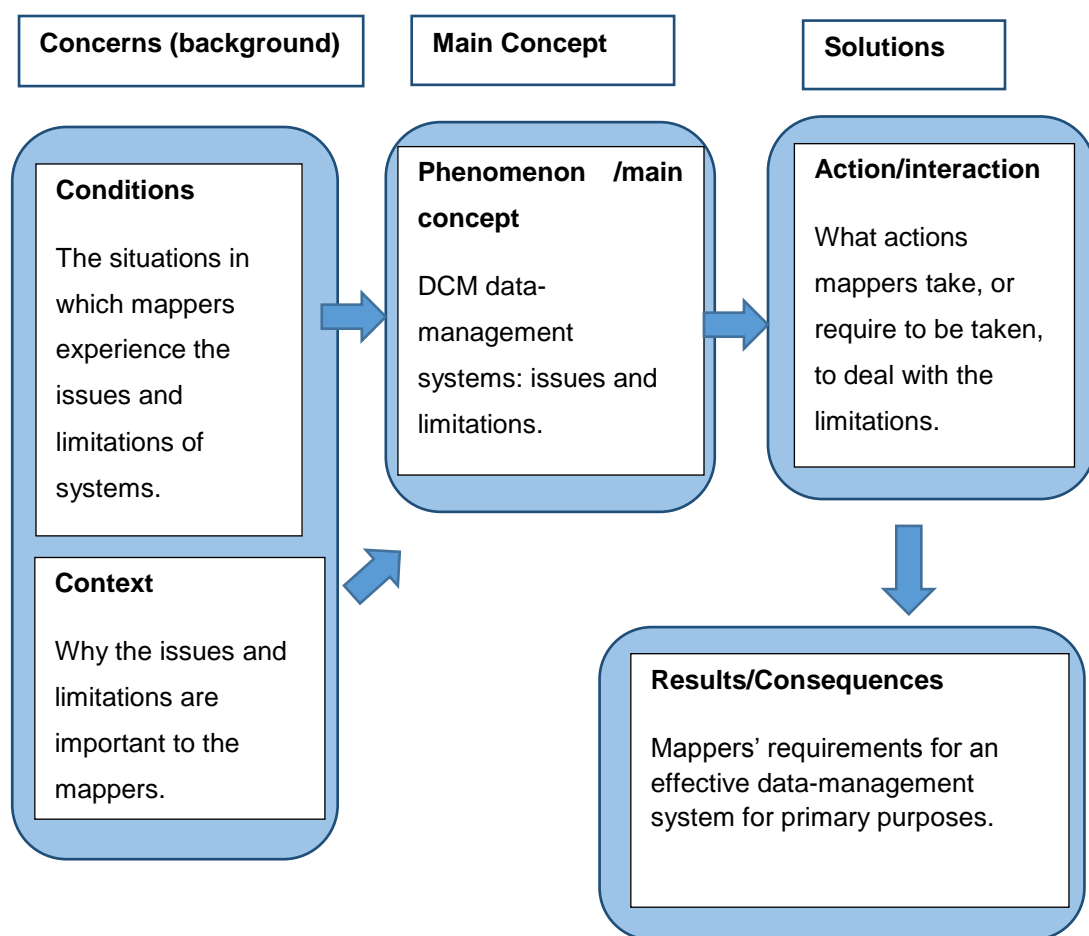


Figure 9: The components of the 'paradigmatic model' and a main concept

Using the same model for exploring categories and their relationships, all categories were related to each other and more data was collected to refine and validate the concept based on which the categories were merged into high-level categories. Some other examples of this high-level category formation are provided in appendices 13 and 14. Following this process, six high-level categories were established, as shown in Figure 10 below.

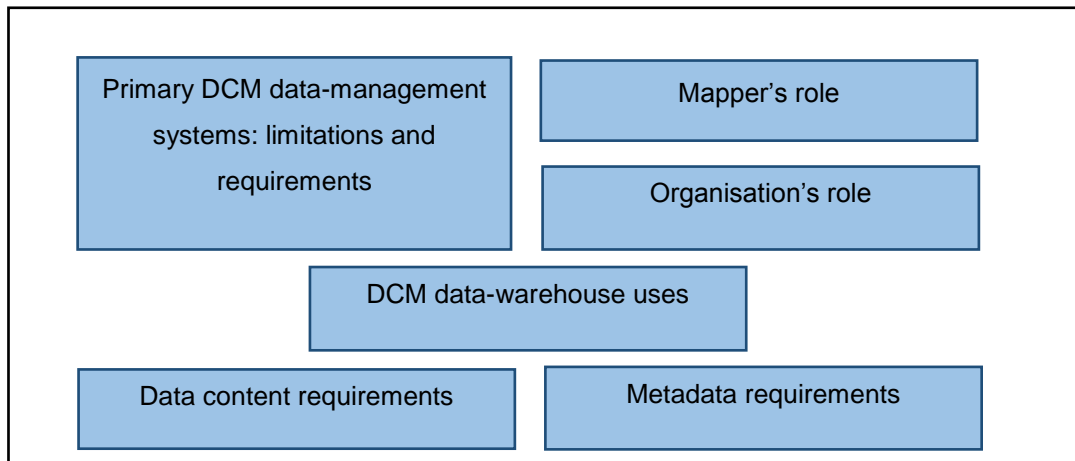


Figure 10: Six high-level categories

At the end of axial coding, there existed high-level categories that could be related at a theoretical level to explain the requirements for the secondary use of DCM data that had emerged from the study participants' perspectives. In order to develop these relations, therefore, the next step of coding commenced. This is called selective coding.

6.10.2.3. Selective Coding; finding a central category

According to Strauss and Corbin (1998), if categories show links that could explain a specific issue within the data, they should go through selective coding. Selective coding is the process of integrating various categories and the underlying concepts to explain a specific issue/phenomenon/event within the data; this is also called a substantive theory. Selective coding is

conducted at a theoretical level. This means that a careful analysis of each category is

conducted to identify the main theme, or to see if the categories are related to each other. The common idea that relates to each category is called the central, or core, category which, according to Strauss and Corbin (1998), provides interpreted abstractions and not the descriptive details of each case (e.g. raw interview data).

In this study, once high-level categories were identified using axial coding techniques, the categories were analysed for the main idea. Three high-level categories such as 'primary data sources: limitations and requirements', 'mapper's role' and 'organisation's role' were pointing towards three main factors that could potentially influence the quality and availability of DCM data for secondary uses. These three main-categories therefore were linked to form a main category depicting issues related to the availability and quality of DCM data for secondary uses (Figure 11). Similarly, two other high-level categories such as 'data content requirements' and 'metadata requirements' were together explaining the users' information requirements for a data warehouse. These two high-level categories were linked on this basis (Figure 12). The remaining category 'potential uses of DCM data' had indirect link with users' information requirements as it directed the process of research into exploring further users' information requirements. However, it does not form part of the main category 'users' information requirements' and therefore will be presented separately.

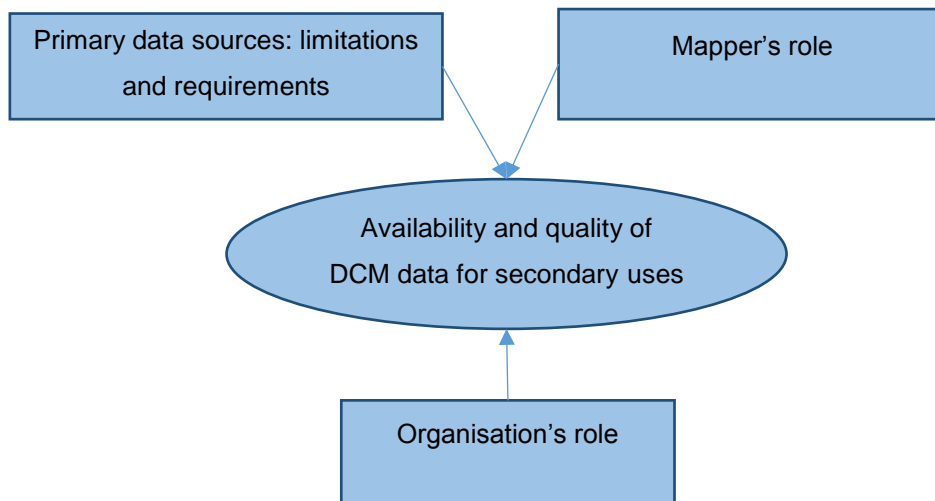


Figure 11: Central category: availability and quality of DCM data for secondary uses

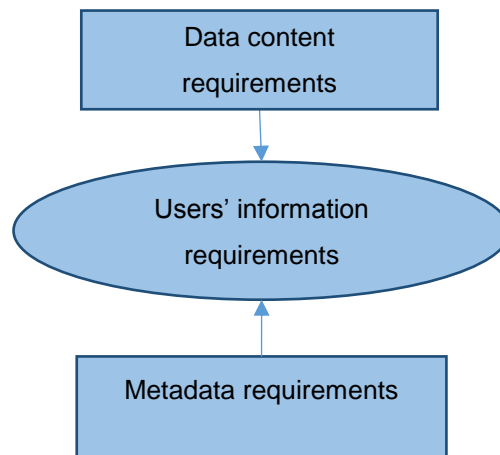


Figure 12: Central category: users' information requirements

The three main categories above present the study findings in chapters 7, 8 and 9. According to Gasson (2003), the grounded theory process and findings can be challenged over their reliability, as these are based on inductive inquiry and conclusions that reflect the researcher's subjective interpretations of the area under study. Like any other qualitative study, the rigour of the data collection and analysis process and the quality of the findings therefore require careful evaluation and justification, which, as

asserted by Gasson (2003), is based on the researcher's perspectives. As mentioned in detail in Chapter 5,

I adopted an interpretive perspective for this study. Therefore, the rigour and quality of the study process should be evaluated within the context of this perspective, the details of which are presented in the following section.

6.11. Evaluating qualitative studies

Several scholars (Guba and Lincoln 1981; Kirk and Miller 1986; Strauss and Corbin 1998; Huberman and Miles 2002; Creswell 2009) agree that qualitative research cannot be evaluated using quantitative procedures of measuring, reliability and validity, but instead needs different methods or techniques. Guba and Lincoln (1989) stress the terms 'trustworthiness' and 'authenticity', rather than 'reliability' and 'validity' in qualitative studies for evaluating rigour and quality. A qualitative approach defines its own criteria to evaluate the rigour and quality of the research as well as the outcomes (Gasson 2003). Therefore, qualitative researchers (Miles and Huberman 1994; Denzin and Lincoln 1998; Gasson 2003) argue using terms that are different to those of quantitative research in assessing validity. These terms are:

- Credibility rather than internal validity;
- Dependability rather than reliability;
- Transferability rather than generalisation or external validity;
- Confirmability rather than objectivity.

6.11.1. Credibility

The credibility of qualitative study findings is ensured through capturing the reality of the participants or phenomenon under scrutiny as accurately as possible (Denzin and Lincoln 1998), thus ensuring that the findings of a particular study are believable. This requires describing the process and

methods of working with the data from its raw form (e.g., interview data) to
the

findings. Charmaz (2006) states the study findings are interpreted statements which go through processing stages from raw data (study participants' interviews) to managed data (findings) that could explain the social phenomenon under investigation or support answering the research question. Miles and Huberman (1994) state that this data-processing stage is part of a process of generating meaning from the raw data. To ensure the credibility of the findings, it is therefore suggested that the data processing be described in detail (Gasson 2003).

This study set out to explore potential secondary uses of DCM data from the users' perspective to identify their needs and expectations as requirements for a future data warehouse. The credibility of the data-collection and analysis process was justified through presenting it in a systematic fashion and demonstrated by providing examples from the data, thus allowing the reader to make judgments about the credibility of the process (Strauss and Corbin 1998).

During data collection, the credibility of the collected data was obtained by ensuring that the question biases were reduced as much as possible. This means that questions were asked that would make sense to a specific group of mappers. Biased questions asked during the interview can influence the interviewee's answers. These biases can be introduced because of the way the researcher asked the questions and the type of questions (leading, unanswerable, or misunderstood by the interviewee) (Holstein et al. 1995; Hoets 2009). It is important to recognise and reduce these biases in order to get information that is of good quality and relevant to the study. I identified these biases during pilot interviews and changed my interview style and

questions. Because of this change, the modified questions were general, open ended, simple, clear, and were usually asked in response to the interviewee's answers. Further, the impact of biased answers was minimised by asking the participants for details and clarifications and by repeating what they had said in order to confirm their point of view.

Further, the credibility of data analysis was ensured by demonstrating systematic analysis of the interview data (such as preliminary analysis and formal analysis presented in Section 6.10) in detail, which enabled me to read and re-read the interview text many times, thus ensuring that nothing was left unattended. Furthermore, through the process of '*constant comparison*' and '*asking questions*', the underlying meanings and concepts within the data were interrogated constantly, which ensured that the explanations of the codes and categories were established within a relevant context. Gasson (2003) asserts that constant comparison of emerged categories with new data increases the credibility of the data. This credibility was confirmed by presenting different views to give in-depth explanation to the emerged categories (as will be seen in chapters 7, 8 and 9).

Miles and Huberman (1994: 65) suggest a process of 'check coding' which refers to the process of validating one's coding process with others, thus ensuring the process's credibility. For this purpose, my supervisors had access to the data obtained and the coding process. This analytic strategy for 'check coding' proves the validity of the coding process. At each stage, the emerged categories and their eventual development were also discussed with my supervisors, the main purpose being to validate my own assumption

and the emerging hypotheses within the data as categories. Further, I also discussed

my interpretations of the interview data with my supervisors to get their opinions and feedback.

Many authors advocate that the credibility of the findings can be achieved by being transparent while explaining the research process to the reader in terms of providing details of study design, implementation, data collection and analysis and reflection appraisal by the researcher (Guba and Lincoln 1981; Shenton 2004; Creswell 2009; Yin 2009). For this purpose, the process of data collection and analysis is explained within this chapter in detail. This explanation provides an understanding about the study development over the period (Shenton 2004), including the decisions taken by the researcher. Further, it demonstrates the dependability/reliability of the study, which is explained next.

6.11.2. Dependability

Dependability refers to the reliability of the process through which findings are achieved. According to Guba and Lincoln (1981: 316), “a demonstration of the former (credibility) is sufficient to establish the latter (dependability)”. Gasson (2003) asserts that the dependability of any study is guided by the researcher’s philosophical position. The interpretivist approach allows the researcher to present socially constructed realities as subjective interpretations of multiple realities, which can be reported differently by different researchers who are reporting the same data and using the same methods (Gasson 2003). The main reason is the subjectivity that each individual holds, which influences her/his interpretation of the data. Whilst the researcher’s subjectivity is widely acknowledged in qualitative studies,

particularly in grounded theory, it is advised that she/he shows reflexivity (Guba and Lincoln 1981; Gasson 2003).

This requires the researcher to reflect on the position she/he takes while performing various actions during data collection and analysis. These actions include: choosing specific participants for data collection; choosing specific categories to focus on; directing the data-collection and analysis process; and deciding when to stop the data-collection process (Gasson 2003). In this chapter, I have shown reflexivity at each stage of data collection and analysis in order to remain transparent about the assumptions I have made and the decisions taken for identifying concepts within the data and for linking various categories. This was to acknowledge my subjectivity and to provide the reader with a transparent account of the study.

6.11.3. Transferability

Usually, findings are considered valid if they can be applied to another population or area/field. This is called generalisation. Qualitative research aims to present the particular description and themes developed in the context of the specific issue, place or population. Caracelli and Greene (1997) describe this aim as providing 'particularity' not 'generalisability'. Shenton's (2004) claims, also supported by Guba and Lincoln (1981) and Creswell (2009), verify that the findings as a result of a qualitative study are impossible to generalise as they are specific to a particular phenomenon, group of people (community) and situation. However, findings from one study can be transferable to another situation given some evidential factors, for example, enough details to support the reader in making judgments about transfer (Lincoln and Guba 1985). For this purpose, Yin (2003) argues that qualitative research can be generalised or replicated only if excellent procedural steps

are documented, thus making it transferable to another somewhat similar situation.

The aim of grounded theory is to produce substantive theories or in-depth explanation, applicable to the particular area of empirical enquiry from which they emerged (Fernandez et al. 2002). Glaser and Strauss (1967) assert that the theory relevance is therefore only within the environment concerned. Using grounded theory, however, generalisability can also be achieved, if the aim is to move from a substantive theory to a formal theory. Formal theory describes the area under study at an abstract level (Gasson 2003). It deals with the conceptual area of inquiry such as – for example – stigma, formal organisation and socialisation (Goulding 2002). This type of theory is not specific to a particular phenomenon, issue or group of people and thus is usually generalisable to other issues, phenomena or groups of people that share similar characteristics (Strauss and Corbin 1998).

The aim of this study was to identify the views, needs, expectations and concerns of the potential users (mappers in this study) and then interpret these into requirements relevant to the future data warehouse. This study is therefore concerned with substantive theory rather than formal theory, where the aim is not to generalise the findings, but to explain user requirements for their potential secondary uses of DCM data for a data warehouse.

6.11.4. Confirmability

Confirmability ensures that the findings are truly coming from the participants' experiences and behaviours and are not contaminated by the researcher's

preconceptions or presence, usually known as biases. Confirmability can be achieved by reducing these biases as much as possible.

Qualitative research, underpinned by interpretivism, encompasses some biases which, according to Patton (1987), are inevitable. However, in order to maintain the quality and trustworthiness of the study, the researcher needs to minimise these as much as possible or at least be aware of such biases and reduce the chances of their having an impact on the quality of the research process. During the research process, there were a few biases, which I recognised and tried to minimise as much as possible. These were: researcher bias (related to the researcher's personal characteristics and professional background); biased questions (which can influence the interviewee's answers); and biased answers (false statements which can be made by the interviewee either in ignorance or on purpose) (Hoets 2009).

In this study, the researcher's bias might emerge due to my technical background, my DCM training as a mapper and my link with School of Dementia Studies. Hoets (2009) maintains that such a bias can restrict the participants' expression of knowledge, as they assume that the researcher is aware of the particular area and will know all the details, which they then do not have to explain. Further, the element of intimidation might also influence the expression of knowledge, as the participants could consider the researcher more expert and knowledgeable in a particular field (Hoets 2009), such as a data warehouse in this study. Such bias can influence the quality of collected data. Therefore, wherever it was possible during the interview process, I made sure not to mention or emphasise my technical background, my link with School of Dementia Studies and my DCM training as a mapper.

Further, to reduce the biased questions, I ensured that interview questions were exploratory rather than directive. For example, I asked the participants how they perceived the use of DCM data within the warehouse. Rather than are you concerned about the quality or security of DCM data within the data warehouse? The exploratory questions helped me to identify the concerns, issues and perceptions, which some mappers felt would concern them while using the data within the warehouse. The bias of providing false answer was minimised by ensuring that study participants were asked to clarify their answers.

6.12. Summary of the chapter

This chapter began by describing sampling, where details were provided to show how the study participants were recognised, identified and finally recruited to take part in the study. Further, it justified the use of a specific number of participants within the study, by arguing that, if the purpose of a study such as qualitative and particularly a grounded theory study is not to generalise but to present a specific area of study, the number of participants is less of an issue.

The chapter then highlighted the ethical approval process for this study. It focused on justifying the data collection method and then showed how the data collection process was begun and completed within three phases. The chapter also presented how and why an interview guide was created and how it was developed alongside the process of data collection. It then moved to describing the process of data analysis in detail. It focused on explaining the use of grounded theory guidelines for data analysis and on highlighting the areas where a number of decisions were taken for the application of the

guidelines due to the nature of the data. Furthermore, this chapter underlined the significance and use of the literature in making sense of the emerged concepts within the data. It also highlighted how the relevant literature helped in understanding and interpreting the users' views, concerns and expectations as data warehouse requirements.

The chapter ended by arguing that qualitative studies required different criteria for evaluating the quality of the study process and findings. It showed how the present study dealt with various data quality issues. This chapter will now lead on to the next chapters where findings will be presented and discussed.

7. Potential Secondary Uses of DCM Data; Users' Views

7.1. Introduction

The previous chapter explained the data collection and analysis methods, underpinned by the guidelines of grounded theory. In keeping with these guidelines, the interview data was coded and three main categories emerged, which are presented in chapters 7, 8 and 9. Chapter 7 (this chapter) explores mappers' views regarding their potential secondary use of DCM data. Chapter 8 presents and analyses the potential users' information requirements for a data warehouse and finally Chapter 9 presents the identified factors that can potentially influence the availability and quality of DCM data for secondary uses. The interviewees' verbatim quotations from the interview transcripts are used to illustrate the analysis and show that findings are grounded within the data. Further, wherever applicable, the information from the relevant literature is drawn in to explain, link, clarify and discuss the findings as potential requirements and their implications for a data warehouse. Each chapter also discusses the main findings with the purpose of highlighting the original contributions to knowledge and arguing the need for further research.

This chapter presents the study findings that explore mappers' views regarding their secondary use of DCM data. This will meet the first objective of the study, which was devised (in Chapter 3) based on the argument that, in order to set the scope of a data warehouse, it is imperative to identify potential uses of the data that can subsequently determine the purpose of a

future data warehouse. This purpose will further direct the design and development of a data warehouse for a specific user group.

The interview data highlighted that the potential secondary uses of DCM data were varied among study participants. How they perceived the uses were mainly influenced by their existing primary uses of data. Mappers, who were using DCM for practice development purpose (DCM practitioners) within their own organisations for improving dementia care at an individual and organisation level, envisaged viewing other organisations' dementia care related data as a way of sharing good practice through benchmarking DCM data. Mappers, who were involved in providing DCM training (DCM trainers), on the other hand, were interested in having access to historic and integrated DCM data to see trends and patterns, to identify and share good care practices and to use a variety of examples from the data for their DCM teaching and training. However, those who were currently using or had an interest in using DCM for research purposes (DCM researchers) saw benefits of using DCM data to support their potential future research within dementia care (secondary research). Based on these differences regarding the secondary use of DCM data, the study participants' views were divided into three sub-categories (Figure 13) and are presented next.

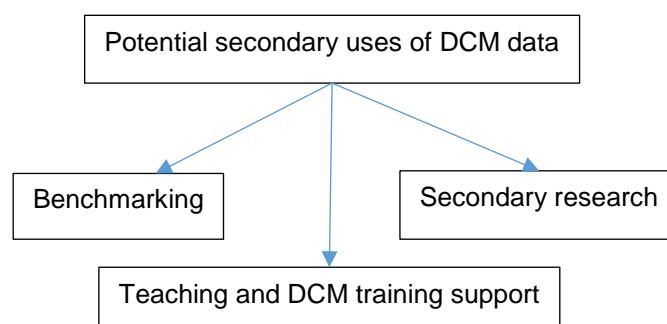


Figure 13: Category 'potential secondary uses of DCM data' and its sub-categories

7.2. Benchmarking

This sub-category explores and analyses the study participants' perceptions, who were using DCM for practice development purposes. Their perceived potential uses of DCM data can be explained within the context of benchmarking. As was mentioned in Section 2.3 of Chapter 2, the concept of benchmarking is associated with organisations' efforts to make continuous quality improvements (Ettorchi-Tardy et al. 2012). Study participants shared the same concept in terms of improving quality of care that they deliver within their organisation and interview data highlighted two ways of achieving this. The first was through using DCM data for internal benchmarking; this means comparing either best practice or current practice over time within their own organisation (Lovaglio 2012). In this context, one of the study participants mentioned the potential use of DCM data within their own organisation to see if the quality of care they deliver was consistent within all parts of the organisation.

"It would be lovely if they [mappers] mapped in surgery and we could compare, yes, because it is a totally different environment and way of working, but if we are working really good, then it would be nice to see if we could transfer it or use it across. And for different areas, I think, like the outpatients because I am not knowledgeable about outpatients, but obviously they are getting people for appointments and things and that has an impact as an approach, you know, everybody is miserable. You know what I mean. It would be interesting seeing what they do and what could we do." (Practitioner, 12)

The above quote suggests the study participant's view of achieving good standards of care within one part of the setting and then replicating these

standards across the setting in other parts. This suggests the attraction of transferability of DCM across different types of hospital wards. According to Kay (2007), if one part of an organisation does well, the knowledge can be replicated in other parts as well. While the above quote indicates the respondent's desire to share their good practices of dementia care within all parts of their organisation through using and comparing DCM data, the comparability of such data might be an issue, as the care environments of an organisation, such as various parts of a hospital (e.g. inpatient and outpatient) are different and thus may not be easily comparable.

In the similar context of internal benchmarking, two study participants from another organisation also mentioned their existing work in measuring the individual's quality of care over a period of time. They established benchmarks using only coded DCM data (BCC, ME, PE/PD). Individual patient-level DCM data was collected on a regular basis over three months and then processed, in an aggregated format, to show how changes occurred over time at both individual-patient and service levels. Using DCM data for internal benchmarking, they showed variability in care quality at an individual and organisational level via three data types or indicators (WIB score, BCC and PD/PE).

"What we have done in the past that we piloted to create a set of Excel spreadsheets that have everybody's data in there and looked at three data points from the period of three months." (Practitioner, 2)

This finding corroborates the idea of Brooker (2005) who suggests that the WIB score could be used as an indicator to assess changes in care over time

that is internal benchmarking. The study participants' indication of the use of DCM data for internal benchmarking supports the argument made in Chapter

2. This stated that the literature provides evidence to demonstrate the use of DCM data for internal benchmarking where the quality of care is compared over time (Brooker et al. 1998). However, the lack of evidence in the literature regarding the use of DCM data for external benchmarking was highlighted. The interview data indicated mappers' perceptions of using DCM data for external benchmarking.

Another way of improving care within their own organisation that some study participants mentioned was the potential use of DCM data for external benchmarking. In this context, the study participants expressed a need to measure their own organisation's performance in providing quality of care against that of other similar organisations.

"I am very interested in the proposed data warehouse to be able to get benchmark data in order to measure against other services, although our service is quite unique, but knowing about other services within the country and further afield, I think, would be really interesting."
(Practitioner 2)

Within the context of external benchmarking, two main views were noticed from the interview data, the comparative and competitive views of benchmarking. The above quote indicates a comparative view, which according to Ettorchi-Tardy (2012), reflects organisations' willingness to learn from others by comparing similar processes of care. It assumes that a data warehouse will provide a set benchmark against which they can measure their performance in providing good care to people with dementia. Further, it also assumes that the warehouse will provide enough information for the organisations not merely to compare themselves against an average number

but also to show the process of how best practice was achieved (Ettorchi-Tardy 2012).

Another participant, however, expressed the use of DCM data for external benchmarking with a competitive view.

"It could help us a lot from a clinical point of view trying to see... you know, there is also a need when you are on a clinical approach to have data for benchmarking... having an idea of when an organisation is using the DCM, one can say, yes, we are a person-centred care organisation. We have reached that level of care, we are providing person-centred care and we can prove this through DCM."
(Practitioner Trainer, 10)

The study participant's view indicated the mappers' perception of DCM data as providing a 'magic number', which could be used to see whether or not an organisation had reached the point of delivering good care. As mentioned in Chapter 2, this view reflects a competitive benchmarking approach where the aim is to meet a target number (Ellis 2006). While this view can also stimulate an organisation's quality-improvement activity, the research shows that its competitive side can be an unhealthy way of benchmarking (Kay 2007; Ettorchi-Tardy et al. 2012). The view presented in the above quote came from a practitioner who was also a trainer and who mentioned that, during DCM training, they were usually asked by participants attending training, how to assess whether or not units or organisations had reached the level of delivering person-centred care. This might be the reason why study participant 10 perceived the secondary use of DCM data within a data warehouse to provide a benchmark for good quality of care.

While using DCM data for either competitive or comparative benchmarking, the practitioners' intentions regarding the secondary use of data were to see other organisations' data so as to identify good practices that had worked for others in improving the quality of dementia care. This finding strengthens what was argued in Chapter 2, that DCM use has always been led by individuals and organisations' willingness to improve the care of people with dementia (Capstick, 2003). A future data warehouse can provide a platform where organisations can share their DCM data for this purpose. By integrating such data and calculating mean averages (such as group WIB scores), a data warehouse can be used to identify potential benchmarks for best practice. However, this requires identifying and storing as much data as possible to support both comparative and competitive views of external benchmarking. It would require quantitative data that can provide an average number indicating high levels of care and qualitative data that can provide additional information to provide in-depth explanation of which organisations have achieved high levels of care and how they have done it (Ettorchi-Tardy 2012). The literature has also highlighted the significance of using a broad range of dataset for understanding the full picture of care while benchmarking healthcare data (Kay 2007).

While DCM provides both quantitative and qualitative data, the data only reflect care experiences of people with dementia and no change processes that have led to improving these experiences. This information is important for the organisations to understand, analyse and then learn from the best processes in the sector (Ellis 2006). In consistent with the argument made in Chapter 2, the additional information explaining the change processes

therefore will be required to be part of DCM data that will be stored within the warehouse to meet organisations' comparative or collaborative benchmarking. This will potentially have implications for organisations in terms of collecting and storing additional information to explain the processes about if and how care quality was improved within their care settings. This is an important issue for future research to establish what and how such information could be made available within a future data warehouse to meet the benchmarking use.

Further, as stated above, the interview data highlighted that the practitioners perceived the use of DCM data for comparison purposes. This includes comparing good practices across various parts of the same setting or across different care settings, where effective comparability is the key. As mentioned in Chapter 2, data comparability across healthcare organisations is reported as a major challenge in benchmarking (Nolte 2010). In this context, Nolte (2010) asserts that a number of confounding variables – including how organisations provide care differently, have different ways of collecting data and also have different types of data – can influence comparisons across organisations which may seem to be similar settings.

In the context of DCM, Chapter 2 argued that, for effective comparability of data, additional information regarding participants' dependency levels and care-setting characteristics should be collected to permit the assessment of comparability, as participant and care-setting characteristics may have an impact on the wellbeing of people with dementia, thus impacting on the WIB score. This also re-emphasises Brooker's (2005) assertion on considering confounding variables alongside DCM data for benchmarking. While the

interview data indicate the study participants' requirements for benchmarking,

the additional data, including participant and care-setting characteristics, need to be collected alongside DCM data to support benchmarking. This study therefore suggests further research into identifying solutions to the question of collecting additional data alongside DCM data within the warehouse to meet practitioners' need for benchmarking.

Further, the data warehouse providers have responsibility for ensuring that the collected DCM data is comparable and suitable for benchmarking. As was argued in Chapter 2, DCM data has the potential to meet these requirements for benchmarking. However, more longitudinal studies are required to assess the suitability and comparability of data for benchmarking before considering the collection of data within a warehouse for this purpose. Chapter 2 also argued that quality and availability of DCM data were important requirements for benchmarking and that currently there was a lack of knowledge in this area. Based on the study data, Chapter 9 explores the factors that can potentially influence the quality and availability of DCM data for secondary uses.

In summary, the DCM practitioners expressed two ways of using DCM data for benchmarking. The first was to use data for internal benchmarking, where the intention was to compare and measure the same processes across various parts of similar organisations or over a period of time. The second one was to compare and measure themselves against other organisations in terms of learning and sharing good practices, referring to external benchmarking. While both types of benchmarking approaches reflect organisations' perception to use DCM data in order to improve the quality of

care provision, there is a need to collect data that is comparable across processes, time and organisations.

Further, the interview data showed that the DCM practitioners, as potential data warehouse users perceived benchmarking for both competitive and comparative purposes. In competitive benchmarking, users may interpret benchmarks only as a quality target number against which to show their performance, thereby not actually learning anything about how to make care improvements. However, the use of a data warehouse for comparative benchmarking will require organisations to know what good practices are and how they can learn from them.

This requires collecting additional information within the warehouse related to the DCM data. While the user requirements are considered valid in terms of their own perceptions of the secondary use of DCM data, the study data suggests the need to collect additional data for effective comparisons and information that can show organisations' change processes following the use of DCM. Further, it is also suggested that the compatibility (in terms of availability, quality and comparability) of DCM data for benchmarking, and its implications for a data warehouse, need further research.

7.3. DCM teaching, training and support

This sub-category explores and analyses the study participants' perceptions of using integrated DCM data for teaching and training and support, as a potential secondary use of data within a future data warehouse. This emerged from those study participants' interview data who, alongside their use of DCM for either practice development or research, were also involved in DCM training. Some of them were also involved in providing consultancy for using DCM in health and social care organisations. During consultancy,

alongside practitioners, they also used DCM data for primary purposes such as, for

individual care planning, organisational change and care improvement planning. For example, they used DCM as a process to show care aspects that needed improving through feedback, and supported staff with care planning at individual and organisational levels. However, the DCM trainers' main role was to provide training to individuals and organisations and then support them as they used DCM.

The DCM trainers' interview data revealed that they saw value in using DCM data for secondary purposes to see trends and patterns within a large amount of historic data stored within a warehouse, for example, to improve training provision and to track the trained mappers' development over time. This refers to trainers' requirements for a data warehouse whereby they can have access to data about the mappers they train as well as the DCM data they generate.

"We give them [mappers] the DCM tool and we say bye-bye to them and we never see them again unless they come to our advanced courses or anything else or you bump into them in practice. The data warehouse can give an opportunity to track what these mappers are doing, or are they doing mapping and if they have done enough mapping. I think there is a great advantage here." (Trainer, 18)

"I think it would be interesting to know the extent to which people [mappers] go on to actually use the tool after the training or particular reasons why they do and do they need support in using them? If they are not using it, then why they haven't managed to use it. I think that would be interesting to know." (Trainer Practitioner, 24)

The above quotes reflect other interviewees' views regarding their requirement of tracking all trained mappers use of DCM data in order to assess their support needs. There is a lack of any specific research around

DCM training, including how trainers can support practitioners and other mappers. There is

some evidence in the study data, as will be presented in Chapter 9, that the DCM practitioners do require support in mapping. However, no further knowledge exists that could explain what exactly mappers need from the trainers. Some evidence comes from Douglass and colleagues' (2010) survey study, where the DCM trainers were involved to devise a survey questionnaire for data collection. One of the main parts of this questionnaire consisted of asking the mappers about the challenges they faced using DCM. The underlying aim of this was to understand mappers' issues and support needs (Douglass et al. 2010). This is the only study that shows that the DCM trainers indicated their requirement to learn about mappers need in supporting the use of DCM.

In the present study, the interview data highlighted that the DCM trainers' requirement points towards having access to the mappers' and their mapping details. For example, they would require to know when mappers received their training, what training levels they had, how many mappings they were conducting and how often. The requirement for accessing mapper and their mapping details implies that a future data warehouse should store mappers' identifiable data such as, their name, contact details, training dates, mapping status etc. Technically, a data warehouse can store such detailed and identifiable data for multidimensional analysis. However, as was argued in Chapter 3, there needs to be a process in place to ensure that ethical and legal aspects are considered appropriately in terms of mappers' consent for access to their identifiable data and a secure storage and access within the warehouse.

Further, the data about mappers and their mapping details should be available to be collected within the warehouse to meet the DCM trainers' requirements. However, only a limited amount of data is collected about mappers. The University of Bradford keeps details of each trained individual mapper, including their names, organisation names and email addresses. However, this data is not linked with their mapping details and therefore it is not possible to assess the mappers' regularity of mapping or any further support in using DCM. The newly built archive DCM database provides the opportunity to store mappers personal (name and contact) and their mapping details in an electronic and linked format. However, it is not known if any, or how many, mappers have registered to date to use this database and whether they are also using the system to input their mapping data on a regular basis. Thus, the availability of data is an issue yet to be resolved.

In addition to having access to mappers and their mapping details, most of the interviewees also agreed that, given the opportunity, the integrated data could also enable them to learn from others' use of DCM whereby they could identify examples of good care practices that they can use during their DCM training and teaching courses.

"I suppose the [DCM] data can enrich the examples given during a DCM course. I think one of the things about the DCM course up until now, as one of the very valuable things that has been incorporated by each trainer, in my experience, is stories from their own experience to illustrate the points that are being made. The fundamental course is sound but the methods of teaching always change. One of the things is that, during the course, you can enrich the evidence with examples, which is possible by having access to the various types of data from this data warehouse." (Trainer, 26)

"I think one of the things that across care settings says that we are not sharing good practice. There is a lot we can do in sharing to trigger ideas and factors." (Trainer, 18)

The above quotes refer to the study participants' requirement to have access to integrated DCM data in order to extract examples of good care practice for dementia care. These examples can potentially help the DCM trainers to explain various mapping scenarios of coding for teaching purposes. The study participants' perceptions of accessing DCM data for teaching purposes do not necessarily require any identifiable data, but an aggregated or summarised dataset that can be used for teaching purposes.

In summary, the DCM trainers showed interest in using DCM data within the warehouse to identify further training needs and support for the mappers, thus referring to the requirement of accessing mappers' identifiable data. Further, they also expressed the requirement to learn about good care practices to use within their training and teaching, thus referring to the access to data for academic or research purposes that can be a requirement of anonymised data. The study findings have indicated that DCM trainers may be the potential users of a future data warehouse if it enables them to have access to data that could help them to improve DCM training, support needs and track-trained mappers' development over time.

7.4. Secondary research

This sub-category explores and analyses the study participants' perceptions of integrated DCM data for (secondary) research purposes. This presents those mappers' responses (DCM researchers) who were involved in using the

data for primary research purposes as well as those who were interested in using data for their potential research studies.

The researchers' main drivers for using DCM data from a data warehouse were as follows: having access to a resource with pre-collected DCM data; having access to DCM data for exploration and having primary validations of the data; and having access to unpublished DCM data. The details are explained next.

7.4.1. DCM data resource with pre-collected data

Like any other observational process, DCM demands a significant amount of time and financial resources for collecting data within research studies. DCM has been criticised for this reason (Thornton et al. 2004) Pre-collected DCM data within a future data warehouse can provide a rich resource, which the study participants welcomed. This finding strengthens the idea proposed, in a previous study (Khalid 2010) which I undertook, for a warehouse as a resource for integrated and historic DCM data.

"It's very expensive to collect any observational data, and if such data is available for secondary analysis, it might be that, later, you have to collect more data; still, it is very useful." (Researcher, 27)

"I would love to use that data [secondary DCM data]. Because, now, so much time is taken doing mapping ourselves, which is enormous in time investment. It would be a great opportunity if you had a sort of a database where everyone put their data in and you can also use it for research purposes." (Researcher, 13)

7.4.2. Exploring DCM data

Integrated DCM data can be used to answer new research questions, to conduct a pilot or an exploratory stage of a project, or to provide a researcher with a wide sample base for testing or validating her/his interpretations (Hox and Boeijs 2005). The study participants showed interest in having access to the integrated and historic DCM data for further analysis. The main purpose was to have access to such a resource for exploring data, collected from various types of care settings, over a period of time, from different geographical locations and for a number of purposes. Such data may provide researchers with access to a variety of situations within which DCM was conducted and data was collected, in order to compare data based on the available characteristics. One of the interviewees, who was interested in exploring available DCM data for comparison purposes, mentioned an interest in reusing the data for secondary analysis.

"I would be very fascinated in being able to extract back the information [DCM data] to see what behaviours people [residents] are experiencing in particular... if there is another organisation that is looking at mapping in a continuing care setting and being able to extract that for comparing and contrasting. So that is linking different organisations and settings together." (Researcher Practitioner Trainer, 17)

The above quote refers to the study participant's two requirements for secondary analyses. The first was to analyse the common type of behaviours experienced by people with dementia, for example, identifying the most common BCC within the data. The second was to compare DCM data sets

collected from various types of care settings. Other study participants also expressed similar types of requirements where they could analyse and

compare the data across various dimensions, a multidimensional view of DCM data. For example, the study participants expressed a need to see DCM data collected from various geographical locations for comparing studies conducted within different countries.

"I think it would be good to see the data of other researchers from different countries. I have collaborated with the German group and the group in Holland, so hopefully we will compare our findings after we have done our separate studies. If the same opportunity is available by the data warehouse, it would be fantastic." (Researcher Trainer, 11)

"It would be really helpful if available data is from various countries." (Researcher, 4)

The above quotes indicate the participants' requirement of accessing DCM data collected from various countries. The above findings confirm what is asserted in the previous studies that I undertook (Khalid 2010; Khalid et al. 2010) regarding the use of a data warehouse to support the secondary use of DCM data, reinforcing inter-organisational and inter-country comparisons. However, such a requirement has significant implications for a future data warehouse in terms of data security and data quality, as each country might vary in the context within which the data is collected, secured and then disseminated for the warehouse (Elger et al. 2010).

The interview data also revealed the study participants' expression to see the available data first (within a data warehouse) in order to make decisions on its reuse or its suitability for specific secondary purposes.

"I think it would have depended on what was available on it [data warehouse]... Yes, if one had been set up, we would have used it [data warehouse]". (Researcher Practitioner, 9)

The response from the above study participant was also shared by others, where they were interested to see the type of data that was available within the warehouse before they could decide its use. The significance of such a requirement has led many others to use a variety of methods to inform users about the type of data that is stored and available to access from the warehouses. For example, the UK Biobank (2012) advertises on their public website showing all the data types (in a grouped format rather than in individual patient level), which are available within the resource and further provides regular updates to inform the users about any new addition of the data type. Further, they also use usual public channels such as publishing articles to highlight the available data within the resource (UK Biobank 2012).

It was also evident from the interviewees' responses that the use of DCM data for secondary purposes is dependent on the type of data available alongside this data within the system. The additional information might help them to ask a variety of questions that could assist in their secondary research. This requirement is further explored in detail in Chapter 8.

"I think it probably depends on other data collected alongside it [DCM data]. I think it might be interesting in understanding relationships ... like particular clinical symptoms that might affect wellbeing or activity patterns. I think it depends what other information is available on the study database." (Researcher, 20)

Further, the study participants were also keen to have access to a large amount of DCM data within the warehouse for comparison purposes. They expressed the need to compare data by various types of care settings with the aim to assess the difference in care provision and outcomes. While such knowledge may help the researchers to design protocols or to propose new

research questions for potential research studies, in terms of a data warehouse, this requires storing information about the type of care settings from where the data was originally collected, thus requiring the need of collecting additional data alongside DCM data within the warehouse.

"It would be quite interesting to see how care provided in care homes differs from hospital care because we know that wards in hospitals are understaffed. Staff do very long shifts and it would be very interesting to compare with the care-home staff." (Researcher Practitioner, 9)

"I would be very interested in using such data for comparing with the data we collect using our tools. Because the DCM data is very detailed data." (Researcher, 27)

The access to a DCM data resource might provide researchers with an opportunity to explore the data for a number of reasons. Hox and Boeije (2005) asserted that secondary data sources might generate new research questions that could be answered via existing datasets. The similar requirement also emerged from the study participants. In addition, in the event of a criteria match between the data available within the warehouse and the proposed research question, the study participants also expressed their interest in using such data to increase their sample size.

"It would be extremely helpful, because now I have only my little sample to study my research question. But it would be really interesting, as it has not been studied before and it's a limitation of my study that there is such a small sample. So I would be really interested in using other mappers' data as well to have a bigger sample to study my research question." (Researcher, 4)

Further, the study participants were also keen to comparing their own findings and data with other similar studies for validation purposes.

"I would love to know what other people had mapped on an acute hospital ward... It would be really, really interesting to know what other people were coming up with." (Researcher Practitioner, 8)

The study participants also found it useful to search for DCM data that reflected various research purposes such as interventions, specific research outcomes. In terms of storing data within the warehouse, this need implies to the requirement of storing DCM data according to various themes reflecting different research purposes. This requires exploring how DCM data is currently being collected and stored so for potential secondary use. Chapter 9 covers the related findings and their discussion.

"It would be really useful, to us as researchers, if we were thinking of using DCM as an outcome measure and to be able to look in the database and find out whether any other researcher has used DCM to look at that specific theme or that particular topic. So, if I wanted to look at other people's mapping on reminiscence groups, it would be really to ask the database whether there are other maps stored here that specifically looked at reminiscence groups or similar topics to those that we are researching." (Researcher Practitioner Trainer, 16)

The study participants were not only interested in exploring raw DCM data for secondary analysis but also mentioned examining the results of the data in order to see the primary data users' interpretation in terms of how they implemented the findings of the study. In terms of DCM data collected through research studies, this requirement indicates the accessibility of the study results and their interpretations stored within the warehouse. However, in a practice-development context, this requirement is potentially related to

the DCM process where feedback and planning is an important step following

each mapping. This implies that, alongside DCM data, it is also important to capture the additional information explaining the change processes to allow the users to interpret the practice-development data within the right context. This finding matches the practitioners' requirement (see Section 7.1.1) for accessing information alongside DCM data that explains change processes in care and thus strengthens the need to store such information within the warehouse that meets both the researchers' and the practitioners' information requirements for their specific purposes.

"It would be great to know what other people's experiences are from looking at activities and themes of mapping and looking at how that has been used. So the links in terms of opportunities for sharing ideas for practice development are enormous." (Researcher Practitioner Trainer, 17)

In summary, the interview data highlighted that those individuals who were using DCM as a tool to collect data within their research studies saw value in exploring data from the warehouse. Their need of accessing DCM data within the warehouse was also intertwined with the requirements of having access to the additional information alongside DCM data in order to analyse the data from various dimensions, including time, type of care settings, locations/area and mapping purpose. Further, they also saw the value of making comparisons between various available datasets, which reflected various research purposes and collected from various countries, to generate new insights. Some however also preferred to view the available data within the warehouse to stimulate their potential future needs of using DCM data for their potential research studies. The researchers' requirements for using historic

and integrated DCM data within the warehouse indicate them as potential data warehouse users.

7.4.3. Access to unpublished DCM data

The aim of research is to add to existing knowledge and, therefore, most research studies require researchers to publish their study data and findings. There is no such requirement for practitioners. Yet, many practitioners have published their experience of using DCM data in journals accessible to practitioners, for example, the Journal of Dementia Care. However, data on the use of DCM outside of a research context remains largely unpublished. As mentioned by one of the study participants, echoing others' views, a data warehouse, as resource for DCM data, would provide access to unpublished data collected for the purposes of practice development as well as research.

"Through this database, you might come across unpublished research data. If you do a literature search, you only see the published research but through this database you might be able to have access to the DCM data that has not been published, that could be useful in your background." (Researcher Practitioner Trainer, 16)

The above quote indicates that one of the requirements is to have access to both type of DCM data that is collected within practice development (e.g. routinely collected data) and research contexts. While DCM data collected within research context is required to adhere to the relevant data quality and security requirements, the use of routinely collected DCM data for secondary research purposes might raise concerns, as will be explained in Chapter 8.

Nearly all study participants were able to perceive the use of DCM data for secondary purposes within the data warehouse apart from Practitioner 23, who

specifically mentioned that she/he could not envisage any use of the integrated DCM data.

"I am not sure, if this data [e.g. DCM data collected from various organisations] be of any use for us as we require the data collected from our own setting." (Practitioner, 23)

A possible explanation of this might be that the role of Practitioner 23 was limited in terms of further analysing and using DCM data, as she/he was not based in a care-providing organisation and had been working on DCM projects with other people only as a supporting mapper with basic-level certification.

"I worked with the lady who was the only one within her organisation who would do the mapping and she wanted to map and she needed someone to work with her and I am always looking at the opportunities to map to further my learning. So I went along and we worked together so it was not a paid arrangement, it was part of my development, and also to support her. So that's how I have used DCM so far." (Practitioner 23)

While only one participant mentioned this, the finding was important in suggesting the use of a prototype, informed by the current study findings, in the future to gather users' requirements. This is suggested following the observation that, while this study has identified a number of requirements, there might be a few individuals or organisations who might not be able to perceive their requirements. In this case, it is suggested that a DCM data-warehouse prototype be used to gather their future requirements. Developing a prototype requires an initial set of requirements, or ideas of users' needs. Due to the lack of any existing relevant knowledge in the field, it was not

possible to develop such a prototype for use within this study for the requirement analysis process.

In summary, the DCM researchers were enthusiastic about having access to integrated DCM data from a resource (a data warehouse). They mentioned that a DCM data resource would provide them with useful data on dementia care without the difficulty of collecting it first, as they find DCM a resource-consuming process. Furthermore, they indicated their interest in having access to integrated DCM and additional data for their research explorations for a variety of purposes, such as multidimensional analysis and comparisons, and to have a resource that could provide them with unpublished DCM data including practice-development data collected within organisations. However, the availability of such data in specific formats is important in meeting the DCM researchers' needs and such a requirement could have data quality implications for a future data warehouse. The above data indicate that while secondary research is one of the potential uses of DCM data, the researcher community will be one of the potential data warehouse users, who would require DCM and additional information in multidimensional format for exploration and comparative analysis from a future data warehouse.

7.5. Key findings and contribution to knowledge

In the absence of any existing knowledge about mappers' views regarding their secondary use of DCM data, this chapter has contributed original knowledge in pursuit of the argument that there could be three potential uses for a data warehouse. By meeting the first objective of the study, one of the key findings, therefore, is that three categories of mappers (trainers, researchers and practitioners) highlighted their intentions to employ DCM data for potential secondary use. This finding can be explained within the

context of the theory underlying the technology acceptance model (TAM),
which

assumes that a system's potential usage can be determined by the users' intentions (Tung et al. 2008). If users see its potential usefulness, they will probably use and accept the future system. Further, He and King (2008: 306), in their meta-analysis identifying the significance of user participation for information-system development, argue that "use intention" is an important outcome, based on which the success of an information-system development (ISD) project can be measured. They define 'use intention' as the "tendency and willingness to use a system" (He and King 2008: 306). He and King cite a number of empirical studies which report users' perceptions regarding a new system and its possible benefits, any concerns they have about it, or any resistance they have towards it. These studies see users' perceptions as important constructs of ISD.

Furthermore, Nieboer and colleagues (2014) add that perceptions and values of care professionals are critical in successfully implementing technology in health care. In this study, mappers' highlighted their intentions to use a future data warehouse for potential secondary uses, implying that they could be the potential users of the system that is specifically designed and built for their identified potential secondary use of DCM data. Their intention to use DCM data for secondary purposes also emphasises the significance of historic and integrated DCM data and its dissemination for secondary use, thus underlining the need for a data warehouse for this purpose. This finding re-emphasises the main argument made in my previous study (Khalid 2010), that there is a need to design and develop a data warehouse for DCM data to support its potential secondary uses.

Further, this is the only study to date that has explored mappers' perceptions of potential secondary uses of DCM data. For example, while there was a mention of the potential use of the DCM data for benchmarking within the literature, this study has gone a step further to explore the potentiality of DCM data for this purpose (Chapter 2) and provide empirical evidence (this chapter) that explains potential user perceptions regarding the use of this data for benchmarking. This information is significant in terms of designing a future data warehouse that meets the practitioners' need to share DCM data for benchmarking for the purpose of sharing and learning good practice. However, the study also suggests collecting additional data alongside DCM data to meet the users' perceptions of using such data for benchmarking.

The data presented in this chapter also indicate the DCM trainers' requirements for accessing mappers' administrative data in order to assess potential issues and to provide support. While only one other study (Douglass et al. 2010) exemplifies the DCM trainers' intention to learn about mappers' issues in using DCM, this study suggests designing a data warehouse for these specific potential users to give access to the historic and integrated data for further analysis. However, further research is required to explore the availability of such information within the warehouse.

While the previous study (Khalid 2010) that I undertook highlighted the technical potential of a data warehouse to provide a DCM data resource for research purposes, this thesis has identified that such a data resource is also a user requirement. DCM researchers' perceptions provide: an in-depth knowledge about their requirement of DCM data for multidimensional

analysis; access to both practice-development and research data; the data that is

required in its raw and aggregated form (data granularity); and an expectation of viewing data collected from various countries and over a period of time. In the absence of any user-identified existing knowledge of secondary use of DCM data, such information is an important first step towards designing a future data warehouse that meets users' data requirements for research purposes.

While the interview data highlighted (as presented in this chapter) that there could be three potential secondary uses for DCM data, additional interview data was collected to further explore the users' information requirements, issues and concerns that emerged regarding the secondary use of DCM data within a research context. Chapter 8 explores these issues and presents users' information requirements for a data warehouse within a research context.

7.6. Summary of the chapter

This chapter explored potential uses of DCM data within a data warehouse, one of the objectives of this study. Three main potential uses were evident in the interview data. The first use was expressed by the DCM practitioners, who envisaged the secondary use of DCM data for benchmarking purposes, where they perceived it for both competitive and comparative purposes. While the DCM practitioners' perceptions of their secondary use of data were evident, the need to have additional data alongside DCM data explaining the change processes was argued.

The DCM trainers, who showed an interest to use DCM data and mappers' administrative data to identify the needs and support for mappers, expressed

the second use. Their interest in accessing aggregated DCM data for learning

about good practices in dementia care for teaching and training was also evident. The DCM trainers' potential secondary use of DCM data also identified the need for collecting and storing mapper related data within the warehouse. The third use was expressed by the DCM researchers, who showed interest in having access to the integrated DCM data within a data warehouse for research purposes. Their perceptions of using DCM data for secondary purposes reflected the need of collecting data from various countries and care settings for comparison and exploration purposes. The identified potential uses of DCM data suggest that there could be three type of potential users for a data warehouse, researchers, practitioners and trainers.

Based on the identified potential secondary uses of DCM data, this chapter also argued the importance of identifying and collecting various types of data that need to be stored within the warehouse. This includes additional data alongside DCM data, such as information about care settings, mapping participants, change processes in care, and mappers' administrative data.

This chapter then highlighted the main findings, the contributions to knowledge that these make and their implications for further research.

8. Users' Information Requirements

8.1. Introduction

The previous chapter explored the potential uses of DCM data for secondary purposes. It highlighted some of the information requirements and identified the type of data that users would potentially need for particular uses within the warehouse. This chapter offers further exploration and analysis of users' information requirements that emerged within the context of the potential use of DCM data for research. It will meet the second objective of the study, which is identifying data that needs to go into the warehouse to meet the warehouse's potential use. While presenting users' information requirements, this chapter also highlights their relevant concerns and issues. This meets the third objective of this study, which is identifying issues and concerns related to the identified secondary use of DCM data within a data warehouse.

While the DCM researchers' views form the major part of the category presented in this chapter, other study participants such as trainers and practitioners' views are also included where applicable. For purposes of clarity, while referring to the study participants, each type of mapper (categorised as DCM researchers, practitioners and trainers – based on their highlighted potential uses of DCM data for secondary purposes) is specified according to their user categories when referring to their specific data requirements. This chapter will now explore the users' information requirements for a future data warehouse for potential research purposes.

Users' information requirements could provide an indication of the type of data that need to be stored within a data warehouse, an objective of this

study. Bruckner and colleagues (2001) assert that such requirements can emerge by exploring the type of information that the users need from the warehouse. The interview data highlighted that the researchers indicated two types of information that could facilitate their use of DCM data for research purposes (Figure 14). The first type of information was required for analytical purposes (data content information requirement) and the second type of information was required to describe the data content stored within the data warehouse (metadata information requirement). Both types of information requirements will have implications on the warehouse. From a technical perspective, the data-content requirements should be part of a data model presented as facts and dimension tables. Metadata should also be stored within the warehouse alongside DCM data. This could be stored in separate files or as part of the data model and could be extracted directly by the user or supplied by the data warehouse providers on request. The details of users' information requirements and their implications in terms of the type of data required to store within the warehouse are explained next.

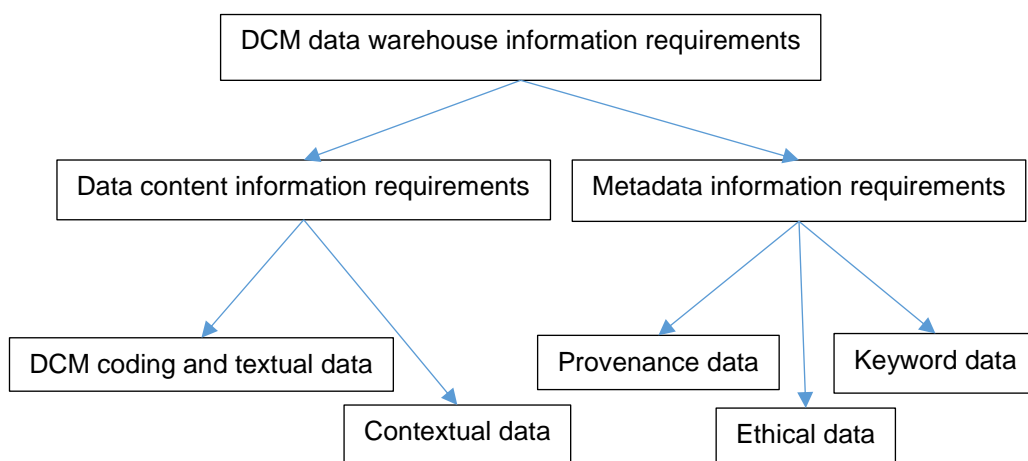


Figure 14: Category 'DCM data warehouse Information requirements and its sub-categories

8.2. Data content information requirements

This sub-category explains the user requirements that are related to the data-content within the warehouse, which can facilitate the users' analytical requirements for the DCM data. The researchers expressed the need to have access to the DCM data (coding and textual data) and contextual data (additional data collected alongside DCM data) that could facilitate their use of data within a data warehouse. These requirements are explained next.

8.2.1. DCM data (coding and textual)

As was mentioned in Chapter 1, in the context of this study, the term 'DCM data' refers to both the coding data (BCC, ME, PDs, PEs) and textual data (qualitative notes) that are captured alongside coding data. The interview data highlighted that the study participants considered both coding and textual data important within the context of its reuse, thus suggesting the requirement of collecting this data within the warehouse to meet the users' information requirements. First, I will explore the coding data items requirement and then I will explore the significance of qualitative data that is collected as part of the DCM data.

There was a mutual consensus among all researchers who took part in this study that a future data warehouse should have all the coding data to facilitate secondary analysis. The below quote reflects others' views in this regards.

"I think whatever is captured during mapping should be stored within the database and the data warehouse. I think it will be good to have all DCM data." (Researcher, 13)

In this case, while a data warehouse needs to store all DCM data including the coding and textual data, it is imperative to explore the levels of data granularity

that need to be stored within the warehouse. The levels of data granularity can be interpreted from users' views about the type of information they would require from the data warehouse (Ponniah 2001).

The interview data highlighted that DCM coding data will be required on both highest and lowest granular levels. The highest granular level means that the users will be accessing data in its detailed format. For example, each type of BCC, ME, PD's and PE's can be accessed according to the individual mapping sessions in a particular type of care setting and in a particular location. For example, from an exemplary quote below, it could be suggested that the study participant would be analysing the common type of behaviours experienced by people with dementia (for example, identifying the most common BCC within the data).

"I would be very fascinated in being able to extract back the information [DCM data] to see what behaviours people [residents] are experiencing in particular". (Researcher Practitioner Trainer, 17)

DCM data stored at highest granular levels can provide the opportunity to the users to manipulate, rearrange and summarise data based on their own needs and requirements.

"I would really like the opportunity to rearrange the data. That is very important because, depending on the research question you have... That is how you want to look at the data." (Researcher, 4)

The data stored in a detailed format can further be processed in aggregated and summarised formats. For example, ME's can further be calculated into the WIB score that can describe the overall wellbeing or ill being of the person with

dementia. Further, DCM data can be aggregated according to its location for comparison purposes.

"I think it would be good to see the data of other researchers from different countries. I have collaborated with the German group and the group in Holland, so hopefully we will compare our findings after we have done our separate studies. If the same opportunity is available by the data warehouse, it would be fantastic." (Researcher Trainer, 11)

This finding strengthens the argument made in my previous study (Khalid 2010) about storing DCM data within the warehouse in both its highest and lowest granularity forms. This suggests that data models proposed in my previous study (Khalid 2010), as shown in figures 4 and 5 (Pg. 70 and 71), seem suitable to meet the user requirement. While the fact tables (FactDCM and FactWIB) in the previously proposed data model for a DCM data warehouse seemed appropriate to meet the user requirements for analysing data at the both highest and lowest granularity form, the dimension tables need to be reconsidered. The dimension tables will be identified from the users' contextual data requirements, as presented in Section 8.2.2.

Alongside DCM data, the researchers also expressed the need to have access to the qualitative data collected during mapping. As some researchers were currently involved in qualitative research studies and were therefore able to foresee the potential use of DCM qualitative notes within their research studies.

"I think the qualitative notes have real potential to use in research studies." (Researcher, 27)

The interview data also indicated that the DCM researchers' interest in qualitative notes was due not only to the significance of data for qualitative

research but also to the fact that the qualitative notes formed an important part of the DCM coding data. Not just the DCM researchers but also nearly all study participants (27/29) mentioned the importance of the qualitative notes, saying that they form a significant part of the observation process. According to the study participants, the notes provided context to the DCM coding data. This means that the notes captured during the observation period were helpful to the mappers in understanding the context within which the DCM codes were chosen.

“I take notes on what is really happening in the PDs and PEs. I think I need a little bit more than code just to present the results to the staff. And then I take some notes... you know, reflections during the mapping, when I see things happening in the environment that may influence the patient behaviour. And sometimes I get some bright ideas that I can use in the feedback, I take notes about that. It’s several types of notes that I take.” (Researcher Trainer, 11)

This is also the reason why the activity of collecting notes alongside the DCM coding data was considered as

“... a scene-setting...” (Researcher Practitioner Trainer, 16)

This finding is in line with what has been reported in other studies, which also signifies the importance of storing qualitative notes alongside the DCM coding data in an electronic format. For example, Jones and colleagues (2014) found in their study that mappers recognise qualitative notes as important as the DCM coding data. However, Jones et al.’s (2014) study emphasised the need of the storage of qualitative notes in an electronic format for the purpose of the primary use of DCM data.

During the DCM training course, all mappers are taught and encouraged to take notes in order to provide additional supplementary information and add richness to the more structured coding data items (Bradford Dementia Group 2005). This information can relate to a variety of different things including supporting information to help them recall why a specific code was recorded in a specific situation. Further, if a mapper is unable to decide on a code during observations, particularly when participant behaviours are complex, they can write the details of what they observed in the notes and can then review these later to support allocation of a code. Regardless of the way in which the DCM is used, all study participants mentioned their collection of notes. However, the type, volume and nature of the notes were subjective and varied among mappers. One of the participants, Researcher Practitioner 9, maintained that she/he collects a vast amount of qualitative data during mapping. She/he responded to the question, regarding the type of information she/he collects as part of the qualitative data, as follows:

“Everything we saw, absolutely everything. So one thing we found was that the dementia care mapping, it can be a little black and white, so with the free notes we would write everything that we saw.”
(Researcher Practitioner, 9)

The notes helped to record the situation in plain natural language, to facilitate the mapper to analyse the whole observation period. This was another reason why note taking was considered as an important activity during DCM observations. Most of the study participants echoed what Practitioner Researcher 3 mentioned about the importance of notes as information alongside the coding data collected during each mapping session.

“So when we start to map what is going on in the environment and what kind of picture is where the mappers are being... that goes into the reports and their notes. We can empirically write down what is PE-related in the environment and we make sure that it goes into the report and it gives the reader and the team the chance to see what is happening in the environment. I think that is what we realised when we map, that if somebody is sat in a room with no music and visual stimulation, then you are writing it down, putting it in a report and feeding it back but it is not in a graph or something. You never know if that is going to be picked up on. I think that is what we try to get down as much as possible because it shows what people were doing and where they were doing it. Yes, that goes in the notes or the information that we gather.” (Practitioner Researcher, 3)

The notes were not just used to capture the mappers' reflections of the coding process but also as a means to capture the contextual information about the environment within which mapping was taking place for example, the noise level, heating, physical features etc. While giving an example from her own experience, Researcher Practitioner Trainer 16 explained that qualitative data could capture the in-depth information about the mapping environment.

“For the reminiscence group, I was describing when the volunteers came in. They set up a table and brought the cabinet in, put out the material... just describing what was happening while I was mapping. So I could see how people with dementia are relating to that. For one group, I was mapping where the TV was left on with the full volume on and it was quite disruptive for the group. That's quite important even though you are not coding that, but it's something that was happening that had quite an impact on the activity. Which is above and beyond your coding, I think.” (Researcher Practitioner Trainer, 16)

The above data show that collecting qualitative data was a common activity of all study participants. However, as mentioned above, the type of information collected within these notes was very subjective and varied among mappers. This can create consistency and quality issues in relation to the data that need to go into the data warehouse. Further, notes are captured in unstructured and textual format, which makes it challenging and complex to de-identify and share for reuse purposes (Meystre et al. 2010; Smith et al. 2013). However, a number of tools exist, and further research is ongoing, to identify ways of anonymising unstructured textual data to make it available for secondary uses. For example, ResearchOne is a health and care research database within the UK that stores de-identified clinical and administrative data and contains up to 28 million records. It has developed a research tool to anonymise free-text data from health records. This tool strips off the identifiable information, such as a patient's or a clinician's name from the medical notes, so that the researcher receives anonymised yet valuable health information from health records (ResearchOne 2012; Smith et al. 2013). In order to enable the use of qualitative notes for secondary uses, further research is required to develop tools and methods for complete anonymisation of the DCM qualitative notes.

Despite the subjective, varied and unstructured nature of qualitative notes, all participants from all three mapper categories (practitioners, researchers and trainers) mentioned the importance of qualitative data alongside quantitative DCM data. They saw the importance of qualitative notes alongside coding data in an electronic and integrated format. This requirement was concerned to the data management systems that store DCM data for primary uses.

However, the researchers also mentioned value of qualitative data within the data warehouse for providing context to use the coding data and for qualitative research. This finding suggests the need to collect qualitative notes alongside

coding data and to store them in an electronic format for secondary analysis. As will be explored in detail in Chapter 9, while a large amount of qualitative data is collected during mapping, there is currently no mechanism to store such data in an effective format.

In summary, the DCM researchers saw value of using all coding data items, and the qualitative notes captured by all mappers during each mapping session. While qualitative data was considered important for both qualitative analysis and providing context to coding data for secondary analysis, storing and analysing such type of data is challenging within a warehouse. For example, qualitative data is unstructured and inconsistent in nature, which can make it complex and challenging to de-identify for effective analysis. The evidence within the data also revealed that qualitative notes are seen as an important part of the coding data and therefore need to be collected and stored in an integrated format to support potential secondary uses for research. Therefore, taken together, coding data and qualitative notes make up the complete DCM data, which all users felt must be included in the future data warehouse.

8.2.2. Contextual data

The DCM researchers expressed their need to have access to additional data alongside DCM data from the warehouse. According to them, the additional data is required to provide context to the DCM data. The interview data highlighted that the contextual data could concern the mapping participants (observed individuals), care setting (where the observations take place) and staff members (who were part of the mapping), which will provide

in-depth exploration of the DCM data. According to the researchers, without this

contextual data, it is very unlikely that they would find the DCM data useful for secondary research.

“I think you should think about this really hard upfront. Otherwise, the data [DCM data] cannot be used, if you don’t have all the required information. Like use of medication and all that information should have been in the database.” (Researcher, 13)

The researchers mentioned that the contextual data alongside the DCM data could provide in-depth and meaningful secondary analysis. It can enhance the value of data as the researchers can ask more questions and explore the data from various angles. When one researcher was asked if she/he would like to access the DCM data from the data warehouse, her/his answer was as follows:

“I think it depends what other information is available on the study databases [data warehouse]. If it was just DCM data with nothing else, then I think from... As a part of the dementia organisation, it would be valuable for us but as a researcher it would not be seen as particularly useful. But if you have neuropsychiatric inventory data, information on cognitive data, functional data, alongside, I think that would allow you to ask some interesting questions.” (Researcher, 20)

The above data indicate the significance of storing contextual data alongside the DCM data within the warehouse. The interview data highlighted that there was a mutual consensus between the researchers in regards to the type of data that could be considered as contextual data and that would need storing within the warehouse. Together, all researchers mentioned the need to see additional data about people with dementia who were mapped, for example, participant’s age, gender, type of dementia, level of dementia, type

of medication they use or any other data that could provide information about
the

mapping participants' stay within the care setting for example, length of stay etc.

"I would like to know about the people with dementia, how they are in a cognitive state and how they are in active daily life... or something about NPI score. As a researcher, I would like to know something more about the people that are mapped." **(Researcher, 4)**

"All of the participant-related characteristics, such as age, gender, type of dementia, are really important. These are taken into account when you analyse your data. You have to have this data. Only patients' DCM data is not enough." **(Researcher, 13)**

According to the researchers, participant related data is usually taken into account while analysing DCM data. This can also be verified from the literature, which highlights that some of the collected attributes which are related to the mapping participant have either direct or indirect impacts on the wellbeing of people with dementia, thus influencing their WIB score (Innes and Surr 2001; Edelman et al. 2004; Kuhn et al. 2005; Sloane et al. 2007). One of these attributes is residents' dependency levels. A study conducted by Edelman and colleagues (2004) with mapping participants from special care facilities, assisted living facilities and adult day care centres, found that low mean individual WIB scores are associated with both high levels of cognitive impairment and increased activities of daily living (ADL) dependency. Similarly, another study (Thornton et al. 2004) reported significant correlations between WIB scores and individuals' total dependency levels and cognitive and behavioural functions. They found that wellbeing levels of those individuals living in continuing care settings and day

hospitals are significantly higher, as they had lower dependency levels and fewer cognitive and

behavioural issues. Chenoweth and Jeon (2007) also report an association between lower WIB score and reduced physical function. Both the researchers' requirement and the literature's indication of the associations of the DCM data with mapping participants' characteristics suggest collecting and storing such data within a future data warehouse to meet users' additional data needs for analytical purposes.

The information about the staff who were on duty during the mapping session was also considered significant for analysis purposes. Agreeing with other researchers, one of the researchers mentioned the importance of knowing the levels of training and education of the staff on shift at the time of mapping. According to her/him, their status as trained for person-centred care can have an impact on the quality of interaction they have with the mapping participant. This can change the whole perspective of looking at various aspects of DCM data that reflect patients' behaviour, engagement and interactions.

"I think it is important to know about the care staff, if they are educated or not, and what kind of education they have, if they are trained in dementia care or person-centred care. Yes, because that would be helpful to know as well because... for example, I can imagine that people who are trained or have a higher education... they interact differently with people with dementia than the people who are not trained and that would be very interesting to study that." (Researcher, 4)

"Maybe you can also think about... I don't know if it goes too far... but a short description of wards, like how many people are there, what the shifts are and what is the education of the nurses on the wards. It does not have to be every single detail but you need to have an

idea... what kind of population is working with the patient.”
(Researcher, 13)

The DCM literature also indicates that the number of staff that are caring for people with dementia during a mapping session can also have indirect impact on mapping participants' wellbeing. Innes and Surr (2001) argue that the lower the staff's involvement with the residents, the fewer the signs of wellbeing shown by the residents, as they will receive less attention from staff or have minimum involvement in activities. Innes and Surr (2001) found that low staffing levels contribute to generating more 'PD: Ignoring' because they are busy or overstretched; meanwhile, more staff means more attention, more PEs and, therefore, higher WIB scores. Hence, staff numbers have an indirect impact on individuals' WIB scores.

The researchers also mentioned that they required access to the information regarding the care setting where DCM takes place. This information consists the type, location and size of the care setting. The literature points out that there exists variations within the wellbeing and ill-being score based on the type of care settings. For example, in day-care centres, the levels of wellbeing are higher than in hospital wards (Kuhn et al. 2002). The researchers, who took part in this study, perceived that this information would help them in analysing DCM data within the right context of care provision.

“I would like to know the type of care settings where the map has... taken place, whether it's a day centre, care home or NHS settings. I guess it might be useful.” **(Researcher Practitioner Trainer, 16)**

Another researcher mentioned the significance of capturing and storing the length of mapping time along the DCM data. According to her/him, this piece of information could provide an opportunity to evaluate the arguments about the quality of data when using long or short maps. Further, she/he also

mentioned the importance of capturing the number of participants who took part in mapping. According to her/his view, these details could provide rich information for analysis in terms of what was working and what was not working.

“I suppose the length of time the maps have taken place, because there is an argument about that, whether you get better quality data from longer mapping period or shorter ones. We think there is value in shorter maps and often that shorter maps are more achievable. So I would like to know how long the mapping data was for, how many participants were mapped for that one map. Because there is huge variation depending on skills and levels, I guess. There might be a six-hour map for one individual but then there might be a six-hour period when mappers are mapping five or more individuals or less.”
(Researcher Practitioner Trainer, 16)

It is indicated that the purpose of mapping derives various aspects of DCM for example, the time of observations, the length of observations and the use of data (Bradford Dementia Group 2014). The review of the literature highlights that there is a great deal of variations in time, length and use of DCM from one study to another. This level of detail alongside DCM data can give a complete context within which the data was collected. The researchers therefore felt there was importance of accessing such details related to each mapping. It could therefore be argued that each DCM study or use should be categorised based on the purpose of mapping and this information should be stored within the warehouse.

The contextual data requirements have highlighted the need of additional dimension tables within the data model for a data warehouse proposed in my previous study (Khalid 2010). As was mentioned in Chapter 3, dimension

tables, linked to the fact table (s), provide information for dimensional analysis. The need of contextual data will provide the users with the ability to conduct multidimensional analysis. This requires further work to enhance the existing data model to include dimension tables containing information about the mapping participants, care settings, mapping details and staff members. Alongside technical implication for enhancing the DCM data model, availability of the contextual data for a future data warehouse is also a potential area for further research.

As was also mentioned in Chapter 1, the additional information alongside the DCM data is very rarely collected. Only research studies using DCM report the collection and use of contextual data, the type of which further depends on the purpose of the study. While Brooker and Surr (2005) also suggest capturing contextual information during DCM evaluations, there is currently no motivation to capture such additional information alongside DCM data, thus hindering the in-depth analysis of data through the consideration of various attributes. The limitations of DCM – in not collecting confounding variables commonly associated with dementia– are also highlighted in the literature (Beavis et al. 2002).

Within this study, the requirement for contextual data alongside DCM data for secondary use re-emphasises the need to collect such important information and to develop mechanisms that can facilitate the collection and storage of additional data alongside the DCM data. Within a secondary analysis, the DCM data and additional information about care settings, patients and staff could be linked and the information explored further to investigate potential

associations that may only be identified through combining and analysing data

from a large number of maps across a wide range of settings. The integration of such information could facilitate the design of specific tailored interventions for similar patient groups to improve their care. Further, DCM data, along with additional information, could provide an opportunity to perform in-depth analysis regarding what, where, why and how the well-being of people with dementia is influenced in formal care settings.

In summary, evidence from the interview data shows that the researchers see importance of contextual data being available alongside the DCM data, thus making the dataset more desirable and useful for secondary analysis within the warehouse. Further, such information will allow the users to conduct multidimensional analysis of DCM data by considering diversity of contextual data. The contextual data includes types of data that concerns mapping participants, the care settings and the staff members. Further, knowing the mapping purpose and details of each mapping session in terms of the number of people mapped, the number of mapping hours and the number of mappers involved was also considered significant by the researchers in making the DCM data a complete dataset for secondary analysis within research context. The evidence not only identifies the potential data attributes of contextual information for the data warehouse but also provides the rationale and context the study participants have identified for such information. Both the DCM data and contextual data form the data-content for a future data warehouse. However, the availability of such data-content is important for a data warehouse to meet the users' requirements, which is suggested an important area for further research.

8.3. Metadata information requirements

This section presents the users' requirements regarding the metadata about data-content that will potentially be stored within a future data warehouse. Metadata is information that describes various aspects of data-content that is stored within a warehouse (Deelman et al. 2010). Deelman and colleagues (2010) argue that metadata can consist of information about the means of creation of the data-content, the purpose of the data-content (why the data was collected), the time and date of creation, the creator or author of the data-content and the location on a computer where the data was created and used. Metadata is significant in providing the ability to interpret a particular data item (Deelman et al. 2010). Such information first describes the data-content and further enables its reuse by providing transparency in data (Simmhan et al. 2005). Simmhan and colleagues (2005) argue that the usefulness of any dataset can be increased by describing the content and context of the data. For example, a webpage might have metadata information that specifies how the webpage was created, what tools were used to create it, what language was used and where to go if you need more information related to the subject area. This allows the computer browsers to enhance their use experience.

Within this study, the need for metadata information became apparent when the DCM researchers mentioned that, to use the data-content from the warehouse, it is important for them to know the relevant information. This information need, in technical terms, could be called a metadata information requirement. There are two types of metadata: structural metadata, and

descriptive metadata. Structural metadata explains the design and specifications of the structures of the data (how data is presented) that is

stored within the data warehouse (Sen 2004). Usually, the technical users (e.g. individuals who manage the data within the data warehouse) are concerned with this type of metadata. Descriptive metadata on the other hand explains the information about the data content of the warehouse (e.g. in terms of what data is available and who has made this available), which concerns those who would probably use the data content, that is the data users (Sen 2004). For example, the UK Biobank provides detailed information about each data field explaining the necessary background about how the specific measure (e.g. participants' exposure and outcomes) was taken (UK Biobank 2007b).

The researchers' requirements were concerned with descriptive metadata rather than structural metadata, as they were the potential data users rather than the technical users of a future data warehouse. The users' metadata information requirements are categorised further based on the type of information required. These are explored and analysed next.

8.3.1. Provenance information

Metadata is used for a range of purposes. One purpose is to provide provenance to the data content. The term provenance refers to establishing information about the origin of any dataset (Glavic and Dittrich 2007). The need for such information is instigated when the distance between the data collector and the user increases; in situations where the data user is not the one who was involved in collecting the data. As a result of this, the issue of data trustworthiness escalates (Janowicz 2009). Goble and Stevens (2008) summarise several applications of provenance information such as, data

quality estimation, audit trail of data, replication purposes, attribution for copyright and ownership and informational, that provide context to interpret

data. The DCM researchers' interview data highlighted their need for provenance information for two main purposes, estimating the quality of DCM data for secondary uses and informational purposes to ensure the data is interpreted within the right context of its original collection. These are discussed next.

According to Simmhan and colleagues (2005), one of the major catalysts of provenance information requirement is that users want to establish the quality of data, as the secondary data mostly encompasses the issues of data trustworthiness regarding the data that is collected by others (Janowicz 2009). Simmhan and colleagues (2005), therefore, argue that in order to trust the quality and reliability of someone else's provided data, the users usually require access to the information that provides provenance of the dataset in terms of who collected the data and where and why it was collected. This is the reason why metadata, as provenance for data-content, is required within a data warehouse that stores secondary data. Within this study, the researchers also mentioned their requirements regarding the provenance of DCM data in terms of 'who' collected the data.

The need for provenance information emerged when the researchers described the DCM data within the warehouse, potentially stored for secondary uses, as 'others' data', referring to data that might not be collected by them as mappers for the specific mapping session conducted for the primary purpose of data collection.

"I am worried that, if I am going to use data from the other mappers, how the quality of data I am going to look at. So that is something that

*bothers me, that if you have to have a big data warehouse in which
you*

have all this data from different sources, if every source can be of good quality for research purposes.” (Researcher, 4)

“You make reference to other people’s [DCM] data but you cannot be sure of the reliability of it. You would never be able to be certain of the reliability of other people’s data.” (Researcher Practitioner Trainer, 16)

The term ‘others’ signifies the ‘mappers’ and ‘organisations’ that will potentially provide the DCM data to the warehouse for secondary research. As argued in Chapter 1, the evidence from the literature shows that DCM is mostly used for primary research, where the mapper is, either directly (through observations) or indirectly (hiring other mappers to conduct observations), involved in collecting and using DCM data. The use of DCM data for (secondary) research is associated with the use of data away from its primary purpose and time of collection. Most importantly, the user of this data might not be the mapper who was involved in the mapping process. This indicates why the word ‘others’ was associated with the data that will potentially be collected from other mappers and used for secondary purposes.

“It would become quite woolly, so when it [mapping data] has been observed and reported by another organisation, the quality of the data can be questionable.” (Researcher Practitioner Trainer, 17)

There is a consensus across studies that, in order to increase users’ belief in the data within the warehouse, the inclusion of metadata about the quality of the data is imperative (Fisher et al. 2003; Moges et al. 2016). Moges and colleagues (2016) further assert that the available metadata information about the quality of data can help users to make decisions about the use of

data for specific purposes. In many studies, the concept of data provenance, as a type

of metadata, is emerging as an important way of enhancing trust and belief in the data for the data-warehouse users (Simmhan et al. 2005).

The interview data pointed to the researchers' concerns being related to the reliability of DCM data provided by other mappers. Data reliability was considered significant for ascertaining the quality of DCM data for secondary uses. The interview data highlighted that a mapper plays an important role in producing and maintaining the quality of DCM data, thus producing reliable data for potential secondary uses. Role of the mapper for providing good quality DCM data is presented and analysed in Chapter 9. The researchers mentioned that access to information about mappers' training level, experience level and their inter-rater reliability (IRR) score alongside the DCM data would provide them with evidence to estimate the quality of the data within the warehouse, thus making the data believable for potential secondary uses.

"I want to see how the mappers have scored, if they are two mappers, have they correlated with each other to see if it [score] matches."
(Researcher, 13)

This finding is in line with what others have suggested to be as important information within a data warehouse for establishing the quality of data. For example, UK Biobank (2012: 2) explains in its data management and sharing plan that "UK Biobank will not impose its own quality criteria on the data, but rather will describe the origin of the data and the methods of data collection so that the data quality can be judged (by the researcher) on the basis of the particular research question being addressed".

Further, the researchers also mentioned their requirements of establishing 'why' the DCM data was collected for original (e.g. primary) purposes. This refers to the need of provenance for 'informational' purposes within the warehouse. Simhann et al. (2005) explain the purpose of this need by maintaining that when data is originally collected for a different purpose, provenance information helps to interpret the data in the context it was intended. The interview data highlighted that knowing the purpose of mapping can help users to make an informed judgment on the reuse of the data. For example, if DCM was conducted for evaluation purposes for developing the care practice, the researchers mentioned that, by knowing this purpose of the mapping, they could make judgments on why, how, when and for what purposes they might use that data. According to Simhann et al. (2005), the information about 'why' data was collected enable the data user to decide if the data meets the requirements of their specific application or reuse of data.

"I think it is really useful to know the themes of the maps, whether the mapping is for particular interaction or whether it's for a general service evaluation or service improvement." (Researcher Practitioner Trainer, 16)

The researchers perceived the importance of the provenance information in terms of mapping purpose to give the DCM data a context for interpretation. For example, if the dataset was collected with the aim of assessing the quality of care provided within a hospital setting, the participants expressed the need to see the information or characteristics of the hospital setting, including its size, ward type, and the type of service provision. This

information would help them to interpret the data and the findings by taking into account the type of care setting from where the data was collected.

“I would want to know what ward type they were from, what hospital they were from, were they from a care home... very much quite specific detail about where the DCM participants were from. And whether it [DCM data] was taken from a general hospital, district hospital, mental health hospital, care home, or a day unit etc. And the staffing levels... Yes, the staffing level, the patient-to-staff ratio, I definitely want to know that to be able to make a judgment.”
(Researcher Practitioner, 8)

The need for accessing information about the ‘mapping purpose’ is also highlighted as contextual data within data-content information requirements (explained in Section 8.2.2). The need for collecting such information for provenance further signifies its importance. It is therefore suggested to categorise DCM data based on its mapping purposes to be able to store such information both as data-content and metadata within the warehouse.

The users’ requirement of provenance information for DCM data to establish ‘who’ and ‘why’ of the data is in line with what Bevan and colleagues (2013: 1757) experienced in their study of investigating issues related to the secondary use of publically available data on associational interests (e.g. Encyclopedia of Associations). They therefore assert that researchers should be concerned with the source quality and potential biases associated with the secondary data. They further maintain that researchers should ask the following questions while accessing the secondary data from the databases: *“what was the original purpose for which the data was collected? and what is the reputation of the data source creator(s)”*? The researchers’ need for accessing the information about the mapper and the mapping purpose suggests that DCM data need to be collected with additional data that provide provenance to the DCM data to ensure its potential secondary use

for research purposes. The availability of such provenance information, however, needs

exploration. Chapter 9 covers this discussion while presenting the role of a mapper and DCM data quality.

In summary, this sub-category explains the researchers' requirement for provenance data within a data warehouse, which could allow them to estimate the quality of data in terms, by ascertaining who collected DCM data (e.g. mapper's reliability, experience level and training) and why or for what purpose the DCM data was originally collected (e.g. mapping purpose). The main requirement was to have access to the information about mapper's experience, training and IRR score and the information about the original mapping purpose as provenance data to ascertain the quality of DCM data.

8.3.2. Keyword information

Metadata also refers to the information offering the opportunity to search the data content stored within the warehouse using specific keywords. This additional information is also significant for secondary data uses, as the right data retrieval is as important as the right data storage (Tablan et al. 2008). A data warehouse stores keywords as metadata to facilitate users' data search within the resource. Each dataset is tagged with some specific keywords that describe the dataset and facilitate the searching activity by providing the right dataset. To meet users' data search requirement, there is a need to store specific keywords information or data within the warehouse that would facilitate the searching and data retrieval activity. The requirement of a searchable database and storing information within this database to facilitate the search are interlinked and therefore are explored and presented together.

The researchers agreed that they required access to a searchable database, an information-retrieval application for a future data warehouse. With this requirement, the usability and information aspects that need to be stored within the warehouse to facilitate the search needed further exploration. The usability aspect is related to how the users would like to search the data. The interview data highlighted that the researchers would like to search for data using some key parameters, thus referring to a single-text input field that was required to communicate with the warehouse in terms of retrieving the data.

“It would be valuable to me if you had a field where you could put in what parameters you were looking at from the data warehouse.”
(Researcher, 20)

There are many solutions to producing a single-text input field for a search facility within a data warehouse. Given the non-technical background of the majority of the users, an interface similar to that of Google may be most appropriate, since these types of interfaces are built for the users who have minimum technical ability for data retrieval (Tablan et al. 2008). The users are usually not familiar with the underlying structure of the data and, therefore, do not have to write a technical query language²⁵ to retrieve the data from the data sources. Considering the users' limited ability to write technical queries for data retrieval, Google and some other search engines/applications provide a facility of a Natural Language Query (NLQ) for users to write their search term/keyword or query in simple language (Tablan et al. 2008). The software then translates these simple language queries into technical language queries

²⁵ The technical language that is used to retrieve data from the databases and data warehouses. This language is only used by the database administrators or individuals who have the ability to write such languages.

that then retrieve the data from the data sources. Therefore, NLQ is considered the most user-friendly data-retrieval method (Tablan et al. 2008).

While NLQ can provide a user-friendly data retrieval facility. To enable this there needs to be an established list of keywords that represent all DCM data instances stored within a data warehouse. The interview data highlighted that the researchers were interested in retrieving data based on the purposes or themes of mapping. This suggests the storage of information or keywords referring to 'mapping purpose' within the warehouse to retrieve the relevant datasets.

"I think you would almost need somewhere to search and say that I want information on people where they have DCM data on two occasions. The required information is there and then it would be able to pick out on, you know, how many people of a database have got that level of information and then be able to sort of pick up the sub-database for the researchers who want to look at that. Because, otherwise, you will have some huge dataset which probably will be a monumental task to search out for the bit you actually want."

(Researcher, 20)

A potentially simple solution would be to categorise each DCM data by the main purpose of mapping and then use these categories as 'keywords' for the users to retrieve the relevant data sets. However, the categorisation of DCM data requires a careful consideration, as the literature indicates a multiplicity of purposes for which the DCM data is collected (Brooker 2005).

When a data warehouse stores large amounts of data, the need of filtering data out based on the user need is a common requirement (Ponniiah 2001).

The researchers also expressed a need to filter the data from the warehouse by its place of origin (where mapping was conducted), subject (purpose of

mapping) and type of care setting (mapping site). This also suggests the requirement for further categorising DCM data, based on 'type of care setting', 'mapping purpose' and 'place of origin'.

"The user should have the option to select the data or refine their search. You should be able to select your data by country, by subject and the type of care setting." **(Researcher, 13)**

In terms of data retrieval and access, the researchers also expressed the need to re-arrange the data, thus requiring the warehouse to store DCM data in its highest granular form, i.e. the raw data without any form and level of processing. The access to raw DCM data will enable the users to manipulate and re-arrange the data based on their potential needs of data analysis.

"I would really like the opportunity to rearrange the data. That is very important because, depending on the research question you have... That is how you want to look at the data." **(Researcher, 4)**

Alongside the requirement to have detailed level data in order to re-arrange it according to their needs, the researchers also required the facility to export the searched results (data) with ease. As one researcher mentioned, the warehouse should be easy to use and it should be easy to export the data from so that they can re-arranged the data according to their specific needs for further analysis. In the quote below, one of the researchers expressed the need to retrieve the data from the warehouse.

"Easy to get export data from." **(Researcher, 13)**

In respect of data retrieval, the easy-to-use interface was also the main requirement among all the study participants. Echoing other researchers'

views, one researcher mentioned that the future data warehouse should be easy to use.

“The advantage would have been, using it, that it is really easy to use so it would have to be quite easy to use and easy to get export data from.” (Researcher Practitioner, 9)

However, what is easy to use is subjective for every individual and based on the users' computer literacy skills. In other words, if the user finds it easy to learn to use the system, remembers to perform various actions and has an understanding of the system, use of the system becomes easy and the system becomes acceptable (Kassim et al. 2012).

In summary, the interview data highlighted the requirements related to the retrieval of data from a future data warehouse. The researchers mentioned a requirement for a search box where they could write their queries or keywords. This suggests allowing the users to provide a text-box where they can search the DCM data resource using non-technical language. However, it is also important that they are aware of what is available within the data warehouse for them to retrieve. Further, the users should be able to import the data from the warehouse into a form where they can manipulate the data according to their further uses. The search process should be easy and user-friendly, as the main requirement in relation to data retrieval. However, it is important to consider the type of data that needs to be stored within the warehouse and its availability as metadata to facilitate the users' search of their required data.

8.3.3. Ethical information

Metadata also refers to the information that provides specific explanation about a dataset. The literature highlights that such type of information varies

according to the users' specific needs or system's specific requirements (Simmhan et al. 2005). This sub-category explores the requirements for the type of information that the researchers felt would be desirable within a future data warehouse.

The researchers were aware of the value of ethical considerations that need to be adhered to in relation to the data stored within the warehouse. Nearly, all researchers who took part in this study mentioned their requirement to be able to find, alongside the DCM data, the ethical considerations including ethical approval and consent from the mapping participants and organisations that were taken while collecting the primary DCM data.

"It's very important that data is collected considering all ethical aspects. As a researcher, I would like to see what ethical considerations were taken when mapping was done." (Researcher, 25)

This finding is in line with data security requirements for research data warehouses (Innovative Medicines Initiatives 2014; UK Biobank 2007a). Ethical considerations are taken into account when storing routinely collected healthcare data, as well as research data, within the warehouse. In both cases, the data provider (e.g., the organisation or the researcher) is responsible for ensuring that data is collected using ethical and legal guidelines or the correct protocol for dealing with a patient's personal data or any data for secondary use. This information is stored as metadata and is made available while disseminating datasets for further research purposes (UK Biobank 2007a).

In general, the researchers comply with all ethical considerations regarding the collection, storage and use of data for research purposes. Similarly, dealing with the DCM data requires consent from the people with dementia

and organisations involved (Bradford Dementia Group 2014). This consent is specific to one's own individual study or purpose of mapping. Further, the literature highlights that each research study must adhere to all ethical and legal requirements. For example, permission should be obtained from local ethics bodies for the research to be conducted. Further, assent should be sought from those who have the capacity to give their consent, from close relatives, or from the key caregiver (Chenoweth and Jeon 2007) and it should be ensured that data is anonymised for use outside the organisation (Brooker et al. 1998). The interview data highlighted that, for the secondary use of DCM data, there was also a need to take into consideration the pertinent ethical issues and their documentation within the data warehouse.

"The only thing, which is important, is that, if you collect data, you are doing it for one purpose and, if you use that data for other purposes, then you have to inform the care organisations where you mapped that you are going to share the data with other researchers. Because that is not the thing that you tell them at the start. So I think it is very important and, besides that, I think I would not include names or organisations and especially the names of the residents if they are included, yes, that is the only thing that is important to not include."

(Researcher, 4)

The above quote reflects two interlinked requirements, first involves gaining consent from the residents (people with dementia) for reusing their DCM data. This requirement seems pertinent, as the users are the researchers who need to follow a systematic process for taking into account the ethical considerations while dealing with any type of data, particularly residents' and organisations' identifiable data. The researchers' requirements regarding the ethical and legal status of the DCM data for secondary research is in line

with a report published in 2012 by Ireland's Health Information and Quality Authority (2012).

They reviewed the policies and regulations of four countries including the UK in relation to their processes of dealing with the secondary use of personal health data. According to this review, the most important and common ethical issues related to the secondary use of healthcare related data are the consent from the patients whose data will be used for the secondary purposes, appropriate de-identification of the data to ensure the privacy of the patient's personal data, and organisations' responsibility of sharing patients' data responsibly.

The researchers showed an element of reluctance to use any dataset from a future data warehouse that lacked adherence to such ethical considerations.

"I would definitely need something around consent within that data warehouse. It would make me feel that we have the right approval to use the data for research." (Researcher, 25)

The second requirement involves viewing de-identified data that do not include resident and organisation's names. A consensus was found among all the researchers that a future research data warehouse should store de-identified (anonymised) DCM data so that the organisations' and mapping participants' names are not visible. This finding suggests that it might not be necessary to gain the consent of the mapping participants. This is in consistent with guidelines on consent from the Health Information and Quality Authority (2012), which maintains that the authorisation from the patients/residents in terms of their consent for the use of their data for secondary purposes is not necessary if data is appropriately anonymised. This means the patient cannot be recognised or their data cannot be re-

identified at any time during its secondary use. While on one side the researchers mentioned that the mapping

participants' and care settings' names could be de-identified, on the other side they showed interest to see the additional data (called contextual data in Section 8.2.2) about the care settings, and the mapping participants. Even though such data has very low chances of re-identification, according to El Emam (2011), if not being careful, such additional data can also lead to the re-identification of the personal data. It is therefore imperative to make sure that the DCM data is anonymised using appropriate methods, yet includes enough information to meet the users' information requirements. This requires further research into investigating effective anonymisation techniques for DCM data within the warehouse.

Further, as DCM produces cross-sectional as well as longitudinal data, there is a need to adopt effective methods to de-identify the longitudinal DCM data. Longitudinal data is collected at various time points as part of a single study. To collect a complete set of data, data should be available and linked appropriately by flagging up the reasons and decisions taken for further data collection as well as the outcome of the study (Schuller et al. 2012). The data linkage requires an identification number that must be specific to each study, but should not identify the details of the care settings and the mapping participants involved in the study. In order to create such identification numbers and ensuring that all data is available as part of a longitudinal study, the existing data warehouse within healthcare provide solutions. For example, the UK Biobank stores longitudinal data and therefore uses a reversible or pseudonymised form of de-identification. Using reversible method, the UK Biobank as a data provider can make sure that data from all

instances is available for dissemination to the users. However, technically and ethically this

process is considered challenging (El Emam et al. 2009). The study participants also recognised such a challenge, such as researcher 20 highlighted the issue of de-identifying DCM data, where data is collected over a period of time.

“I think, as long as you don’t name the individuals and name... the care home, I think probably you would have a care-home number that you could link to. I mean the other problem with that, of course, would be that, if you have got longitudinal data, it’s very hard to anonymise it until you have follow-up data available. I think, if you want to do it in a confidential way, there would be problems. You might have to have a two-face system to one where you get the data and put it on to the database in a linked kind of way and then that is imported within an anonymised database. Otherwise, it would be impossible for you to have longitudinal data on it.” (Researcher, 20)

There was a concern raised among the researchers that if the warehouse stores DCM data collected by the practitioners, which they collect for their own practice development purpose, the data might not comply with all necessary ethical considerations, which are required for secondary use of data.

“There are ethical issues around sharing the data collected from practice development by those who are care workers within the same care service. I think that would be the other concern I would have. As researchers, we comply with very, very strict rules of... you know... ethical rules and I suppose, as practitioners, you are not restrained by the same things.” (Researcher, 25).

When Researcher 25 was asked to elaborate on this point, she/he explained it as follows.

“So, for example, you imagine if it was your service and you know you were a key worker for five or six people and you measure their care

using DCM, then I suppose, then you might not get ethical approval and you might not even get consent. You just do it as part of the care planning and I think there are issues around then what you can do with your data. The data stays as part of your care plan within your service and, if it is not shown to anybody else, then it's fine. But when you start sharing all that anonymously, I think there will be ethical considerations around that, which need careful thought.”
(Researcher, 25)

On the emergence of this concern from the researchers, as part of the theoretical sampling approach, I approached the DCM practitioners to ask about the ethical procedures they adhered to while dealing with DCM data. The practitioners' responses reflected that they would deal with the DCM data as if they would do with any other patient/resident related data they were using within their organisations. According to a general policy of dealing with patients' sensitive data, all health and social care providing organisations are obliged to follow specific rules, which include appropriate data access and security issues (Department of Health 2011).

However, there is limited evidence available to assess how and whether data collected within organisations for their own practice development adheres to all ethical and legal requirements. While the practitioners who took part in this study confirmed that they adhered to all data-protection and privacy issues while dealing with patient data, they deal with ethical issues as primary care-givers. There is no evidence that they ever asked for the permission of the mapping participants about the potential secondary use of their DCM data for future research purposes. Developing a future data warehouse to support secondary uses of DCM data will require tackling the issue of how consent can be gained from the mapping participants, at the

time of their DCM data collection, to support any future secondary research.

As argued in Chapter 3,

the use of routinely collected data, and the gaining of consent for secondary uses, are also major areas of concern reported in healthcare (Elger et al. 2010; Lamas et al. 2015).

The researchers' concern indicate the requirements of establishing the ethical status of DCM data collected within a practice-development context and whether the status allows the data use for secondary analysis. This finding suggests further research to design a consent-management process that addresses the DCM data-related consent and privacy issues. Further, data security is highlighted as a major challenge in the literature (Chapter 3), where alongside consent-management and privacy issues, data anonymisation is also highlighted as an area of concern in regard to warehousing healthcare data. While this study has identified that the DCM researchers required anonymised data for research purposes, it is nevertheless important to establish the practicality of anonymised DCM data for research and to develop anonymisation techniques that will ensure complete anonymisation. The literature highlights the role of the data providers (such as a healthcare organisation) as data custodians to ensure that a level of security governance on the data is demonstrated. This includes defining the purposes of the secondary use of data, identifying why the data is required and for what secondary use, assessing the ethical considerations such as managing the consent process and establishing the data-security controls (Hovenga and Grain 2013). However, the role of data custodian can be taken by those who will store and process data for secondary uses such as a data warehouse providers. In the context of DCM,

it is important to establish the role of data custodian who will make the decisions on the secondary use of DCM data.

In summary, the interview data highlighted the researchers' views regarding the importance of ethical considerations for DCM data that is deemed for secondary research purposes. While they highlighted their concerns of using DCM data collected for practice development purposes, the researchers further indicated the need to have access to the information about the ethical considerations taken for collection of DCM data within the warehouse. Further, majority of the researchers mentioned the need of accessing anonymised data, thus suggesting storing anonymised DCM data within the warehouse for research purposes. Future studies with more focus on DCM data anonymisation, consent-management and the role of data controller/processor within the warehouse are therefore recommended.

8.4. Key findings and contribution to knowledge

This chapter presented novel findings that explain the DCM researchers' information requirements in terms of collecting various types of data-content and metadata within the warehouse. This includes the requirements for provenance, ethical, keyword and contextual data, and qualitative notes. These findings have implications for both enhancing the existing DCM data models proposed in (Khalid et al. 2010; Khalid 2010) and ensuring that the additional data is collected alongside DCM data to meet the users' data needs for research purposes.

Another important finding is users' concerns regarding the quality of DCM data. While significance of data quality was highlighted in the previous study that I undertook (Khalid 2010), this study brings data quality related users' concerns at the forefront. According to Kumar and Thareja (2013) in the context of a data warehouse, users' views are important to establish the data

quality requirements. While Chapter 9 will be discussing DCM data quality in detail, this chapter has identified that in order to make the DCM data trustworthy within the warehouse, the collection and storage of provenance data is significant. This suggests ensuring the availability of provenance information within the warehouse prior to commence its technical design and development.

8.5. Summary of the chapter

This chapter provided original knowledge by focussing on data requirements for the potential secondary use of DCM data for research. The researchers highlighted their information requirements, which were categorised as data-content and metadata requirements. These requirements identify various types of data that need to go into the warehouse. As data-content, the study participants mentioned the importance of all coding data and textual notes for their potential secondary uses. While coding data requires storage in various granular forms within the warehouse, the storage and analysis of textual notes may create challenges as the qualitative data is usually of an unstructured and inconsistent nature.

Further, the researchers expressed a need to have access to the contextual data alongside the DCM data to make it more desirable and a richer source for potential secondary research purposes. The most desired contextual data included details of the mapping, the care staff, the mapping participants and the care settings. The interview data highlighted that the diversity of the contextual data would allow users to analyse DCM data from various dimensions, which is important in terms of explaining the activities, behaviour and interaction types of the mapping participants.

Furthermore, the researchers also expressed the need to access metadata within the DCM data contained in the warehouse. The researchers' interview data highlighted the need for three types of metadata, provenance, keyword and ethical information. Provenance information includes the 'who' and 'why' of DCM data in order to estimate the quality of the data for secondary purposes. Keyword information and ethical information also seemed important to the researchers in terms of searching for the right data within the warehouse and the data's ethical position for secondary use. This suggests that the metadata stored within the warehouse should include provenance, keyword and ethical information. However, further work is required to ensure the availability and storage of this information within the data warehouse.

9. Factors Affecting the Availability and Quality of DCM Data

9.1. Introduction

The previous chapter examined the users' information requirements for the secondary use of DCM data within a research context, arguing that, for this purpose, a future data warehouse needs to collect data-content and metadata information, including provenance, ethical and contextual data. The previous chapter also highlighted the researchers' concerns regarding the quality of DCM data. In relation to this concern, this chapter presents study findings which develop an argument that there are three factors which can potentially influence the quality of DCM data for secondary uses. Further, this chapter also argues that the trilogy of factors can also potentially influence the availability of DCM data for secondary uses.

As was mentioned in Chapter 3, a data warehouse needs to store quality data, taken from primary data sources, that is complete, accurate, consistent and relevant to meet the user's information requirements for secondary uses (Ballou and Tayi 1999). In relation to this study, the researchers' information requirements are presented in Chapter 8. Whilst the availability of quality data within a warehouse is a major requirement for a usable and successful data warehouse (Kumar and Thareja 2013), it is important to understand the issues influencing the availability and quality of data by exploring the situations within which the primary data is produced and managed.

The interview data highlighted a range of factors that can potentially influence the availability and quality of DCM data. It is extrapolated from the interviewee's responses that an individual mapper (who makes observations

and collects DCM data for primary purposes), the organisation (where DCM is conducted and data is used for primary purposes) and an electronic data management system (that stores and analyses DCM data) play significant roles in maintaining the availability and quality of DCM data (Figure 15). The roles of mapper, organisation and primary data management system are presented and explained next. This is followed by a discussion that brings these three factors together at conceptual level to discuss their implications for the availability and quality of DCM data for secondary uses.

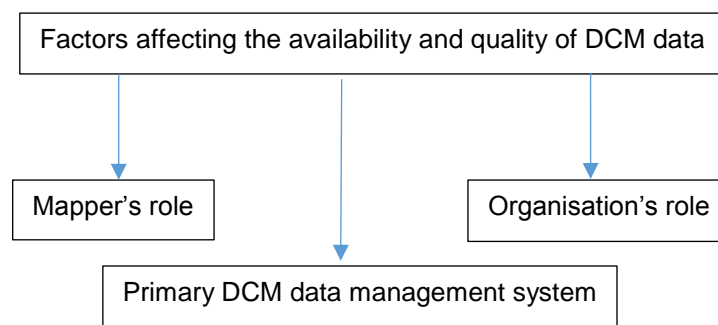


Figure 15: Category 'factors affecting the availability and quality of DCM data' and its sub-categories

9.2. Mapper's role

The interview data indicated that the mapper plays a fundamental role in producing DCM data and in establishing and maintaining the quality of that data. A number of aspects can potentially influence the quality of data that mappers produce. These are explained in detail.

9.2.1. Mental and physical presence

During mapping, a mapper performs various actions, for example, preparation to use the method, conduct of observations, capturing and processing of data and further, interpretation and presentation of the data

through verbal feedback and/or written reports. All these steps demand mappers' time,

presence (mentally and physically), and attention to collecting mapping data. The interview data highlighted that if a mapper is physically or mentally exhausted then there is a risk of quality compromise in the produced data.

As one of the participants mentioned:

“... Each day [mapping day], it is a completely different day and actually a quite exhausting day. In that case, then chances are that your map is not going to be very good, but if you know this day requires a lot of your time... you need to prepare yourself. You need to be focused, then you have a good map.” (Researcher, 13)

This finding is in line with what is asserted by Woods and Lintern (2003: 30) as “observation (during mapping) is hard work, and reliability of recording may suffer when the rater (mapper) becomes tired or feels uneasy regarding what is being observed”.

9.2.2. Consulting the DCM Manual

DCM produces observational data and mappers go through intensive training in learning to use the method empathetically and objectively (Capstick 2003). There are rules and guidelines on how to attain and maintain the quality of the collected data. The study participants mentioned that they follow these guidelines by consulting the DCM Manual regularly before and during mapping in order to ensure the quality of collected data. For some study participants, following the DCM Manual provides assurance that they will achieve a good map, which can lead towards good outcome (data).

“When you are mapping, you have to be very focused and you need to go over your Manual on several occasions. Even then, you always encounter situations when you are not very sure how what [data]

should be coded. For me, that is the one when you map very well and exactly

as the Manual says, so far, which is the tool we all have to follow.”(Practitioner Trainer, 10)

9.2.3. Mapper’s Inter-rater reliability

The quality of observational data is also assessed in terms of its validity and reliability (Patton 2002). Reliability is assessed using the degree of agreement between different observers of observational data collected at a single point from the same place (called inter-observer agreement) or by the same observer at different times in the data collection process (called intra-observer agreement). In the context of DCM, the inter-observer agreement is calculated where two or more mappers conduct mapping on the same people with dementia at the same time and compare the similarities and differences within their coded data (Bradford Dementia Group 2005). They work out the percentage of concordance for BCCs and MEs and a minimum of 70% concordance is recommended for use of DCM for practice development purposes and 80% for research. This is called inter-rater reliability (IRR). It is recommended within the DCM 8 Manual that mappers should check their IRR prior to each mapping cycle in order to produce accurate codes (data) and to ensure that data collected on different mapping participants, by the two mappers, is comparable when combined into a single report for the setting (Bradford Dementia Group 2005), thus retaining the quality of data.

The study participants were aware of the importance of their IRR and most of them mentioned that they try to conduct an IRR often with an experienced mapper or with a colleague.

“We make sure that they [mappers] meet with one, or map with an experienced mapper.”(Practitioner, 7)

The interview data suggested that, in order to estimate the quality of DCM data, IRR is one of the measures that can provide evidence regarding data quality. According to the study participants, if a mapper had achieved IRR against an experienced mapper, it is more likely that her/his DCM data will be more reliable to use than data of a mapper who has not conducted IRR either with someone experienced or at all. However, a recent study by Jones and colleagues (in press), reported that only few practitioners reported they regularly conduct IRR, with many reporting their organisation would not support additional time to the mapping process for them to conduct this ahead of each map. Those who were willing and able to conduct IRR identified that lack of an experienced mapper or in some cases any other mapper within the organisation, with which to conduct IRR was a barrier to achieving this. While a mapper's IRR score is considered an important indicator for quality data within a warehouse, the results of Jones et al.'s (in press) study raises a point of concern regarding the potential availability of DCM data that can be said to have met this quality indicator, given the reported levels of IRR implementation in regular mapping.

9.2.4. Mapper's training and experience

Mapper's training level and experience of using the method were also mentioned by the study participants, as potential indicators of quality data provision. This reinforces Woods and Lintern's (2003) statement regarding experienced mappers enhancing the reliability of generated data. The participants expressed a high level of trust in the generation of quality data by those mappers who were experienced or were qualified to an advanced or

trainer level to use DCM. For example, one study participant mentioned the words 'gold standard' to describe the DCM trainers as experienced mappers.

Researcher 4: *"She [friend] is a DCM trainer but with gold standard. But I am worried that, if I am going to use data from the regular mappers, how the quality of data I am going to look at..."*

SK: *What do you mean by a trainer with "gold standard"?*

Researcher 4: *"I mean she has a lot of experience of mapping and she has trained many mappers so far and she is lead of DCM here in this country."*

One of the study participants who was the lead of DCM within her/his organisation expressed a notion of trust for those who were experienced and considered them as 'reliable mappers'.

"...We assumed that they [experienced mapper] continue to do what they are doing and continue to be reliable."(Practitioner, 2)

However, what makes a mapper experienced is arguable. Participants in this study associated the concept of 'experienced mapper' with mappers' regularity of mapping.

"The fact that all of our mappers have to map regularly because we have got everybody to map means that people are well practised... so we have got very experienced mappers here, who have been mapping for a long time and who have been mapping for longer."(Practitioner, 2)

This finding strengthens Woods and Lintern's (2003) definition of an experienced mapper. They state that a mapper becomes experienced when she/he conducts or uses the method regularly and experiences a variety of situations at the time of mapping.

On another dimension, the concept of being an experienced mapper was also associated with the level of training the mappers had. The higher the level of training a mapper had the more experienced she/he is considered to be. This may relate to the need to map more to gain advanced user or higher levels of DCM certification and qualification (Douglass et al. 2010). Further, as reported by Douglass et al (2010) in their survey study comparing UK and USA mappers' experiences and views of DCM use, the advanced users from both countries were 23 times more likely to map than the basic users after their DCM training, as they have to map in order to achieve advanced certification. Therefore, the advanced mappers could be considered as more experienced than the basic users, thus most likely to be perceived as producing quality DCM data that is trustworthy. On the other hand, a novice or inexperienced mapper is perceived as more likely to make mistakes and is therefore, seen as less trustworthy in providing quality data.

“... [an inexperienced] mapper might be defaulting to using a small group of codes they use all the time. It is that less experienced mapper, who never maps the plus-5 ME values because they don't feel confident about that or they don't feel sure of their way of finding that coding value...” (Practitioner, 2)

The evidence from the interview data shows that an experienced mapper, who maps regularly either for enhancing their training levels or for practice development or research purposes, is most likely to be perceived as producing quality DCM data. As was mentioned in Chapter 8, the information about the mapper's experience and training level was also required to establish the provenance on quality of DCM data for secondary uses. The

availability of provenance information within the warehouse however raises a number of key

issues that need to be considered. They include, whether the information about mapper's experience, training level and IRR score is available for provenance to the DCM data for secondary uses. And whether inexperienced mappers have accessed support from an experienced mapper in order to check data quality.

The first issue requires exploring the availability of mapper related data within the warehouse. However, in relation to the second issue regarding mapper experience and available support, if the mappers (e.g. new and less experienced mappers) are not experienced, interview data highlighted that they often try to seek support from individuals who are experienced and trained at higher levels to oversee their use of the DCM method and this would mean that the generated data was checked for quality.

"It gives confidence if somebody with the highest level of mapping experience is there." (Practitioner, 6)

However, the availability of such support from an experienced mapper is questionable. A recently published survey study by Jones and colleagues (in press) reports that the UK based practitioners recognise the need of a 'buddy' or someone with whom they can conduct their IRR. However, due to the lack of another mapper, Jones and colleagues further reported that half of the total practitioners, who took part in their online survey, never conduct IRR.

9.2.5. Mapper's link with care setting

Some of the interview participants such as the researchers stated their concern regarding the quality of DCM data that is collected by practitioners who use DCM for developing their own personal and organisational care

practice. Interview participants assumed that, as the mappers are familiar with

those they are mapping, there may be a risk of bias with mappers potentially only recording and then feeding back the good aspects or practices within the data.

“I think there is another issue, whether the mapper works within the environment [data-collection site] or not. I think potentially whether the mapper works or not works in the environment is quite a big bias. Well, it might not be a problem for benchmarking, although it might be, as people might rate their own solution better than if they would do blind mapping somewhere else.” (Researcher, 20)

“There’s issues around reliability and validity of data. The problem I have with lots of DCM data collected is that it is collected by practitioners, so they are using DCM within their own care services and I think there is inherent bias in that. So don’t get me wrong, as a researcher, I would like to see the data, but I suppose we would draw conclusions with caution, I suppose. Whereas, if it was collected by somebody outside of the service, then I might consider it. Like, if any trainer from BDG or me or my colleagues are collecting data, then we would not have any motive that could influence the score or anything...if you know what I mean. But whereas, when your own care workers are making their own ratings for themselves and their colleagues... I think that is an inherent bias.”(Researcher, 4)

When the above issue was explored from the practitioners’ perspective, it was evident that wherever possible, they avoid mapping participants within the wards/areas, where they work or where they know the mapping participants very well. However, even when this was not possible, from practitioners’ responses it was evident that they conduct observations carefully and then feed the data back to the staff as accurately as possible.

However, practitioners also acknowledged the benefits of mapping in their own setting, for example, if they know a participant they are mapping, they may

have a better chance of coding that person accurately. On the contrary, if a mapper does not know someone they are observing, it can be a challenge sometimes to decide on a code that reflects their correct behaviour, particularly when that behaviour is complex.

“...It is good in a more fundamental way because, to a point, mapping is subjective. It’s like having your plus and minus bar on your data point. While there are lots of different reasons, if you know the resident, that’s going to change how you do something, ... like giving the right code.”(Practitioner, 5)

This finding raises intriguing questions regarding the researchers’ trust in the data generated within a practice-development context. This issue has been explored in Chapter 8 where the interview data highlighted the researchers’ concerns relating to the reliability of DCM data within the warehouse, a factor that was considered significant for ascertaining the quality of DCM data for secondary uses. This means, therefore, that the data created for practice development requires adherence to quality. The researchers highlighted the fact that, when it is research data, compliance with the quality of that data becomes very important. It might be that research is usually conducted for generalisation purposes and researchers are required to assure themselves and others that there has been a robust quality check in all aspects of the study, from data collection to reporting the findings.

“...Because I think if you are doing it for practice development, it is also very important that you have good quality data. But then it’s not like it matters if PE 6, 7 or 11 and it’s more like that it is PE and it’s not so much about your real scoring or thing, so if you want to do it for research purposes... you expect other things... when you use it for

research purposes. In some organisations, what you score as a PE will not be

scored as a PE in other organisations because it is normal care for them. But if the data will be used for research purposes, this does matter.” (Researcher, 4)

The interview data indicated that the mapper’s role is considered significant in producing quality DCM data. This is the reason why the researchers expressed the need to have provenance information about the mapper – such as IRR score, experience and training level – to estimate the quality of DCM data for secondary research within the warehouse (Chapter 8). Based on these findings, it could be argued that a good reliability score is associated with a mapper’s experience and advanced-level training, or with the mapper having gained help from those who are experienced and trained at an advanced level. This is in line with what Cooke and Chaudhury (2012) found in their literature review, that those studies which report a good IRR score are those where help was gained from BDG trainers (e.g. individuals who provide training to national and international mappers and who are considered champions of DCM). Further, when referring to good quality data, Brooker (2005: 16) uses the term “gold-standard mapper” for the provision of quality data and recommends to ‘formalise’ such a role by providing advanced training and accreditation. In the absence of any measure or criteria for a ‘gold-standard mapper’, it could be argued that such a mapper is one who has a good IRR score, effective experience of mapping and advanced-training certification.

While the mappers’ IRR score is considered significant for establishing the quality of DCM data, it is not conducted consistently across all uses of DCM. While the literature indicates that IRR is conducted for research studies, there is also evidence that it is conducted inconsistently within a practice

development context. Jones and colleagues (in press) provide alarming figures showing that only half of their sampled practitioners were conducting IRR in the UK. This will have significant implications for the provision of quality DCM data, as well as the mapper's IRR information alongside the DCM data for provenance purposes. The lack of information on mappers' IRR scores means a lack of evidence regarding reliable or good quality DCM data storage for secondary use within the warehouse. It is therefore suggested that, before designing and developing a data warehouse, it is important to establish a set of criteria or a tool for measuring data quality and for defining the term 'gold-standard mapper'. The arc|hive DCM database (Surr et al. 2015) has subsequently built a video resource for mappers to check their IRR against various coding scenarios. However, the implementation of the functionality whereby mappers can code the scenarios by watching the video is not yet complete.

In summary, the interview data highlighted that the mapper plays an important role in generating and maintaining the quality of DCM data. They encompass a range of characteristics that can indicate their ability as a mapper for producing quality data. For example, their experience of mapping on a regular basis, training and IRR conducted against another experienced mapper can be data quality indicators. The researchers were concerned with the potential quality of practitioners' DCM data as some assumed that since practitioners' primary aim is to develop their own practice, there might be an element of bias that can influence the quality of their data and they questioned whether it meets the quality standards to be used for research

purposes. While interview data suggested that mappers' IRR score, training and experience levels can

provide evidence of quality, the availability of such information is an important issue for future research.

In addition to these individual factors, the interview data also highlighted the organisational support for mappers as an important factor that can influence the quality of DCM data. This factor is explained next.

9.3. Organisation's role

The interview data highlighted that the organisations where DCM is used also play an important role in providing support for implementing it and maintaining the quality of its generated data. First, I will present the organisation's role in implementing DCM, as perceived by the study participants. Then, I will explore the techniques that organisations employed to support mappers to use DCM and improve the quality of data.

9.3.1. Organisational/management support in implementing DCM

The interview data highlighted that the organisation's management support in implementing DCM can play a key role in enhancing its use and consequently the data production. The study participants such as the practitioners (n=5) from two out of four organisations stated that they mapped regularly and saw the use of DCM as an important part of the organisations' regular care-monitoring activities. Within these organisations, the role of management was strong and supportive in embedding the use of DCM across all units. In these situations, the use of DCM was on a routine basis, as mentioned by one interviewee who was in a management role.

"DCM is used practically... I would say almost on a daily basis. The current structure is that, every month, we have a DCM day that is set

aside for mappers to do mapping and also, on that day, they would meet

and reflect on practice and talk about any concerns and plans as well as about mapping in terms of a home or a particular unit. We are currently working with the Bradford team regarding culture change in the organisation and they are actually using the mapping to set benchmarks for practice and they also use the maps to evaluate the culture change and the progress as well. So it is commonly used.”

(Practitioner, 7)

In both organisations, where DCM was used on a regular basis, there were individuals from management positions who were themselves trained as mappers. They were enthusiastic about the regular use of DCM within their organisations and therefore were keen to provide support for the mappers to conduct DCM regularly (approaches for supporting mappers are presented in Section 9.3.2 of this Chapter).

Contrary to the above, the interview data identified the lack of management support as the main factor influencing the use of DCM within one organisation. This finding supports what Douglass and colleagues (2010) found in their study to gather experiences of mappers from the US and UK. They reported that the major reason for mappers not mapping was the lack of support from their organisational management. Similar results were also reported by a recent multi-method study (Jones et al. in press), where the authors found that 80% of those who mapped within the UK mentioned the role of organisational support as a major driver for using DCM.

One practitioner highlighted the importance of members of the management team attending DCM training. According to the participant, if management has someone within it who is DCM trained they will be able to understand the

importance of its use and implementation. However, if someone is not DCM trained at management level, it indicates the unsupportive management.

“Basically, over here [in the organisation], it’s just been so difficult to gain support really from the management on the wards. We do the dementia-care mapping course and we would actually specify very clearly that, even at a very high managerial level, actually, all the ward managers need to go on that. Because we identify that, really, unless you have got the ward manager on board, it’s really difficult to do anything. ...Where we really struggled were the two places where the two ward managers didn’t go on the course even though they were invited. ...It’s been a real obstacle and there’s mappers, particularly over here, where I have gone over and arranged a support for them, for some of the briefing, some of the relatives and something like this. And actually they just haven’t had the support from the manager.”
(Practitioner Trainer, 14)

The interest from the management also resulted in supporting mappers to map on a regular basis by providing them with time and resources. However, with weak management support, the participants saw taking the time out to map as a huge challenge alongside the demands of their normal job role and duties.

Some interview participants mentioned that they were struggling to conduct maps on a regular basis as they found it difficult to embed the use of DCM within their busy working hours. This finding seems to be consistent with Douglass and colleagues’ (2010) study where they report a lack of time for DCM as a major issue expressed by mappers from both the US and the UK. One of their assertions is that it could influence the use of DCM, thus impacting on the generation of data for secondary uses.

“Time-wise, there is no protected time to map. It is something that we have to make happen.” (Practitioner, 3)

In response to the question about why were mappers not mapping on a regular basis in spite of their desire so to do, another participant answered:

***Practitioner, 2:** “I think we have enough mappers across the trust. There are a few more mappers in medicine than in surgery. I think, at the moment, we are constrained by our own roles and daily work.”*

An additional factor influencing the mapping frequency was the professional job role of the mappers. Within the organisations, mappers take specific time out for mapping, as they have to embed mapping within their usual roles.

“As all are busy, so it seems like a secondary job rather than embedding it into the unit.” (Practitioner, 5)

“It is quite challenging to incorporate DCM with my role... my role is very wide. I am involved in direct clinical work, I am involved in families and I am involved in quite a lot of research that happens, so DCM is only one part of what I do. I find that I have to work quite hard to make sure that I hold on to it.”(Practitioner, 3)

This finding reflects those of Jones and colleagues (in press) regarding the lack of resources, including time, for conducting DCM within organisations. By realising that one of the biggest challenges is to get time for mapping, one mapper, who was the organisation manager, reported that they had set days for mappers to map within their organisation.

“So it’s a day where it is protected for the mappers to carry out mapping, because... one of the challenges is getting the time to actually map. So these DCM days protect the mapper, allow the time to map.” (Practitioner, 5)

Whilst advice is provided in the DCM Manual (Bradford Dementia Group 2005) and British Standard (BSI 2010) on how DCM can be effectively

implemented within the organisations by providing resources to the mappers,
the study data

highlighted that the mappers were still struggling to find time to map. This generates potential data availability issues for a future data warehouse.

While it was identified in the interviews that organisational support to mappers plays an important role in implementing DCM, the next sub-category explores how organisations were supporting mappers to implement DCM.

9.3.2. Organisational support for mappers

Organisations which mentioned their support for mappers for implementing DCM were taking various approaches to give that support in mapping and generating quality DCM data. This sub-category will explore these approaches taken.

9.3.2.1. Further training

As highlighted earlier, according to the study participants, the mappers who had training at advanced level or above were considered most likely to produce reliable and thus quality data. The two organisations, which were actively supporting the use of DCM realised the need to arrange basic training for more individuals to become DCM trained and encourage the existing basic mappers to enhance their training to an advanced level.

“We encourage new staff to go for training and existing mappers to go for advanced training.” (Practitioner, 5)

“We have got very experienced mappers here who have been mapping for a long time and who have been mapping for longer than I have. But what we hopefully benefit from is the further training.” (Practitioner, 2)

The aim of these organisations was to enable more staff to have DCM training, upgrade the existing mappers' training level to an advanced level and ensure that all mappers collect quality DCM data.

9.3.2.2. Peer support

The interview data highlighted that the mappers who were supported by their peers were more confident in using DCM and therefore were most likely to produce data that was perceived to be of good quality. It was also identified that the supportive organisations were encouraging a culture of peer support for their mappers. Their main reason in doing this was to support mappers in using DCM in an accurate way. Participants highlighted how, in their organisations, the mappers were meeting as a group and were encouraged to support each other to share their mapping experience and to examine queries or questions and so learn from each other. One of the participants, who was from their organisation's management team, mentioned the culture of peer support within their organisation.

“We make sure that we have a team of mappers, so that they can work with each other and support each other.” (Practitioner, 5)

Peer support was considered an important step towards producing quality DCM data, as reflected in other study participants' interview data.

“One benefit of having such a small number of mappers is that you can manage quality much better. Everybody is trained and we meet as a group.” (Practitioner, 2)

“In order to ensure the quality of data, we make sure that they [mappers] meet with one another or map with another experienced mapper.” (Practitioner, 7)

This finding reflects what Jones and colleagues (in press) concluded in their survey-based study that peer support, or a 'buddy system, may enhance the chances of mappers conducting IRR on regular basis, thus improving the quality of DCM data.

9.3.2.3. Regular mapping

It was identified in the interview data that the study participants perceived experienced mappers as most likely to produce good quality DCM data. As mentioned previously, a mapper was considered experienced if she/he mapped on a regular basis. Therefore, it was identified from the interview data that in order to maintain the quality of DCM data, organisations were also trying to ensure that mappers were well practiced by encouraging them to map on a regular basis. For example, the management in one care setting were strongly in favour of mappers to map every month.

“One of the things that we did after the training is that... each mapper had to do mapping for a certain period of time.” (Practitioner, 7)

SK: *What about the quality of data? How do you make sure within the organisation that the data you capture is of good quality?*

Practitioner, 5: *“Within this organisation it is really about our monthly meetings”*

“The fact is that all of our mappers have to map regularly. Because we have got everybody to map means that people are well practiced, which means they will produce good quality data.” (Practitioner, 2)

9.3.2.4. Help from experts

In order to provide support to the mappers, organisations also gained help from experts within the DCM field. The aim was to make sure that mappers were adhering to DCM procedures correctly.

“We also had DCM experts from Bradford, mapping with individual mappers and comparing information just to make sure that they fully understand and actually follow the procedure and the outcomes and everything.”(Practitioner, 7)

The mappers within organisations, who had been using DCM for a long time or had an advanced level of DCM training, were also considered as experts and a useful supervisory support for the novice and less experienced mappers in their mapping activities.

“The advanced mapper at intervals maps with other mappers and talks with them about the results and the exercise and everything else and just makes sure that they are coaching and continually developing.” (Practitioner, 7)

9.3.2.5. Mapping in pairs

Another approach for managing the quality of data was to encourage mappers to map in pairs for conducting IRR between them. The main aim was to make sure that they capture the DCM data in a more unified way.

“In order to enhance the reliability of the data, we try whenever possible to map in pairs anyway.”(Practitioner, 2)

“...That the mappers do the sessions where they compare the information that they had collected and about the same resident within the same time-frame. So that it is a kind of measure that are we getting, a completely different picture.”(Practitioner, 7)

Alongside enhancing the implementation of DCM, the organisational support is also paramount in enhancing the quality of DCM data. The mappers from the organisations who supported DCM implementation explained various approaches that they were using to ensure the quality of data. These approaches were related to ensuring basic and further training for staff to use DCM, peer support; regular mapping, mapping in pairs and seeking support from experts to help in implementing DCM within their organisation.

The above data reflect the organisations' views that are supportive to DCM and its implementation. While this study has limitations, in not capturing the view of those organisations where DCM is not conducted or is conducted irregularly, the evidence from Jones et al.'s (in press) study suggests that not all mappers within the UK organisations where DCM training has been delivered are mapping regularly. One of the reported reasons in their study is the lack of support from organisations. While there is evidence that mappers need organisational support to conduct DCM following their DCM training, further research is required to establish how organisations can be attracted to the use of DCM data and consequently provide support to mappers to conduct DCM on a regular basis. The regular use of DCM can enhance the chances for generating large amount of DCM data for the potential secondary uses. This is in line with suggestions made by others. For example, Douglass and colleagues (2008), in a qualitative study, suggest that the use of more widespread training and organisational support might enhance the use of DCM in the US. To some extent, Jones and colleagues (in press) have made similar suggestions for the UK. However, the research into how this can be achieved is yet lacking.

In summary, the evidence presented in the above sub-category shows that management within an organisation can play an important role in providing resources such as time for mappers to map on a regularly basis, thus generating DCM data on regular basis. This can ensure the availability of DCM data for a data warehouse. However, lack of management support and irregular mapping can influence the amount of specific data generated and its availability within the warehouse for secondary uses, thus referring to the

issues related to the availability and quality of data that is fit for secondary purpose. Further research is required to establish how organisations can be attracted to use DCM on a regular basis so as to generate data for potential secondary uses.

9.4. Primary DCM data management system; limitations and requirements

A data warehouse stores data taken from various data sources and therefore its success and workability depend on the availability of the required data (by users) within the data sources. Data sources for a data warehouse collect and store data for primary uses, primary data management systems. The main issue reported within the literature is the inability and incompatibility of the primary data management systems to capture complete, accurate, consistent and thus quality data in an electronic format to enable data sharing and transferability for secondary uses (Batini et al. 2009; Health and Social Care Information Centre 2012).

The interview data from this study highlighted that the mappers were not satisfied with the system they were using for managing DCM data for primary purposes in terms of a complete, consistent and integrated data storage solution. As was argued in Chapter 3 (data quality section), this can create data quality issues in terms of its availability and compatibility for secondary uses. According to the study participants, the lack of an effective data management system was causing huge implications for their current use of DCM data. This section will explore the study participants' views regarding their current systems for managing primary DCM data, both in terms of its

limitations and their requirements for an effective system. This exploration will

provide insights into the issues that can have a significant impact on the availability and quality of DCM data.

At the time of data collection for this study (2012-2014), the Excel programme was the only tool provided (on request) by the University of Bradford to the mappers for managing their primary DCM data for storing and analysis purposes. Therefore, the majority of the study participants (n=27) reported the use of the Excel programme for managing their data. Nearly all of those study participants who were using the Excel system mentioned its limitations as a data storage and analysis solution. Two participants however appeared to be satisfied with functionalities provided by this programme for managing their primary DCM data. They were conducting only very basic analysis techniques, which may have been the reason why they did not envisage any further uses for their data and therefore, could not identify any limitations in their existing systems.

“It [the Excel programme] meets the needs, because, when you produce graphs, you obviously need to interpret in the report itself. Yes, I don’t have any problems with it. It is very straightforward, very easy to use and it produces graphs that I understand.” (Practitioner, 23)

The interview data revealed that nearly all the participants found the Excel programme insufficiently competent to deal with their data storage and analysis needs. This finding is in accord with previous studies indicating limitations of the currently used Excel programme and the need for an effective data-management system for primary uses of DCM data (Khalid 2009; Khalid 2010; Khalid et al. 2010; Jones et al. 2014).

The researchers' interview data indicated that the main issue was the limitations of the Excel programme for dealing with data-analysis activity. Therefore, they preferred to use other analytical tools. One of the researchers highlighted how she/he imported the data from the Excel programme into SPSS to permit further data manipulation and complex analysis.

“Today, I got an Excel from A and her colleagues and I have now converted it into SPSS to be able to analyse the data. I think the Excel programme is good for improving the quality of care from the practice side. It is OK, but for the research side it is a little bit too shallow maybe and I always like to have converted my data into SPSS because it is very nice to work with. For my research purpose... I really have to rearrange the data and the Excel is not very helpful, so there is a lot of work to do.” (Researcher, 4)

Some researchers, however, preferred to transfer their data directly from DCM raw data sheets into SPSS for their desired analysis.

“We did not use the Excel tool; we only used SPSS for our data analysis.” (Researcher, 20)

In order to deal with their particular needs, some researchers and practitioners also mentioned that they were developing their own data management solutions for complete and integrated data storage. As mentioned by one researcher, they developed their own database to fulfil their storage and analysis requirements for data collected in their study.

“We created our own Access databases to store and analyse these [DCM] data in a statistical package. It works for our study because it was designed for our study.” (Researcher Practitioner, 8)

Some study participants such as the practitioners reported that within their organisations they were interested in having an integrated view of the DCM

data in one place for the purposes of analysis and sharing with other in-house mappers, with the aim of identifying good practices. In the absence of any effective system, mappers from one organisation mentioned strategies they had in place for storing data and then sharing it with all those concerned, mostly individuals at a managerial level. For example, currently all mappers within their organisation store DCM data within a central system in a secure format. This system was just for storage and did not allow the mappers to share their data with other mappers and it did not allow any analysis at an organisation level. In order to see trends and patterns within the data they had set organisational DCM days when all mappers get together, discuss and share their mapping data.

“We created like an intranet, a system for all the mappers to store their data centrally. You know, it’s a software system and we also have hard copies and each mapper has their own copies of mappings and other things they have done. The central system on the computer that we created in terms of practising, evaluation and reflection. So that if somebody wants to see what other mappers have done and if they want to see any trends that have been around. It’s not something I think at the moment that all the mappers are accessing or anything. I think mainly, when mapping is done, people keep their own data or information and then discuss it through DCM days, which are actually held on a monthly basis, so that’s how they are sharing information at the moment and also when they do the feedback, it is a direct feedback and the data they keep themselves at the moment.”

(Practitioner, 7)

Further, DCM trainers were also interested in viewing the mapping data that they had collected over time to see trends and patterns. This was not possible with the Excel programme. One trainer mentioned that various

mapping sheets could be amalgamated to view the mapping conducted over time. However,

using the Excel programme, this process was not easy and the analysis techniques meant it was not possible to examine change for a specific mapping participant within a specific unit over time. Therefore, they indicated, there was a lack of ability to conduct multidimensional analysis of the data, since the Excel programme did not support such analyses.

Currently, the Excel programme only allows the input of some of the numerical DCM data (BCC and ME) and does not allow the storing of Personal Enhancers (PE) and Personal Detractors (PD) or qualitative mapping notes. In order to make the DCM data useful and meaningful, the participants mentioned the need to view all DCM data items within one single system for storage and further analysis. The participants saw great value in being able to link PDs and PEs with other DCM data items within the analysis process.

“This system [Excel] does not capture the person enhancers and personal detractors either and I think that, in itself, what staff are really interested to hear is that what they have done has made the difference. So we capture that sort of elsewhere and feed that back slightly differently, so yes, that’s another limitation really.”
(Practitioner, 3)

“There is no way to record the PD and PEs and it would be nice to be able to have the option to do that and also on the tracker graphs for us to be able to identify PD and PEs, so that we could put all the four codes together. That would be really interesting.” **(Trainer, 19)**

“The Excel programme does not do anything with PD and PE and staff interactions but together they would formulate a full report and we would then look at trends that people were in over time, what activities are people engaging in overtime and the quality of the staff

interactions that are helping people with dementia in the facility.”
(Trainer, 1)

Therefore, the participants expressed their requirement to have access to the integrated data (all DCM data items in one place in one system) that they collect over time in order to compare them from various dimensions (individual, unit, organisation and time).

“I find we are often having to work like part of the data that we import from elsewhere to answer some of these questions. It would be lovely if it is all in a package.” (Practitioner, 3)

The participants also mentioned the importance of qualitative notes, which are part of the DCM data and which give them the opportunity to conduct more in-depth analysis of the numerical data. The current Excel programme does not provide the facility to add notes within the system. The participants stated that they would like to be able to use these notes alongside DCM data to answer many other questions that they cannot answer at present. One participant mentioned that they want to use more data from the notes, as during mapping a lot of information is not coded but is noted down in qualitative form. However, currently there is no way of dealing with this data.

“The data that you can’t use is the data that is interesting. For example, you are in a setting and somebody is slipping down from U-3 and you know where they are heading and you stop mapping but you know, in the next 10 time-frames, it will be U-5 and that is where you have been intervening and trying to make a difference, so that comes out in qualitative notes...” (Trainer, 18)

“I think the possibility to take notes is lacking in the system that I use. And if there was a possibility to put notes and maybe possibly treat them as qualitative data ... analysing them afterwards, I would like that.” (Researcher Trainer, 11)

“You can’t do anything else; there is no option for writing your qualitative notes.” (Researcher Practitioner Trainer, 16)

“We could also use some more data than just the dementia care mapping, as the noise levels, we had the number of staff on the bay so we developed our own one to include everything. I think the problem for us with the Bradford Dementia Group spreadsheet was that it did not record everything. It’s easy just to have everything at one place.” (Researcher Practitioner, 8)

The study participants also mentioned the requirement of comparing DCM data overtime to see trends and patterns within the data. According to the practitioners, the Excel system is not compatible with the ever more complex types of analysis they want to perform on the DCM data. These needs were mostly related to the analysis of the longitudinal DCM data they collect.

“If you have data to look at comparatively over a six-month period and then see if there is any change of ME values or BCC codes and in the PD and PE numbers, then that would be really useful just to give you the numbers and give you the general picture of the home.” (Practitioner, 6)

Most of the participants (including both practitioners and trainers) showed a strong interest in comparing their mapping data over time to track the same samples at different points in time. The sample might consist of individual patients, or a group of patients who take part in the mapping, and a particular unit or organisation where mapping is conducted over a period of time. This refers to the multidimensional analysis of the data for internal benchmarking purposes. Chapter 7 has discussed the use of DCM data for internal benchmarking.

“We have been doing global [organisational] analysis of all the information at once. We will sit down and go through the feedback... actually we created groups and each group would discuss their mapping sessions and the information that they have collected and also discuss that with the rest of the team. And during those times they will be picking up if there is any training need or any particular thing that is happening in units and then they are using that information either for feedback with the manager to talk about the developments or any other concerns and so on. So that is how we have been using the data. So it has been actively used but we have not had a system where we collect the data all at once and look at it globally [organisationally].” (Practitioner, 7)

“If you could compare the ranges of behaviour category codes... like in some particular time what were the percentages of these codes as high medium and low? And comparing the ME values and also PDs and PEs and directly link them to the psychological needs, I think it would be good to compare those. I don't think I have those skills to do that at the moment.” (Trainer, 19)

It was also identified from the interview data that participants felt they were being restricted in their use of DCM data due to the unavailability of an effective electronic data management system, and they attached great importance to such a system in order to explore the use of DCM from various angles. For example, one of the participants mentioned that an effective system would change the way they currently use DCM and would make its use more meaningful.

“Once we have got our IT team to do it [developing a data-management system for DCM data], which is not everybody putting one map into one document as is done in Excel but keeping it per patient and pulling out the data what we need for whatever purposes.

Then, I think the way that we use DCM will change really quite dramatically and it will be far

more meaningful and far more joined up and we will be able to direct service development, staff training, all of those things, much, much better.” (Practitioner, 2)

The above data indicate that the study participants are currently unable to collect all the DCM data (both coding data and qualitative notes) in one system for the purposes of integrated analysis. Further, the lack of a system to support historic analysis of DCM data to assess care improvement is also highlighted. A number of previous studies also reported this issue (Khalid 2009; Khalid 2010; Khalid et al 2010; Jones et al. 2014). As a consequence of which, the University of Bradford developed and launched a web-based database system (the arc|hive DCM database) (Surr et al. 2015), with enhanced functionalities for complete DCM data storage. This addresses some of the study participants’ expressed requirements for their primary DCM data and its use. For example, the new system is web-based and allows the entry and storage of all DCM data, such as coding data and notes, in an integrated format. However, the system does yet not provide the individual mappers and organisations with the facility of data linkage, such as linking maps collected over a period of time, and its multidimensional analysis, both of which are major requirements related to the primary DCM data-management system identified above. This system is only available in the UK because of varying data-protection and ethical requirements regarding the collection of international data. Further, the system has not been evaluated in terms of its usage and user acceptance since its launch in June 2015. Therefore, it is not evident how many people are using the database application, nor whether it is providing an improved data-storage and analysis solution for the mappers.

As was mentioned in Chapter 3, data quality is a key issue in data warehousing (Verma et al. 2014). Chapter 3 also argued that data quality within a warehouse is associated with technical processes of the system, the intrinsic nature of the data and secondary uses of data. During technical processes, data quality issues can emerge at various levels including data sources and integration (Singh and Singh 2010; Mohammed and Talab 2014). The study data highlighted that DCM data quality issues can emerge at both data source and data integration level. The study findings show that the current data management system such as the Excel programme is not collecting complete and integrated DCM data and therefore cannot be characterised as a quality data source for a data warehouse. Further, there is a lack of evidence to show that the archive DCM database is in a regular use, thus providing quality DCM data for potential secondary uses. Identifying good quality data sources for data warehouses is emphasised in the literature (Singh and Singh 2010). However, the study data indicate that the current situation of DCM data management and storage for primary use is yet not up to the level where mappers are content with the system and it is being used regularly to collect DCM data. This indicates the data incompatibility and availability as quality issues for the future data warehouse. These findings are in accord with other studies, where the issues of data availability and incompatibility are highlighted for this purpose, while developing a data warehouse solution for managing other type of healthcare data for secondary uses (Sandra and Garmon 2007; Botsis et al. 2010).

The study data also indicated that some mappers were developing their own systems for managing DCM data for primary uses. Jones and colleagues'

(2014) survey study with UK based practitioners also highlighted that at present some organisations are using other systems such as RIO (Servelec 2016) and SystemOne (2016) to record DCM data. While DCM or relevant data can be obtained from other data sources for the warehouse, due to diverse formats and locations of these data sources it can create quality issues at a data integration level (Mohammed and Talab 2014). This requires further research into developing effective tools and solutions for DCM data integration to create a unified view within a data warehouse to facilitate secondary analysis.

In summary, nearly all the study participants were using the Excel system and most of them mentioned its limitations as a data storage and analytical tool for their increasing requirements for dealing with the DCM data. The researchers were mostly interested in a data analysis tool that could conduct complex statistical analyses. The practitioners were interested in a system that provides them both a facility to provide an integrated view of DCM data they collected over a period of time about one patient and a facility to analyse this data from various dimensions. These requirements might have emerged for different reasons but act as conditions for escalating the need for an effective data management system to fulfil mappers' individual and organisational requirements to deal with the primary DCM data in a meaningful way.

The requirement for an effective data storage system was evident. This would need to provide an electronic storage of data that captures all the DCM data attributes and the relevant data collected during each map, stored in one format. The current system stores incomplete DCM data that is

collected during a single mapping. This raises data availability and quality issues in

terms of its completeness and sufficiency for specific purposes. The requirement for a primary data management system for mappers also becomes a requirement for the data warehouse as the acceptability and usability of a data warehouse depends on the availability of the DCM data. An effective system that stores primary data in an electronic format that is standardised will make the DCM data sharable and usable, and thus available for the secondary use within a data warehouse.

9.5. Key findings and contribution to knowledge

While the roles of mappers and organisations, and the limitations of the existing data-management system for managing DCM data, have been identified as separate issues in the literature, this is the only study that has brought these together at a conceptual level to discuss their implications for the quality and availability of DCM data for secondary uses.

Data quality and availability are two important issues in data warehouses. They are also different but linked issues. For example, Wang and Strong (1996) assert that, if user-required data is not available within a data warehouse for a specific use they identify as important, this is associated by users with data incompleteness, which is considered a data quality dimension. Thus, this study has indicated that data availability issues with a future data warehouse may also influence whether users consider data that is within the warehouse to be of quality. The data presented in this chapter indicate that both mappers and organisations can influence the actual and perceived quality of DCM data within a future warehouse. The study has indicated that if organisations provide better support for mappers through

encouraging conduct of IRR maps and regular use of DCM, this will support generation of

trustworthy data in consistent quantities. However, currently there remains a lack of evidence about how many, and how often, individual mappers and organisations are using DCM within and outside the UK and ways in which regular mapping activity within individual organisations can be supported. This study indicates that addressing this in future research will be an essential component of building a successful data warehouse.

Further, there is also a lack of systematic research to find out how mappers are managing their DCM data and in what formats. The University of Bradford's vision of replacing the existing Excel system with a web-based database also reflects their efforts to manage the use of DCM, its data, the number of mappers and their regularity of mapping. While such a system can technically support data management, there is no evidence that individuals or organisations will adopt this system. While studies have shown that mappers welcome such a web-based system (Jones et al. 2014), it has also been pointed out in general literature that adoption of a technical system at individual and organisational level is a huge challenge. The most reported organisational factors as barriers to implement IT within care settings are planning, project management, training, technology support, turnover rate, clinical workload, and communication (Lluch 2011; Cresswell and Sheikh 2013; Yusof 2015). In specific to social care settings, iCareHealth (2014), one of major technology support providers to the social care settings within the UK, in their recent report reveals three major issues/barriers to implementing technology within social care: budget constraints, lack of internal resources and resistant to technology. Considering social care

settings as the hub of various services with staff from various skills and knowledge, the

implementation of IT systems can be challenging and therefore require in-depth understanding.

Further, many studies also report individual factors that can influence IT implementation within care settings. Individual factors are related to individual's own perceptions, values and attitude towards the use of IT systems (Nieboer, 2014). For example, a study reports nurses' perception and attitude to using and adopting IT within their practice (Hung et al. 2014). They conclude that if nurses find a system trustworthy and useful, their attitude will change towards the technology. In agreement with this, Nieboer and colleagues (2014) assert that the perceptions and values of care professionals are critical success factors in successfully implementing technology in health care. In the context of DCM, systematic research is required to investigate the core issues in implementing technical systems for managing DCM data within organisations and in further assessing mappers' willingness to use this system.

9.6. Summary of the chapter

The chapter examined three main factors that can potentially influence the quality and availability of DCM data for secondary use. These were: the mapper who collects DCM data; the organisations where DCM is conducted; and a data resource, which collects and stores data for primary purposes. The mappers' role was considered significant in providing reliable data that would be considered quality data for secondary use. It was suggested that the mappers' IRR scores, mapping experience and training at advanced level could provide indicators of the quality of DCM data. The interview data also suggested that organisational support plays an important role in providing

support to mappers in producing quality DCM data. The supporting organisations provided the time to map, the opportunity to map on a regular basis and also offered help in the form of the expertise of experienced mappers. It was argued that organisations' support for implementing DCM and the support to mappers using it on a regular basis can provide the chance for DCM data availability to facilitate secondary uses within the future warehouse.

The third factor contributing to the availability and quality of DCM data was found to be the data-management system for primary use of DCM data. The study participants' lack of satisfaction with the existing system and their need for a new system, with improved functionality of integrated data storage and multidimensional analysis, were explored. It was argued that an effective system storing primary DCM data in an electronic standardised format would make the data sharable, usable, and thus available for secondary use within a future data warehouse.

10. Future Work, Limitations and Conclusions

10.1. Introduction

The previous chapters (7, 8 and 9) presented and discussed study findings, highlighted the original contributions to knowledge proposed in the thesis and underlined the need for potential future work. In pursuit of concluding the study, this chapter first provides a summary of the key findings based on the objectives of the study. It then compares a user-driven approach with a data-driven approach in the context of the study findings. The chapter then goes on to reflect on the suitability of the user-driven approach adopted within this study. This includes providing a brief reflection on the use of modified grounded theory as a user-driven approach. Further, this chapter provides a summary of the recommendations for future practical work and research for designing a future data warehouse for DCM. It ends by discussing the limitations of this study and reaching a conclusion.

10.2. Summary of the findings based on the study objectives

This study aimed to explore requirements for the secondary use of DCM data using a user-driven approach. The scope of the study was established by setting three main objectives following a detailed background literature review presented in chapters 2, 3 and 4. Using modified grounded theory techniques, this study has met its set objectives. Chapter 7 has explored the potential secondary uses of DCM data to develop an argument that there could be three potential uses of DCM data within a future data warehouse, thus meeting the first objective of this study, which was to identify the potential secondary uses of DCM data within a future data warehouse.

The second objective of this study was to identify the type of data that needs to go in the warehouse to meet the users' identified potential use of DCM data. While Chapter 8 focused on exploring the information need within the context of using DCM data for research purposes, Chapter 7 argued the need for additional data collection within the context of benchmarking and mappers' training and support. Based on the evidence from both chapters 7 and 8, it was then argued that additional data needed to be stored within the warehouse to meet the users' specific secondary uses. The main findings emerged regarding the need for storing data that can provide provenance as well as keyword and ethical information within the warehouse. The need for storing contextual data, such as mapping participants and care settings' characteristics, and mapping details, such as 'purpose of mapping' was also identified. While the second objective was met by identifying the user-information requirements, Chapter 8 also discussed the data availability issue, as currently there is a lack of evidence of such additional data collection as part of the mapping process or of its recording within existing data-management systems available for primary uses of DCM data.

The third objective of this study was to identify issues and concerns, if any, related to the secondary use of DCM data. While both practitioners and trainers highlighted their potential secondary uses of data, they did not express any related concerns or issues. Researchers, on the other hand, identified their concerns regarding the quality and security of DCM data within the warehouse (Chapter 8). A general issue, however, highlighted by all study participants, concerned the lack of an effective data-management system for

the primary uses of DCM data. This issue can potentially influence the quality and availability of DCM data for secondary uses (Chapter 9).

10.3. User-driven approach vs. data-driven approach

The previous study (Khalid 2010) which I undertook adopted a data-driven approach and proposed a technically workable data model for the warehouse. Jukic and colleagues (2010: 381) name the data-driven approach as the “shrug the shoulders approach”, which means that data-warehouse designers and developers, in the absence of user identified requirements, only focus on the technicality of data integration. In my previous study (Khalid 2010), however, this approach was concerned with demonstrating the data flow between data models, as explained in Chapter 3. Jukic et al. (2010: 381) further assert that a data-driven approach originates from the assumption that, if a system is built that is technically sound, “its users will become apparent”. However, a technical working system also fails due to its unacceptability by the users (Winter & Strauch, 2003). One way of making a system user-acceptable is by involving users using an effective user-driven approach and identifying their views as requirements for the new system, the aim of this study. This study employed a user-driven approach and identified a number of requirements, which were not apparent previously during the use of a data-driven approach in my previous study (Khalid 2010) and are crucial to designing and developing a user required data warehouse.

The findings of this study, presented in Chapter 8, have shown that study participants, such as the researchers, expressed their concerns regarding

the quality and security of data for research purposes. These concerns consequently led to the identification of the information requirements that will

inform the users regarding the quality and security of data within a future data warehouse. These issues and concerns were not visible during the data-driven approach employed in my previous study (Khalid 2010) and therefore the need for storing additional data within the warehouse was also imperceptible.

Further, as argued in Chapter 4 that users' information requirements are important in understanding the type of data that has to be collected, integrated and stored within the warehouse for secondary use (Kujala et al. 2003). It is also asserted that user requirements illuminate the aspects that influence every part of system development, from the system's design to its development to its implementation and finally to its user-acceptance and user-satisfaction (Winter & Strauch 2003). The findings of this study show that users' information requirements have also led to identifying the issues that can potentially impact on design, development and successful working of a data warehouse. One such an issue was identified as data availability for the secondary use of DCM data within a future data warehouse. For example, the requirement for additional data, such as mapping participants' characteristics, care-setting's characteristics, provenance information and mapping details, implies collecting and storing this data within the warehouse to meet user requirements. However, currently such data is not collected as part of the mapping process. Further, there is also a lack of any functionality within the existing data-management systems to record, store and link such additional information alongside DCM data to meet requirements for secondary uses. The requirement for additional data alongside DCM data, and the lack of any processes and systems for collecting such information,

underline the issues of data availability for a future data warehouse. These issues need careful

consideration for designing and developing a working and user acceptable DCM data warehouse.

In summary, this study demonstrates that a user-driven approach for DCM has identified a number of novel requirements which were not evident during the use of a data-driven approach in the previous study (Khalid 2010) and which are significant for the potential secondary uses of DCM data.

10.4. User-driven approach; reflections on the use of modified grounded theory

This study has demonstrated that the use of an effective methodology can help in eliciting the user views and their interpretations as requirements for a new system. User-driven approaches are encouraged for system analysis and design (Raab 1998; Kujala et al. 2001; Teixeira et al. 2012). However, as argued in Chapter 4, users are often criticised for not being aware of their requirements, having communication issues, not being aware of the technical demands of the system and therefore providing requirements that cannot be translated into workable systems (Christel and Kang 1992; Kujala et al. 2001; Abai et al. 2013). This might be the reason why user requirements are ignored or poorly defined for system designing (Kujala et al. 2001). Considering the potential users' non-technical background and limited knowledge of the data-warehouse systems, techniques from grounded theory were used in this study.

The use of grounded theory guidelines has been successful in part in terms of identifying the requirements of users who had a limited understanding of a data warehouse. The inductive approach helped me to understand what was important to the users in terms of their issues, concerns and requirements

rather than having to ask them questions based on an existing conceptual framework from the literature. Further, the interpretive nature of this study helped me to extrapolate requirements that were not directly expressed by the study participants but which repeatedly emerged from their interview data as main issues of concern (Chapter 9) and were considered as having a potential to influence the secondary use of DCM data. For example, the limitations of the current primary DCM data-management system were of concern to all study participants. The main reason for this was that they wanted to explore the use of DCM data and to see the value of multidimensional analysis of longitudinal data. While the issue was related to the need of a data management system for primary uses, it was imperative to explore it as the literature highlights that incompatibility of such systems can influence the availability of data for secondary uses (Wang and Strong 1996). In this sense, this category was contributing to requirements for the secondary use of DCM data at conceptual level.

The feasibility of looking at the literature during the data collection and analysis (theoretical sensitivity in grounded theory) enabled me to see if the issues or concerns that emerged would have any impact on the secondary use of data within a data warehouse. For example, the researchers' need to access additional data alongside the DCM data was classified as two types of data requirements for a future data warehouse, content data and metadata. Furthermore, mapper's and organisation's roles were recognised important for data provision and data quality for the secondary use of DCM data.

This study, being of an exploratory and interpretive nature, raises a number of opportunities for future research, both in terms of requirement validation and

of developing an in-depth understanding of the issues identified. More research will, in fact, be necessary to refine and further elaborate the exploratory findings of this study. The recommendations for some of the future work are provided next.

10.5. Recommended future work

While this study has initiated novel work in the field by identifying mappers' views regarding the secondary use of DCM data, further research is required to clarify, explain and validate these views in more detail so as to gather focused requirements concerning each specific secondary use (identified through this study) of DCM data. This would require recruiting a large sample from each user category (researchers, practitioners and trainers) and designing a specific set of questions to illuminate further the topics of concerns that emerged in this study. The findings of this study have implications for further research and for practical work for designing and developing a future data warehouse. Where possible, the implications of future work were highlighted in the previous three chapters, along with the emerged findings. Next, however, I will summarise the main areas for future work that are significant for a future DCM data warehouse.

10.5.1. Identifying, collecting and linking additional data

This study has identified the need for additional data collection within the warehouse to meet users' need for all three identified potential secondary uses of DCM data (Chapter 7). Further research, however, is required to investigate how the required additional data can be made available, in terms of its identification, collection and linkage with DCM data within the

warehouse, to meet the user requirements. This has both governance and technical

implications for a future data warehouse. For example, data-warehouse providers will need to ensure that policies and procedures are in place for identifying, collecting and storing additional data within the warehouse. Further, based on users' new data requirements, there is a need to modify the existing data model for the warehouse (Khalid 2010) and to build applications to facilitate user access to the data for particular uses.

10.5.2. Data quality

As stated in Chapter 3, data quality is a major issue in warehousing healthcare data. DCM data need to be of good quality to be part of a future data warehouse for secondary uses. DCM data quality is an under-researched area. While, for the primary use of data, the literature highlights that data quality is associated with the mapper's IRR score (Brooker et al. 1998; Thornton et al. 2004), there is a lack of any criteria for the quality of DCM data for secondary uses. The findings of this study have contributed to original knowledge by showing that, according to the study participants, the mapper's reliability, along their experience and training levels, are also significant for the secondary use of DCM data, particularly for research purposes. However, further research is required to identify how mappers' reliability for producing good quality DCM data can be regulated and monitored.

Further, as discussed in Chapter 9, based on user requirements for additional data, data quality issues can emerge at integration level when such data will be collected and integrated with DCM data within a future data warehouse. Whilst this study offers an opportunity to refine and validate the requirements that emerged from inductive analysis of the interview data, the

issue of data quality will need further refinement and elaboration to find out
which other

quality issues can impede or enhance the secondary use of DCM data within the warehouse. Further, it is also important to research how the data quality issues at data-source and integration levels can be dealt with.

The study findings underline the significance of the quality of DCM data for secondary research, as study participants identified it as an important issue to be considered for the use of data for research. As argued in Chapter 2, data quality is also an important requirement for benchmarking DCM data. In order to ensure the quality of data for secondary uses, it could be suggested that a framework should be developed for assessing the quality of DCM data for this purpose. The literature indicates (Chapter 3) the lack of any agreement on a single framework, or on quality dimensions based on which the quality of healthcare data for secondary use can be assessed and evaluated. However, it has been argued in Chapter 3 that the identification and definition of data-quality dimensions depend on the nature of the data, the context within which it is used and the system which will manage the data. Based on this argument, it could be suggested that a DCM data-quality framework should take into consideration the users' perception of data quality (the findings of chapters 8 and 9) and the fact that the intrinsic aspects of DCM data are complex.

10.5.3. Data security

Data security is a key issue in warehousing healthcare data (Kaplan 2014; Lamas et al. 2015). As data is shared across various individuals and organisations, it has implications for ensuring the security and anonymity of data. Legally, any data needs to be secured before it is shared across

organisations and countries through an information system. However, this issue becomes even more important when the data is related to people's

health and personal identity (Department of Health 2011). There are strict government rules and NHS guidelines relating to patients' data security and the sharing of protocols. These require the data warehouse to be physically secured as well. The user requirements determine the data-security models that need to be within the context of government and NHS policies on patient data security and anonymisation before they become part of the design process. These models also define the users' data-access roles and data-anonymisation concerns within the system. The study findings suggest the requirement for anonymised data for research use of DCM data (Chapter 8). This requires further technical research to identify data anonymisation techniques for both structured and unstructured data collected within a future data warehouse for DCM. This will also require adherence to the data anonymisation rules provided by the NHS (Chester 2011). However, it is also important to establish who will be responsible for data anonymisation. The literature shows that both data providers (e.g. health and social-care organisations) and data controllers/processors (e.g. data-warehouse providers) can share the responsibility. In the context of DCM, it is therefore also important to establish ethical and legal policies to ensure the appropriate secondary use of DCM data by both data providers and data controllers.

As discussed in Chapter 3, warehousing data taken from various countries for secondary uses has been reported as challenging, as countries have their own national legislation for dealing with privacy and data protection (Elger et al. 2010), including their own set of guidelines for data anonymisation within

each country (Lamas et al. 2015). The user requirement for inter-country comparisons using DCM data also implies the need to ensure that each

country's ethical and legal issues are taken into consideration while collecting international data within the warehouse.

10.5.4. System usability

While the present study suggests the provision of interfaces for the user through a data-search facility (as discussed in Chapter 8), it is important to gather further requirements to refine the usability aspects of such interfaces. The study findings have important implications for developing system interfaces that are user friendly, considering the users' limited technical knowledge.

10.5.5. DCM data-management systems for primary purposes

Further research is required to understand the in-depth issues and factors influencing the development, implementation and management of a DCM database within organisations and its adoption by users. The successful use and adoption of the archive database can ensure the availability of quality DCM data within the warehouse. Further, resources are also required for the development and implementation of a future data warehouse by the system provider.

10.5.6. Development of the methodology

This was a qualitative study, in which an inductive approach was employed to identify mappers' views, perceptions, expectations and needs regarding their potential secondary use of DCM data and then to translate these into requirements for a data warehouse. The methodology used to identify the views of users who were unfamiliar with the new system can potentially be used in similar situations within any system-design process. Further, it would

also be interesting to see if the guidelines of grounded theory could be used to identify requirements from a large sample and whether this would reproduce the study findings. This study, therefore, could be used as a basis for the development of a quantitative survey questionnaire to gather focused requirements from a specific DCM data warehouse user group.

10.6. Limitations of the study

This study is limited to exploring only the requirements of those individuals who were recruited based on their knowledge of DCM data such as the mappers. In the absence of any existing knowledge about who could be the potential users of a future data warehouse, this study began with an assumption that mappers could be among them. While the findings of this study have validated this assumption (Chapter 7), it is important to gather requirements from those who might not be mappers but who would like to use DCM data for quality improvement and research purposes. These could be care-quality improvement organisations, universities and dementia charities. However, it is recommended that a prototype (based on the findings of this study) should be developed for eliciting requirements from those who are not familiar with DCM data to demonstrate the type of information that they can retrieve from the warehouse.

Another possible limitation of this study is that the one-to-one interviewing method was the only one used for data collection. It would have been interesting to use other data-collection methods, such as focus groups, for two main reasons. The first would have been to triangulate the data to see how different methods could produce data in regard to similar concepts. The second would have been that, through use of a focus group method, the

organisational view would have been potentially identified as well. The organisation's view of the secondary use of DCM data and related concerns and issues would have conveyed knowledge of the requirements for a data warehouse in which organisations could have been the potential users. However, as mentioned in Chapter 6, it was not feasible to conduct a focus group within this study (please see Section 6.4).

Another possible limitation could be time and resources. According to Strauss and Corbin (1998), the use of grounded theory can identify a number of concepts within the data, which could be pursued further. During data analysis, all three potential secondary uses of DCM data could have been taken further to explore users' information requirements in detail. However, a pragmatic decision was taken to focus on the potential use of data for research purposes. The main reason was also that the researchers were more expressive in stating their information needs and associated issues and concerns alongside their perceptions of using DCM data.

As this study has taken an interpretive approach, where the researcher's role is significant for understanding and making sense of users' views of the system requirements, the researcher's background knowledge and skills play an important role in the sense-making process. Therefore, as this study has been carried out with the same participants, but with a different researcher with different knowledge and background, her/his interpretation of the data might have produced different findings. Gasson (2003) acknowledges this by asserting that different researchers can report the same data and methods differently, as each individual holds her/his subjective interpretations and

views towards data. Such subjectivity can be minimised by showing reflexivity

at each stage of data collection and analysis. This study detailed the data collection and analysis process in Chapter 6.

10.7. Conclusions

DCM provides rich data that is in widespread use internationally in order to extract knowledge about how to improve the quality of care for people with dementia. However, the use of such important, rich data is only limited at a local level, which means the individual mappers or organisations collecting the data uses it within the setting or research project, for the purpose it was originally collected. The underdeveloped data sharing culture in DCM community reflects the unavailability of integrated data for secondary uses. Secondary uses of DCM data can enhance understanding about care improvements of people with dementia within formal care settings such as hospital wards and residential care settings. This, however, requires an effective technical solution such as a data warehouse that supports the provision of integrated and historic DCM data for a range of potential secondary uses. This thesis argued that designing such a system requires an understanding of the potential uses of DCM data from users' perspectives. This includes identifying their information requirements and associated issues and concerns. Together, such information can provide the requirements for designing a user-acceptable data warehouse. The study aim, therefore, was to explore the requirements for the secondary use of DCM data using a user-driven approach.

In order to achieve this aim, mappers were recruited as potential users of a future data warehouse. Given mappers' unfamiliarity with the data

warehouse and their non-technical background, a methodology was needed to support

the identification of user needs and their interpretation as data warehouse requirements. Further, the lack of any existing knowledge within the field also demanded an exploratory and inductive methodology. The choice of a modified grounded theory, underpinning interpretive philosophical assumptions, was appropriate in this context. The use of a modified grounded theory identified a number of requirements deemed significant for the secondary use of DCM data, thus suggesting that grounded theory is suitable for requirement analysis, particularly when the aim is to identify mutually concerned requirements at a conceptual level.

The key findings of this study address the research question (and study objectives) established, that is identifying requirements for the secondary use of DCM data from potential users' perspectives. First, the study has identified three potential uses of DCM data and has argued that practitioners, trainers and researchers can all be potential data warehouse users if given a system designed to meet their specific needs. This also showed mappers' intention to use a future system that will permit the secondary use of DCM data.

Second, the study has identified a set of information requirements for all three potential uses of data. The information requirements for a future data warehouse for research purposes were explored in detail. DCM researchers require contextual data and metadata, particularly the additional data that can inform them of the ethical status and quality level of the DCM data stored within the warehouse. The researchers' requirement for contextual data is also similar to the additional data requirements identified for benchmarking. The availability of this data was also discussed, highlighting the need for further

research into exploring ways of finding and collecting such information and linking it to DCM data.

As data quality is reported in the literature as being a major issue for a data warehouse in terms of storing secondary data, it was also identified as a concern by the researchers in terms of using DCM data for research purposes. Therefore, the data quality issue was further explored and the study found that three factors could potentially influence the quality and availability of DCM data, which are essential to consider in the future design of a data warehouse. These are as follows: the mapper who collects DCM data; the organisation where DCM is conducted; and the technical system that enables the storage of the collected data for primary purposes. The exploration of these factors revealed that mappers' training at advanced level, their experience and their IRR score can collectively provide an indication of the quality of DCM data for secondary uses, particularly for research purposes.

Further, the study suggested that DCM supportive organisations can potentially encourage the mappers to conduct the method on a regular basis by providing them with advanced-level training, time to map and also with support when needed. Further research is required to investigate how the use of DCM can be encouraged within organisations, thus providing an important step towards generating more DCM data and more opportunities for its secondary uses. The study also identified mappers' expressed limitations and requirements regarding existing data-management systems for primary use of DCM data, such as the Excel programme. In order to

ensure that DCM data is collected in an effective electronic format at the point of its primary uses, which

can consequently support easy and efficient secondary uses, it is suggested that primary DCM systems be designed to meet user requirements.

A DCM data resource, such as a data warehouse, could serve as an agent of change in the data-sharing culture of the DCM community and increase the demand for data, both in terms of its collection and its accessibility for secondary uses. However, before designing and technically developing such a system, this study has indicated that further research into a number of areas is needed. These include: ensuring the availability of additional data including contextual and metadata alongside DCM; increasing the use of DCM to generate more data for secondary uses; developing data-quality criteria for secondary uses; defining ethical and legal policy for secondary uses; and investigating how DCM data could be collected from the organisations and individual mappers. Further, this study also suggests increasing the use of DCM across the world and ensuring that sufficient funding and human resources are available to design, develop and implement technology for DCM data management, for both primary and secondary uses. This will consequently enhance the national and international data-sharing culture in the DCM community and opportunities to make DCM data part of big-data for dementia research.

References

- Abai, N. H. Z., Yahaya, J. H. and Deraman, A. (2013) User requirement analysis in data warehouse design: a review. *Procedia Technology* 11 (0), 801-806.
- Abidi, S. S. (2001) Knowledge management in healthcare: towards 'knowledge-driven' decision-support services. *International Journal of Medical Informatics* 63 (1-2), 5-18.
- Administrative Data Liaison Service (2012) *Supporting administrative data research in the UK*. Administrative Data Liaison Service. <http://www.adls.ac.uk/secure-anonymised-information-linkage-databank/> Accessed 20/02/2012.
- Aleem, S., Capretz, L. F. and Ahmed, F. (2015) Security issues in data warehouse. *Recent Advances in Information Technology*, 15-20.
- Alzheimer's Society (2009) *Counting the cost-caring for people with dementia on hospital wards*. https://www.alzheimers.org.uk/site/scripts/download_info.php?fileID=787
- Alzheimer's Society (2012) *Formal care of people with dementia*. http://www.alzheimers.org.uk/site/scripts/documents_info.php?documentID=402 Accessed 01/08/2012.
- Amber, A., Bajwa, S. I., Naweed, S. M. and Bashir, T. (2011) Requirement elicitation methods. *IEEE, 2nd International Conference on Mechanical, Industrial, and Manufacturing Technologies (MIMT)*.
- Ancker, J. S., Shih, S., Singh, M. P., Snyder, A., Edwards, A., Kaushal, R. and the, H. i. (2011) Root causes underlying challenges to secondary use of data. *AMIA Annual Symposium Proceedings 2011*, 57-62.
- Armour, P. G. (2000) The five orders of ignorance. *Communications of the ACM* 43 (10), 17-20.
- Audit Commission (2000) *Forget me not: mental health services for older people*. London: UK's Audit Commission.
- Bahr, A. and Schlunder, I. (2015) Code of practice on secondary use of medical data in European scientific research projects. *International Data Privacy Law* 5 (4), 279-291.
- Ballard, C., Powell, I., James, I., Reichelt, K., Myint, P., Potkins, D., Bannister, C., Lana, M., Howard, R., O'Brien, J., Swann, A., Robinson, D., Shrimanker, J. and Barber, R. (2002) Can psychiatric liaison reduce neuroleptic use and reduce health service utilization for dementia patients residing in care facilities. *International Journal of Geriatric Psychiatry* 17 (2), 140-145.
- Ballou, D.P. and Tayi, G.K., 1999. Enhancing data quality in data warehouse environments. *Communications of the ACM*, 42(1), pp.73-78.
- Barnes, S., Wasielewska, A., Raiswell, C., Drummond, B. (2013) Exploring the mealtime experience in residential care settings for older people: an observational study. *Health and Social Care in the Community* 21 (4), 442-450.
- Barquin, R. (1996) On the first issue of the journal of data warehousing. *Journal of Data Warehousing* 1, 2-6.

- Batini, C., Cappiello, C., Francalanci, C. and Maurino, A., 2009. Methodologies for data quality assessment and improvement. *ACM Computing Surveys (CSUR)*, 41(3), 16.
- Batini, C., Lenzerini, M. and Navathe, S. B. (1986) A comparative analysis of methodologies for database schema integration. *ACM Computing Survey (CSUR)* 18(4), 323-364.
- Beavis, D., Simpson, S. and Graham, I. (2002) A literature review of dementia care mapping: methodological considerations and efficacy. *Journal of Psychiatric & Mental Health Nursing* 9 (6), 725-736.
- Berndt, D. J., Hevner, A. R. and Studnicki, J. (2003) The catch data warehouse: support for community health care decision-making. *Decision Support Systems* 35 (3), 367-384.
- Bevan, S., Baumgartner, F. R., Johnson, E. W. and McCarthy, J. D. (2013) Understanding selection bias, time-lags and measurement bias in secondary data sources: putting the Encyclopedia of Associations database in broader context. *Social Science Research* 42 (6), 1750-1764.
- Birks, M. and Mills, J. (2015) Planning grounded theory study. *Grounded Theory: A Practical Guide*. 2nd edition. Sage Publications. 15-31
- Blumer, H. (1969) *Symbolic interactionism: perspective and method*. New Jersey: Prentice-Hall.
- Boslaugh, S. (2007) *Secondary data sources for public health; a practical guide*. Parctical Guide to Biostatistics and Epidemiology. Cambridge, UK: Academic, Cambridge, UK.
- Botsis, T., Hartvigsen, G., Chen, F. and Weng, C. (2010) Secondary use of EHR: data quality issues and informatics opportunities. *Summit on Translational Bioinformatics* 2010, 1-5.
- Bradford Dementia Group (1997) *Evaluating dementia care: the DCM method 7th edition*. University of Bradford, Bradford, England, UK:
- Bradford Dementia Group (2005) *DCM 8 user's Manual*. University of Bradford, Bradford, UK: University of Bradford.
- Bradford Dementia Group (2014) *Dementia Care Mapping™: process and application*. Bardford, UK: University of Bradford.
- Breckenridge, J. and Jones, D. (2009) Demystifying theoretical sampling in grounded theory research. *Grounded Theory Review; an International Journal* 8 (2).
- Bringer, D. J., Johnston, H. L. and Brackenridge, H. C. (2006) Using computer-assisted qualitative data analysis software (CAQDAS) to develop a grounded theory project. *Field Methods* 18, 245-266.
- British Medical Association (2007) *Guidance on secondary uses of patient information*. British Medical Association, London.
- Brooker, D. (2005) Dementia care mapping: a review of the research literature. *The Gerontologist* 45 Spec No 1 (1), 11-18.
- Brooker, D. and Duce, L. (2000) Wellbeing and activity in dementia: a comparison of group reminiscence therapy, structured goal-directed group activity and unstructured time. *Aging & Mental Health* 4 (4), 354-358.
- Brooker, D., Foster, N., Banner, A., Payne, M. and Jackson, L. (1998) The efficacy of Dementia Care Mapping as an audit tool: report of a 3-year British NHS evaluation. *Aging & Mental Health* 2 (1), 60-70.

- Brooker, D. and Rogers, L. (2001) *Dementia Care Mapping (DCM) think tank transcripts*,. Bradford University, England UK, Bradford University.
- Brooker, D. and Surr, C. (2005) *Dementia care mapping: principles and practice*. University of Bradford: University of Bradford, UK.
- Brooker, D. and Surr, C. (2010) *Dementia care mapping : principles and practice*. 3rd edition. Bradford: University of Bradford.
- Brooker, D. J. and Surr, C. (2006) Dementia Care Mapping (DCM): initial validation of DCM 8 in UK field trials. *International Journal Of Geriatric Psychiatry* 21 (11), 1018-1025.
- Browne, G. J. and Ramesh, V. (2002) Improving information requirements determination: a cognitive perspective. *Information & Management* 39 (8), 625-645.
- Browning, D. and Mundy, J. (2001) *Data warehouse design considerations*. Microsoft SQL Server 2000. [http://technet.microsoft.com/en-us/library/aa902672\(v=sql.80\).aspx](http://technet.microsoft.com/en-us/library/aa902672(v=sql.80).aspx) Accessed 16/11/2012.
- Bruckner, R., M, List, B. and Schiefer, J. (2001) Developing requirements for data warehouse systems with use cases. In *Seventh Americas Conference on Information Systems*. 329-335.
- Bryant, T. and Charmaz, K. (2007) *The SAGE handbook of grounded theory*. London: SAGE.
- BSI (2010) *Use of Dementia Care Mapping for improved person-centred care in a care provider organisation*. London: BSI.
- Buneman, P., Khanna, S. and Wang-Chiew, T. (2001) Why and where: a characterization of data provenance. *Database Theory—ICDT 2001*. Springer. 316-330
- Byrd, T. A., Cossick, K. L. and Zmud, R. W. (1992) A synthesis of research on requirements analysis and knowledge acquisition techniques. *MIS Quarterly* 16 (1), 117-138.
- Böhnlein, M. and Ulbrich-vom Ende, A. (2000) Business process oriented development of data warehouse structures. *Data Warehousing 2000: Methoden, Anwendungen, Strategien*. Heidelberg: Physica-Verlag HD. 3-21 http://dx.doi.org/10.1007/978-3-642-57681-2_1
- Campbell, S. M., Braspenning, J., Hutchinson, A. and Marshall, M. N. (2003) Research methods used in developing and applying quality indicators in primary care. *British Medical Journal* 326, 816-819.
- Capstick, A. (2003) The theoretical origins of dementia care mapping. In Innes, A. (editor) *Dementia Care Mapping; applications across cultures*. Baltimore, Maryland: Health Professions Press, Inc. 11-24
- Caracelli, V. J. and Greene, J. C. (1997) *Crafting mixed-method evaluation designs*. Jossey-Bass Publishers.
<http://www.academyhealth.org/files/phsr/Caracelli%20and%20Greene.pdf> Accessed 15/12/2012.
- Care Quality Commission (2016) *Using data to monitor services*. Care Quality Commission, 2016. <http://www.cqc.org.uk/content/using-data-monitor-services#data-examples> Accessed 02/01/2016.
- Chakraborty, S. and Dehlinger, J. (2009) Applying the grounded theory method to derive enterprise system requirements. In *Proceedings of the 2009 10th ACIS International Conference on Software Engineering, Artificial Intelligences, Networking and Parallel/Distributed Computing*

- (SNPD '09). IEEE Computer Society, Washington, DC, USA.
<http://dx.doi.org/10.1109/SNPD.2009.102>.
- Charmaz, K. (2006) *Constructing grounded theory: a practical guide through qualitative analysis*. London: Sage.
- Chaudary, S. and Dayal, U. (1997) An overview of data warehousing and OLAP technology. *ACM, SIGMOD Archive* 26 (1).
- Chen, E. T. (2012) Implementation issues of enterprise data warehousing and business intelligence in the healthcare industry. In *Communications of the International Information Management Association (IIMA)*. Vol. 12. 39-50.
- Chen, T.-J., Chou, L.-F. and Hwang, S.-J. (2003) Application of a data-mining technique to analyze coprescription patterns for antacids in Taiwan. *Clinical Therapeutics* 25 (9), 2453-2463.
- Chenoweth, L. and Jeon, Y. (2007) Determining the efficacy of Dementia Care Mapping as an outcome measure and a process for change: a pilot study. *Aging & Mental Health* 11 (3), 237-245.
- Chenoweth, L., King, M. T., Jeon, Y.-H., Brodaty, H., Stein-Parbury, J., Norman, R., Haas, M. and Luscombe, G. (2009) Caring for aged dementia care resident study (CADRES) of person-centred care, dementia-care mapping, and usual care in dementia: a cluster-randomised trial. *Lancet Neurology* 8 (4), 317-325.
- Chester, C. (2011) *NHS business services-pseudonymisation and anonymisation of data policy*. Services, N. B. England: NHSBSA.
- Chisholm, M. (2008) *Six architectural styles of data hubs*. BeyeNetwork. <http://www.b-eye-network.com/view/8109> Accessed 21/06/2015.
- Christel, G. M. and Kang, C. K. (1992) *Issues in requirement elicitation*. Pittsburgh, Pennsylvania 15213: Software Engineering Institute.
- Clarke, A. E. (2003) *Situational analysis: grounded theory mapping after the postmodern turn*. California, USA, University of California Press.
- Completo, J., Cruz, R. S., Coheur, L. and Delgado, M. (2012) Design and implementation of a data warehouse for benchmarking in clinical rehabilitation. *Procedia Technology* 5 (0), 885-894.
- Cooke, H. A. and Chaudhury, H. (2012) An examination of the psychometric properties and efficacy of Dementia Care Mapping. *Dementia*.
- Corbin, J. and Strauss, A. (2007) *Basics of qualitative research: techniques and procedures for developing grounded theory*. 3 edition. SAGE Publications, Inc.
- Coyne, I. T. (1997) Sampling in qualitative research. Purposeful and theoretical sampling; Merging or clear boundaries? *Journal of Advanced Nursing* 26, 623-630.
- Cresswell, K. and Sheikh, A. (2013) Organizational issues in the implementation and adoption of health information technology innovations: an interpretative review. *International Journal of Medical Informatics* 82 (5), 73-86.
- Creswell, J. W. (1994) *Research design : qualitative & quantitative approaches*. Thousand Oaks, Calif.: Sage Publications.
- Creswell, J. W. (2007) *Qualitative inquiry and research design: choosing among five approaches*. 2nd edition. London: Sage Publications.

Creswell, J. W. (2009) *Research design : qualitative, quantitative, and mixed methods approaches*. 3rd edition. Thousand Oaks, CA. ; London: Sage.

- Crook, C. W. and Kumar, R. L. (1998) Electronic data interchange: a multi-industry investigation using grounded theory. *Information & Management* 34 (2), 75-89.
- Cummings, E. and Turner, P. (2009) Patient self-management and chronic illness: evaluating outcomes and impacts of information technology. 143, 229-234.
- Dalal, N. P. and Yadav, S. B. (1992) The Design of a knowledge-based decision support system to support the information analyst in determining requirements. *Decision Sciences* 23 (6), 1373-1388.
- Darke, P. and Shanks, G. (1995) Defining system requirements: a critical assessment of the NIAM conceptual design procedure. *Australasian Journal of Information System* 2 (2), 50-62.
- Davis, C. H., Tremblay, M. C., Fuller, R. M. and Berndt, D. J. (2006) Communication challenges in requirements elicitation and the use of the repertory grid technique. *Journal of Computer Information Systems* (Special Issue), 78-86.
- De Moor, G., Sundgren, M., Kalra, D., Schmidt, A., Dugas, M., Claerhout, B., Karakoyun, T., Ohmann, C., Lastic, P. Y., Ammour, N., Kush, R., Dupont, D., Cuggia, M., Daniel, C., Thienpont, G. and Coorevits, P. (2015) Using electronic health records for clinical research: the case of the EHR4CR project. *Journal of Biomedical Informatics*, 162-73.
- Deelman, E., Berriman, B., Chervenak, A., Corcho, O., Groth, P. and Moreau, L. (2010) Metadata and provenance management. In Arie Shoshani, D. R. (editor) *Scientific Data Management: Challenges, Existing technology and Deployment*. CRC Press.
- Deetjen, U., Meyer, E. T. and Schroeder, R. (2015) *Big data for advancing dementia research: an evaluation of data sharing practices in research on age-related neurodegenerative diseases*. Paris, France, OECD Publications.
- del Hoyo-Barbolla, E. and Lees, D. (2002) The use of data warehouse in the healthcare sector. *Health Information Journal* 8 (1), 43-46.
- Denzin, N. K. and Lincoln, Y. S. (1998) *Strategies of qualitative inquiry*. SAGE Publications, Inc.
- Department of Health (2009) *Living well with dementia: a national dementia strategy*. London: http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_094051.pdf Accessed 13/2/2013.
- Department of Health (2010) *Essence of Care 2010: Benchmarks for the fundamental aspects of care*. England, UK, Department of Health.
- Department of Health (2011) *Data protection Act 1998*. http://webarchive.nationalarchives.gov.uk/+www.dh.gov.uk/en/Managingyourorganisation/Informationpolicy/Recordsmanagement/dh_4000489 Accessed 22/12/2012.
- Department of Health (2012) *The prime minister's challenge on dementia: delivering major improvements in dementia care and research by 2015: a report on progress*. Department of Health. <https://www.wp.dh.gov.uk/dementiachallenge/files/2012/11/The-Prime-Ministers-Challenge-on-Dementia-Delivering-major-improvements-in-dementia-care-and-research-by-2015-A-report-of-progress.pdf> Accessed 01/12/2012.

- Department of Health (2013a) *A review of the potential benefits from the better use of information and technology in health and social care: a final report*. London, UK: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/217012/DoH-Review-of-Information-and-Technology-Use-Final-Report-V2.pdf
- Department of Health (2013b) *Dementia: a state of the nation report on dementia care and support in England*. UK: Department of Health. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/262139/Dementia.pdf
- Department of Health (2013c) *G8 Dementia summit declaration*. London, UK: Department of Health. <https://www.gov.uk/government/publications/g8-dementia-summit-agreements/g8-dementia-summit-declaration> Accessed 05/03/2015.
- Department of Health (2015a) *2010-2015 Government policy: dementia*.
- Department of Health (2015b) *Prime minister's challenge in dementia 2020*. UK: Department of Health. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/414344/pm-dementia2020.pdf
- Detmer, D. E. (2000) Information technology for quality health care: a summary of United Kingdom and United States experiences. *Quality Health Care* 9 (3), 181-189.
- Dey, I. (1999) *Grounding Grounded Theory: Guidelines for qualitative Inquiry*. 1 edition. Emerald Group Publishing Ltd.
- Douglass, C., Keddle, A., Brooker, D. and Surr, C. (2010) Cross-cultural comparison of the perceptions and experiences of dementia care mapping "mappers" in the United States and the United Kingdom. *Journal of Aging and Health* 22 (5), 567-588.
- Earle, C. C., Neville, M. B., Landrum, J. M., Souza, J. C., Weeks, S. D., Block, E., Grunfeld, J. Z. and Ayanian, J. Z. (2005) Evaluating claims-based indicators of the intensity of end-of-life cancer care. *International Journal for Quality in Health Care* 17 (6), 505-9.
- Edelman, P., Kuhn, D. and Fulton, B. R. (2004) Influence of cognitive impairment, functional impairment and care setting on dementia care mapping results. *Aging & Mental Health* 8 (6), 514-523.
- Edgren, G., Hjalgrim H., Reilly, M., Tran, T. N., Rostgaard, K., Shanwell, A., Titlestad, K., Adami, J., Wikman, A., Jersild, C., Gridley, G., Wideroff, L., Nyren, O., and Melbye, M. (2007) Risk of cancer after blood transfusion from donors with subclinical cancer: a retrospective cohort study. (1474-547X (Electronic)).
- Edvardsson, D. and Innes, A. (2010) Measuring person-centered care: a critical comparative review of published tools. *The Gerontologist* 50 (6), 834-846.
- Einbinder, J. S. and Scully, K. (2002) Using a clinical data repository to estimate the frequency and costs of adverse drug events. *Journal of American Medical Association* 9 (6 Supply. 1), 34-38.
- El Emam, K. (2011) Methods for the de-identification of electronic health records for genomic research. *Genome Medicine* 3 (4), 25.

El Emam, K., Dankar, F., Issa, R., Jonker, E., Amyot, D., Cogo, E.,
Corriveau, J. P., Walker, M., Chowdhury, S., Vaillancourt, R., Roffey,
T. and

- Bottomley, J. (2009) A globally optimal k-anonymity method for the de-identification of health data. *Journal of American Medical Informatics Association* 16, 670-682.
- El Mohajir, M. and Latrache, A. (2012) Unifying and incorporating functional and non functional requirements in datawarehouse conceptual design. *Information Science and Technology (CIST), 2012 Colloquium in.* 22-24 Oct. 2012.
- Eleveld, M., Schrimpf, W. and Siegert, A. (2003) User requirements and information definition for a virtual coastal and marine data warehouse. *Ocean and Coastal Management* 46 (6-7), 487-505.
- Elger, B. S., Iavindrasana, J., Lo Iacono, L., Müller, H., Roduit, N., Summers, P. and Wright, J. (2010) Strategies for health data exchange for secondary, cross-institutional clinical research. *Computer Methods and Programs in Biomedicine* 99 (3), 230-251.
- Ellis, J. (2006) All inclusive benchmarking. *Journal of Nursing Management* 14 (5), 377-83.
- Elsheikh, Y. M. A. (2011) *A model for the adoption and implementation of web-based government services and applications: a study based in grounded theory validated by structural equation modelling analysis in a Jordanian context.* Bradford: University of Bradford, <http://bradscholars.brad.ac.uk/>
- Ettorchi-Tardy, A., Levif, M. and Michel, P. (2012) Benchmarking: a method for continuous quality improvement in health. *Health Policy* 7 (4), 101-119.
- EU Directive (1995) *EU Directive 95/46/EC-The data protection directive.* Data Protection Commissioner: <https://www.dataprotection.ie/docs/EU-Directive-95-46-EC-Chapter-1/92.htm> Accessed 12/01.
- European Commission (2004) *e-Health—making healthcare better for European citizens: an action plan for a European e-Health area.* Brussels: http://ec.europa.eu/information_society/doc/qualif/health/COM_2004_0356_F_EN_ACTE.pdf Accessed 12/03/2012.
- European Medical Information Framework (2013) *Innovative Medicine Initiative (IMI),* <http://www.imi.europa.eu/content/emif> Accessed 17/06/2015.
- Faria, P. L. d. and Cordeiro, J. V. (2014) Health data privacy and confidentiality rights: crisis or redemption? *Revista Portuguesa de Saúde Pública* 32 (2), 123-133.
- Fernandez, D. W., Lehmann, H. and Underwood, A. (2002) Rigour and relevance in studies of IS innovation: A grounded theory methodology approach. In *European Conference on Information Systems (ECIS).* Poland. 110-119. <http://is2.lse.ac.uk/asp/aspecis/20020149.pdf> Accessed: 07/08/2012.
- Finnamore, T. and Lord, S. (2007) The use of dementia care mapping in people with a learning disability and dementia. *Journal of Intellectual Disabilities* 11 (2), 157-165.
- Fisher, C. W., Chengalur-Smith, I. and Ballou, D. P. (2003) The impact of experience and time on the use of data quality information in decision making. *Information Systems Research* 14 (2), 170-188.

- Ford, D., Jones, K., Verplancke, J.-P., Lyons, R., John, G., Brown, G., Brooks, C., Bodger, O., Couch, T. and Leake, K. (2009) The SAIL databank: building a national architecture for e-health research and evaluation. *BMC Health Services Research* 9 (157).
- Fossey, J., Lee, L. and Ballard, C. (2002) Dementia Care Mapping as a research tool for measuring quality of life in care settings: psychometric properties. *International Journal of Geriatric Psychiatry* 17 (11), 1064-1070.
- Galal-Edeen, G. H. (2005) Information systems requirements engineering: an interpretive approach. *The Egyptian Informatics Journal* 6 (2), 154-174.
- Galal-Edeen, G. H. and Paul, R. J. (1999) A qualitative scenario approach to managing evolving requirements. *Requirements Engineering* 4 (2), 92-102.
- Gallivan, M. J. and Keil, M. (2003) The user-developer communication process: a critical case study. *Information Systems Journal* 13 (1), 37-68.
- Gallo, C., De-Bonis, M. and Perilli, M. (2010) *Data warehouse design and management: Theory and practice*. Foggia, Italy, Dipartimento di Scienze Economiche, Metematiche e Statistiche.
- Gam, I. and Salinesi, C. (2006) A requirement-driven approach for designing data warehouses. *Foundations of Software Quality*. Luxembourg.
- Gardyn, E. (1997) A data quality handbook for a data warehouse. In Strong, D. M. A. K., Beverly K. (editor) *IQ*. MIT. 267-290 <http://dblp.uni-trier.de/db/conf/iq/iq1997.html#Gardyn97>
- Gasson, S. (2003) *Rigor in grounded theory research - an interpretive perspective on generating theory from qualitative field studies*. Hershey PA, Idea Group Publishing.
- Geissbuhler, A., Safran, C., Buchan, I., Bellazzi, R., Labkoff, S., Eilenberg, K., Leese, A., Richardson, C., Mantas, J., Murray, P. and De Moor, G. (2013) Trustworthy reuse of health data: a transnational perspective. *International Journal of Medical Informatics* 82 (1), 1-9.
- Ghaffarian, V. (2011) The new stream of socio-technical approach and main stream information systems research. *Procedia Computer Science* 3 (0), 1499-1511.
- Giannoccaro, A. and Shanks, G. (1999) Stakeholder perceptions of data quality in a data warehouse environment. In *10th Australian Conference on Information Systems*. 344-355.
- Gigliotti, C. M., Jarrott, S. E. and Yorgason, J. (2004) Harvesting health: Effects of three types of horticultural therapy activities for persons with dementia. *Dementia (14713012)* 3 (2), 161-180.
- Giorgini, P., Rizzi, S. and Garzetti, M. (2005) Goal-oriented requirement analysis for data warehouse design. In *DOLAP' 05*. Bremen, Germany. ACM.
- Giorgini, P., Rizzi, S. and Garzetti, M. (2008) GRAnD: A goal-oriented approach to requirement analysis in data warehouses. *Decision Support Systems* 45 (1), 4-21.
- Glaser, B. and Strauss, A. (1967) *The discovery of grounded theory: strategies for qualitative resaerch*. Aldine Transaction.

Glaser, B. G. (1992) *Basics of grounded theory analysis: emergence vs forcing*. Mill Valey, Calif.: Sociology Press.

- Glaser, B. G. and Strauss, A. L. (1966) *Awareness of dying*. London: Weidenfeld and Nicolson.
- Glaser, B. G. and Strauss, A. L. (1968) *The discovery of grounded theory: strategies for qualitative research*. Observations. London: Weidenfeld & Nicolson.
- Glaser, B. G. and Strauss, A. L. (1999) *The discovery of grounded theory: strategies for qualitative research*. New York: Aldine Transaction.
- Glaser, G. B. (1978) *Theoretical sensitivity: advances in the methodology of grounded theory*. Mill Valley, California: Sociology Press.
- Glavic, B. and Dittrich, R. K. (2007) Data provenance: a categorization of existing approaches. *Datenbanksysteme in Business, Technologie und Web (BTW)*. Aachen, Germany.
- Goble, C. and Stevens, R. (2008) State of the nation in data integration for bioinformatics. *Journal of Biomedical Informatics* 41 (5), 687-693.
- Goldstein, H. and Spiegelhalter, D. J. (1996) League tables and their limitations: statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society* 1 (Series, A), 385-443.
- Golfarelli, M. (2010) From user requirements to conceptual design in warehouse design: a survey. In Bellatreche, L. (editor) *In Data Warehousing Design and Advanced Engineering Applications: Methods for Complex Construction*. 1-16
<http://bias.csr.unibo.it/golfarelli/papers/surveygolfarelli08.pdf> Accessed 08/10/2012.
- Gosain, A. and Singh, J. (2008) Towards data warehouse business quality through requirement elicitation. In *IEEE*. 464-469.
- Goulding, C. (2002) *Grounded theory: a practical guide for management, business and market researchers*. London: SAGE.
- GPRD (2010) *The general practice research database*. London, UK:
http://www.erskinepractice.scot.nhs.uk/website/S11486/files/GPRD_PatientLeaflet.pdf
- Gray, D. E. (2009) *Doing research in the real world*. 2nd edition. Sage Publications Ltd.
- Gray, G. W. (2004) Challenges of building clinical data analysis solutions. *Journal of Critical Care* 19 (4), 264-270.
- Gray, P. and Watson, H. (1998) *Decision support in the data warehouse*. Prentice-Hall, Upper Saddle River.
- Greg, H. (2013) *Top 10 countries for EHR adoption*,. Becker's Healthcare.
<http://www.beckershospitalreview.com/healthcare-information-technology/top-10-countries-for-ehr-adoption.html> Accessed 07/06/2015.
- Guba, E. G. and Lincoln, Y. S. (1981) *Effective evaluation*. The Jossey-Bass higher education series. San Francisco; London: Jossey-Bass.
- Guba, E. G. and Lincoln, Y. S. (1989) *Fourth generation evaluation*. Newbury Park, Calif.; London: Sage Publications.
- Guba, E. G. and Lincoln, Y. S. (1994) *Competing paradigms in qualitative research: handbook of qualitative research*. Thousand Oaks, California CA: Sage.

Guo, Y., Tang, S., Tong, Y. and Yang, D. (2006) Triple-driven data modeling methodology in data warehousing: a case study. *In Proceedings of the 9th ACM international workshop on Data warehousing and OLAP*

- (DOLAP '06). Arlington, Virginia, USA. ACM, New York, NY, USA. <http://doi.acm.org/10.1145/1183512.1183524>. Accessed 11/05/2012.
- Halaweh, M. (2012) Using grounded theory as a method for system requirements analysis. *JISTEM J.Inf.Syst. Technol. Manag.* 9 (1), 23-38.
- Hansen, H. B. and Kautz, K. (2005) Grounded theory applied- studying information systems development methodologies in practice. *Proceedings of the 38th Hawaii International Conference on System Sciences*. Hawaii.
- Harjinder, G. and Rao, P. C. (1996) *Data warehousing*. Hispanoamericana, S. A: Prentice-Hall.
- Hartig, O. and Zhao, J. (2009) Using web data provenance for quality assessment. *CEUR Workshop proceedings*.
- Hass, K. B., Horst, V. R., Ziemski, K. and Lindberg, L. (2007) *From analyst to leader: elevating the role of the business analyst*. Management Concepts Press.
- Health and Social Care Information centre (2012) *Supporting transparency and open data*. UK: <http://www.hscic.gov.uk/transparency> Accessed 20th May 2015.
- Health and Social Care Information Centre (2012) *The quality of nationally submitted health and social care data, England-First Annual Report, 2012, Experimental Statistics*. England: Health and Social Care Information Centre, Information Governance and Data Quality. V1.1.
- Health and Social Care Information Centre (2013) *A guide to confidentiality in health and social care; treating confidential information with respect*. Health and Social Care Information Centre. <http://www.hscic.gov.uk/media/12822/Guide-to-confidentiality-in-health-and-social-care/pdf/HSCIC-guide-to-confidentiality.pdf>
- Health and Social Care Information Centre (2015a) *Information and technology for better care; Health and Social Care Information centre Strategy 2015-2020*. London, UK: Health and Social Care Information Centre. http://www.hscic.gov.uk/media/16232/HSCIC-Draft-Strategy-2015-2020-Information-and-technology-for-better-care/pdf/80435_HSCIC_Strategy_2015-2020-v1g_%281%29.pdf
- Health and Social Care Information Centre (2015b) *Spine services*. England, UK: Health and Social Care Information Centre. <http://systems.hscic.gov.uk/spine> Accessed 11/07/2015.
- Health and Social Care Information Centre (2015c) *Hospital Episode Statistics (HES)*. London, UK: Department of Health. <http://www.hscic.gov.uk/hes> Accessed 12/06/2015.
- Health Canada (2001) *Towards electronic health records*. Office of Health and the Information Highway, Health Canada. http://www.hc-sc.gc.ca/hcs-sss/alt_formats/pacrb-dgapcr/pdf/pubs/ehealth-esante/2001-towards-vers-ehr-dse/2001-towards-vers-ehr-dse-eng.pdf Accessed 01/01/2012.
- Health Care Reform (2010) *Using information technology to support better health care: One infrastructure with many uses*. <http://www.brookings.edu/~media/events/2010/5/14%20health%20information/final%20issue%20brief%2051310.pdf> Accessed 15/01/2012.

- Health Industries (2009) *Transforming health care through secondary use of health data*. <http://robertlu.med.ncku.edu.tw/data/02092.pdf>
- Health Information and Quality Authority (2012) *International review of secondary use of personal health information*. Ireland, Health information and Quality Authority.
- Hermann, R. C., Chan, J. F., Provost, S. F. and Chiu, W. T. (2006) Statistical benchmarks for process measures of quality of care for mental and substance use disorders. *Psychiatry Service* 57 (10), 1461-7.
- HiMSS (2009) *Data warehousing, a new focus in healthcare data warehousing*. The Healthcare Information and Management Systems Society (HiMSS).
- Hirschheim, R. (1992) Information systems epistemology: An historical perspective. In Galliers, R. D. (editor) *Information Systems Research-Issues, Methods, and Practice Guidelines*. Henley-on-Thames, England: Alfred Walter. 61-88
- Hoets, H. (2009) *What is bias in qualitative research?* , FocusGroupTips.com.
- Holstein, A. J. and Gubrium, J. F. (1995) *The active interview*. Thousand Oaks, California: Sage Publications Inc.
- Hoppszallern, S. (2012) Good-bye to paper. *Hospitals & Health Networks / AHA* 86 (2), 22-22.
- Hovenga, E. J. S. and Grain, H. (2013) *Health information governance in a digital environment*. (Studies in Health Technology and Informatics) Vol. 193 Amsterdam: IOS Press. <http://ezproxy.ecu.edu.au/login?url=http://ebooks.iospress.nl/volume/health-information-governance-in-a-digital-environment> Accessed 2013.
- Hox, J. J. and Boeije, R. H. (2005) *Data collection, primary vs, secondary*. Elsevier Inc.
- Huberman, A. M. and Miles, M. B. (2002) *The qualitative researcher's companion*. London: Sage Publications.
- Hughes, J. and Jones, S. (2012) *Reflections on the use of grounded theory in interpretive information systems research*. <http://is2.lse.ac.uk/asp/aspecis/20030065.pdf> Accessed 27/07/2012.
- Hung, S.-Y., Tsai, J. C.-A. and Chuang, C.-C. (2014) Investigating primary health care nurses' intention to use information technology: An empirical study in Taiwan. *Decision Support Systems* 57 (0), 331-342.
- iCareHealth (2014) *3 common barriers to the uptake of technology (and ideas for solving them)*. iCareHealth, UK.
- IGGI, I.-g. G. o. G. I. (2005) *The principles of good data management*. London, UK: Queen's Printer and Controller of Her Majesty's Stationery Office. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/14867/Good_dataMan.pdf Accessed 06/06/2015.
- Iivari, J. and Hirschheim, R. (1996) Analyzing information systems development: a comparison and analysis of eight development approaches. *Information Systems* 21 (7), 551-575.
- IMIA: International Medical Informatics Association (2012) *2012 European summit on trustworthy reuse of health data*. International Medical

Informatics Association. <http://euhealthdata2012.imia.info/> Accessed 05/05/2012.

- Information Commissioner's Office (2015) *Data protection principles*. <https://ico.org.uk/for-organisations/guide-to-data-protection/data-protection-principles/> Accessed 16/01/2016.
- Information commissioner's Office (2016) *Data sharing code of practice*. https://ico.org.uk/media/for-organisations/documents/1068/data_sharing_code_of_practice.pdf Accessed 14/02/2016.
- Inmon, H. W. (2012a) *Building the data warehouse: Getting started*. <http://inmoncif.com/inmoncif-old/www/library/whiteprs/ttbuild.pdf> Accessed 03/11/2012.
- Inmon, W. H. (1996) *Building the data warehouse*. 4 edition. John Wiley and Sons.
- Inmon, W. H. (2012b) *Data Warehousing in the Healthcare Environment*. <http://inmoncif.com/registration/whitepapers/DATA%20WAREHOUSING%20IN%20THE%20HEALTHCARE%20ENVIRONMENTR1.pdf> Accessed 10-06.
- Innes, A. and Surr, C. (2001) Measuring the well-being of people with dementia living in formal care settings: the use of Dementia Care Mapping. *Aging & Mental Health* 5 (3), 258-268.
- Innovative Medicines Initiatives (2014) *Code of practice on secondary use of medical data in scientific research projects*.
- ISO 9241-2010 (2010) *Ergonomics of human-system interaction-part 2010: Human-centred design for interactive system*. http://www.iso.org/iso/catalogue_detail.htm?csnumber=52075
- Janowicz, K. (2009) *Trust and provenance; you can't have one without the other*. University of Muenster, Germany: Institute of Geoinformatics, University of Muenster.
- Jantunen, S. and Gause, D. C. (2014) Using a grounded theory approach for exploring software product management challenges. *Journal of Systems and Software* 95, 32-51.
- Jarrott, S. E. and Bruno, K. (2003) Intergenerational activities involving persons with dementia: An observational assessment. *American Journal of Alzheimer's Disease and Other Dementias* 18 (1), 31-37.
- Jaycock, S., Persaud, M. and Johnson, R. (2006) The effectiveness of dementia care mapping in intellectual disability residential services: a follow-up study. *Journal of Intellectual Disabilities* 10 (4), 365-375.
- Jensen, P. B., Jensen, L. F. and Brunak, S. (2012) Mining electronic health records: towards better research applications and clinical care. (1471-0064 (Electronic)).
- Johnston, H. L. (2004) Technical and methodological learning curves: reflections on the use of QSR-NVivo in doctoral research. In *Strategies in Qualitative Research: Methodological issues and Practices in using QSR NVivo and Nud*ist*. University of Durham, UK.
- Jones, M. and Alony, I. (2011) Guiding the use of grounded theory in doctoral studies-an example from the Australian film industry. *International Journal of Doctoral Studies* 6, 95-114.
- Jones, S., Hamilton, J. and Surr, C. (2014) Dementia Care Mapping: building up a bigger picture. *Journal of Dementia Care* 22 (6), 19-21.
- Jones, S., Hamilton, J. and Surr, C. (in press) Monitoring and improving the quality of person-centred care in health and social settings using

- Dementia Care Mapping (DCM): staff experiences of implementation barriers and supports. *European Journal of Person Centred Healthcare*.
- Kahn, B. K., Strong, D. M. and Wang, R. Y. (1997) Model for delivering quality information as product and service. In *Conference on Information Quality*. Cambridge, MA. 80-94.
- Kaldeich, C. and Sá, J. O. e. (2004) Data warehouse methodology: a process driven approach. *Advanced Information Systems Engineering: 16th International Conference, CAiSE 2004, Riga, Latvia, June 7-11, 2004. Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg. 536-549 http://dx.doi.org/10.1007/978-3-540-25975-6_38
- Kalra, D. and Ingram, D. (2006) Electronic health records. *Information technology solutions for healthcare*. Springer. 135-181
- Kaplan, B. (2014) How should health data be used? Privacy, secondary use, and big data sales. *Institution for Social and policy Studies ISPS14-025*, 1-39.
- Kaplan, B. and Maxwell, J. A. (1994) Qualitative research methods for evaluating computer information systems. In J.G. Anderson, C. E. A. A. S. J. J. (editor) in *Evaluating Health Care Information Systems: Methods and Applications*. Thousand Oaks, CA: Sage. 45-68
- Kaplan, B. and Maxwell, J. A. (2005) Qualitative research methods for evaluating computer information systems. In Aydin, J. A. A. C. (editor) *Evaluating the organizational impact of health care information systems*. 2nd edition edition. New York: Springer.
- Kapplerman, L. A. (1994) User engagement in the development, implementation and use of information technologies. *Proceedings of the twenty-seventh Hawaii International conference*. Wailea, HI, USA, Vol. 4.
- Kassim, E. S., Jailani, S. F. A. K., Hairuddin, H. and Zamzuri, N. H. (2012) Information system acceptance and user satisfaction: the mediating role of trust. *Procedia - Social and Behavioral Sciences* 57, 412-418.
- Kay, F. L. J. (2007) *Health care benchmarking*. The Hongkong Medical diary.
- Kelle, U. (1995) Introduction: An overview of computer-aided methods in qualitative research. In U. Kelle, G. P., and K. Bird (editor) *Computer-Aided Qualitative Data Analysis: Theory, Methods and Practice*. London: Sage. 1-17
- Kelly, M. (2010) The role of theory in qualitative health research. *Family Practice* 27 (3), 285-290.
- Kerlinger, F. N. and Lee, H. B. (2000) *Foundations of behavioral research*. 4 edition. University of California, USA: Harcourt College Publishers.
- Khalid, S. (2009) *A web-based DBMS for an international data repository on the quality of dementia care*. Bradford: University of Bradford,. <http://virtualscholars.brad.ac.uk:8080>
- Khalid, S. (2010) *Towards data governance for international Dementia Care Mapping (DCM) : a study proposing DCM data management through a data warehousing approach*. Bradford: University of Bradford,. <http://bradscholars.brad.ac.uk/>
- Khalid, S., Surr, C. and Neagu, D. (2010) DCM data management framework: A data warehousing approach. *Information Technology in Bio- and Medical Informatics, ITBAM 2010*. Bilbao, Spain, Vol. 6266.

Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-642-15020-3_4.

- Kim, Y. (2010) The pilot study in qualitative inquiry: identifying issues and learning lessons for culturally competent research. *Qualitative Social Work*.
- Kimball, R. and Ross, M. (2002) *The data warehouse toolkit; the complete guide to dimensional modeling*. Canada: John Wiley and Sons, Inc.
- Kimball, R., Ross, M., Thornwaite, J., Mundy, J. and Becker, B. (2008) *The data warehouse lifecycle toolkit*. Indianapolis: Wiley publishing Inc.
- Kirk, J. and Miller, M. L. (1986) *Reliability and validity in qualitative research: qualitative research methods*. 1 edition. SAGE Publications.
- Kitwood, T. M. (1997) *Dementia reconsidered: The person comes first*. Buckingham: Open University Press.
- Kitzinger, J. (1995) Introducing focus groups. *British Medical Journal* 311, 299-302.
- Klein, H. K. and Myers, M. D. (1999) A set of principles for conducting and evaluating interpretive field studies in information systems. *MIS Q.* 23 (1), 67-93. DOI=<http://dx.doi.org/10.2307/249410>
- Knapp, M., Comas-Herrera, A., Somani, A. and Banerjee, S. (2007) *Dementia: international comparisons, Summary report for the National Audit Office*. Personal Social Services Research Unit. <http://www.pssru.ac.uk/pdf/dp2418.pdf> Accessed 13/11/2012.
- Kossarova, L., Blunt, I. and Bardsley, M. (2015) *Focus on: international comparisons of healthcare quality, what can the UK learn?* Quality Watch.
- Krauss, S. E. (2005) Research paradigms and meaning making: a primer. *The Qualitative Report* 10 (4), 758-770.
- Kritsotakis, E. I., Dimitriadis, I., Roubelaki, M., Vounou, E., Kontou, M., Papakyriakou, P., Koliou-Mazeri, M., Varthalitis, I., Vrouchos, G., Troulakis, G. and Gikas, A. (2008) Case-mix adjustment approach to benchmarking prevalence rates of nosocomial infection in hospitals in Cyprus and Greece. *Infection Control & Hospital Epidemiology* 29 (08), 685-692.
- Kroeger, T. A., Davidson, N. J. and Cook, S. C. (2014) Understanding the characteristics of quality for software engineering processes: a grounded theory investigation. *Information and Software Technology* 56 (2), 252-271.
- Kuhn, D., Edelmann, P. and Fulton, B. R. (2005) Daytime sleep and the threat to well-being of persons with dementia. *The International Journal of Social Research and Practice* 2 (4), 233-247.
- Kuhn, D., Fulton, B. R. and Edelmann, P. (2004) Factors influencing participation in activities in dementia care settings. *Alzheimer's Care Quarterly* 3 (5), 144-152.
- Kuhn, D., Kasayka, R. E. and Lechner, C. (2002) Behavioral observations and quality of life among persons with dementia in 10 assisted living facilities. *American Journal of Alzheimer's Disease and Other Dementias* 17 (5), 291-298.
- Kujala, S., Kauppinen, M. and Rekola, S. (2001) Bridging the gap between user needs and user requirements. *Proceedings of the Panhellenic*

- Conference with International Participation in Human Computer Interaction PC-HCI 2001*. Typorama Publications.
- Kumar, V. and Thareja, R. (2013) A simplified approach for quality management in data warehouse. *International Journal of Data Mining and Knowledge Management Process (IJDMP)* 3 (5), 61-69.
- Kuo, A. M. (2011) Opportunities and challenges of cloud computing to improve health care services. *Journal of Medical Internet Research* 13 (3).
- Lai, C. K., Chi, I. and Kayser-Jones, J. S. (2004) A randomised controlled trial of a specific reminiscence approach to promote well-being of nursing home residents with dementia. *International Psychogeriatric* 16 (1), 33-49.
- Lamas, E., Barh, A., Brown, D. and Jaulent, M. C. (2015) Ethical, legal and social issues related to the health data-warehouses: re-using health data in the research and public health research. *Studies in Health technology and Informatics*, 2010, 719-23.
- Lankton, N., McKnight, D. H. and Thatcher, J. B. (2014) Incorporating trust-in-technology into Expectation Disconfirmation Theory. *The Journal of Strategic Information Systems* 23 (2), 128-145.
- Law, M. (2005) Reduce, reuse, recycle: Issues in the secondary use of research data. *IASSIST Quarterly* 29 (1), 5-10.
- Lincoln, Y. S. and Guba, E. G. (1985) *Naturalistic inquiry*. Newbury Park, CA: Sage Publications.
- Linder, J. A., Kaleba, E. and Kmetik, K. S. (2009) Using electronic health records to measure physician performance for acute conditions in primary care: empirical evaluation of the community-acquired pneumonia clinical quality measure set. *Medical Care* 47 (2), 208-216.
- Lindgaard, G., Dillon, R., Trbovich, P., White, R., Fernandes, G., Lundahl, S. and Pinnamaneni, A. (2006) User needs analysis and requirements engineering: theory and practice. *Interacting with Computers* 18 (1), 47-70.
- Lintern, T., Woods, R. T. and Phain, L. (2002) *Before and after training: a case study of intervention*. In *Dementia Topics for the Millennium and Beyond*. London: Hawker Publications.
- List, B., Bruckner, M. R., Machaczek, K. and Schiefer, J. (2002) A comparison of data warehouse development methodologies: case study of the process warehouse. In *DEXA 2002*. Springer-Verlag Berlin Heidelberg. 203-215. Accessed: 09/09/2012.
- Lluch, M. (2011) Healthcare professionals' organisational barriers to health information technologies—a literature review. *International Journal of Medical Informatics* 80 (12), 849-862.
- Lo Iacono, L. (2007) Multi-centric universal pseudonymisation for secondary use of the EHR. *Studies in Health Technology and Informatics* 126 (239-47), 239.
- Lovaglio, G. P. (2012) Benchmarking strategies for measuring the quality of healthcare: problems and prospects. *The Scientific World Journal* 2012, 1-13.
- Lowe, A. (1995) The basic processes of entrepreneurial innovation. *The International Journal of Entrepreneurial Behaviour and Research* 1 (2),

54-76.

- Lyman, J. A., Scully, K. and Harrison Jr, J. H. (2008) The development of health care data warehouses to support data mining. *Clinics in Laboratory Medicine* 28 (1), 55-71.
- Maguire, M. and Nigel, B. (2002) User requirements analysis: a review of supporting methods. *Proceedings of IFIP 17th World Computer Congress*. Montreal, Canada. Kluwer Academic Publishers. http://nigelbevan.com/papers/WCC_UserRequirements.pdf. Accessed 12/07/12.
- Malterud, K. (2001) Qualitative research: standards, challenges, and guidelines. *Lancet* 358 (9280), 483-488.
- Martikainen, S., Kopela, M. and Tihonen, T. (2014) User participation in healthcare IT development: a developer's view point in Finland. *International Journal of Medical Informatics* 83 (3), 189-200.
- Martin, J. L., Clark, D. J., Morgan, S. P., Crowe, J. A. and Murphy, E. (2012) A user-centred approach to requirements elicitation in medical device development: a case study from an industry perspective. *Applied Ergonomics* 43 (1), 184-190.
- Mavetera, N. and Kroeze, J. H. (2009) *Practical considerations in grounded theory research*. Sprouts: Working papers on Information Systems: <http://sprouts.aisnet.org/9-32> Accessed 09/10/2011.
- Mavin, A. and Maiden, N. (2003) Determining socio-technical system requirements: experiences with generating and walking through scenarios. *11th International Conference on Requirement Engineering*. IEEE CS Press.
- Mazón, J.-N., Trujillo, J., Serrano, M. and Piattini, M. (2005) Designing data warehouses: from business requirement analysis to multidimensional modeling. *Requirements Engineering for Business need and IT alignment: REBNITA* 5, 44-53.
- Medical Research Council (2011) *UK e-health records research capacity and capability*. <http://www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC007896> Accessed 12/11/2012.
- Meehan, T. J., Stedman, K. E., Neuendorf, I. D., Francisco, M. G. and Neilson, M. G. (2007) Benchmarking Australia's mental health services: is it possible and useful? *Australian Health Review: A Publication of the Australian Hospital Association* 31 (4), 623-7.
- Meissner, W., Ullrich, K. and Zwacka, S. (2006) Benchmarking as a tool of continuous quality improvement in postoperative pain management. *European Journal of Anaesthesiology* 23 (02), 142-148.
- Mental Health Foundation (2012) *Dementia*. <http://www.mentalhealth.org.uk/help-information/mental-health-a-z/D/dementia/> Accessed 20/11/2011.
- Meystre, S., Friedlin, F., South, B., Shen, S. and Samore, M. (2010) Automatic de-identification of textual documents in the electronic health record: a review of recent research. *BMC Med Res Methodol* 10, 70.
- Miles, M. B. and Huberman, A. M. (1994) *Qualitative data analysis: An expanded sourcebook*. 2nd edition. Thousand Oaks ; London: Sage Publications.

- Moges, H.-T., Vlasselaer, V. V., Lemahieu, W. and Baesens, B. (2016) Determining the use of data quality metadata (DQM) for decision making purposes and its impact on decision outcomes — an exploratory study. *Decision Support Systems* 83, 32-46.
- Mohammed, R. O. and Talab, S. A. (2014) Clinical data warehouse issues and challenges. *International Journal of u-and e-Service, Science and Technology* 7 (5), 251-262.
- Monks, R. (2006) *A grounded theory study of the experiences of illicit drug users and nurses caring for them on medical wards: the experiences of nurses and illicit drug users of the care management within medical wards and the Medical Admission Unit in a large hospital trust*. Bradford:
- Moody, L. D. and Kortink, A. R. M. (2000) From enterprise models to dimensional models: a methodology for data warehouse and data mart design. *Proceedings of the International workshop on design and management of data warehouse (DMDW'2000)*. Stockholm, Sweden, Vol. 28. http://ssdi.di.fct.unl.pt/bddw/material_apoio/artigos/files/2000-Moody.pdf. Accessed 03/10/2012.
- Mosel, D. and Gift, B. (1994) Collaborative benchmarking in health care. *The Joint Commission Journal on Quality improvement* 20 (5), 239-249.
- Mundy, J., Thornthwaite, W. and Kimball, R. (2006) *The Microsoft data warehouse toolkit: with SQL Server 2005 and the Microsoft business intelligence toolkit*. Indianapolis: Wiley Publishing, Inc.
- Myers, D. M. (1997) *Qualitative research in information systems*. Association for Information Systems. <http://www.qual.auckland.ac.nz/> Accessed 23/03/2012.
- Nasiri, A., Wrembel, R. and Zimányi, E. (2015) Model-based requirements engineering for data warehouses: from multidimensional modelling to KPI monitoring. *Advances in Conceptual Modeling*. Springer. 198-209
- National Audit Office (2007) *Improving services and support for people with dementia*. London: The Stationary Office: http://www.nao.org.uk/publications/0607/support_for_people_with_dementia.aspx Accessed 13/11/2012.
- National Audit Office (2010) *Improving dementia services in England: an interim report*. London:
- National Information Board (2014) *Personalised health and care 2020: using data and technology to transform outcomes for patients and citizens-a framework for action*. London, UK: NHS, National Information Board. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/384650/NIB_Report.pdf
- National Institute of Health Research (2006) *The research capability program (RCP)*. National Institute of Health Research. http://www.nihr.ac.uk/systems/Pages/Research_Capability_Programme.aspx Accessed 06/12/2011.
- Neupert, P. (2009) White Paper: Investing in health IT: a stimulus for a healthier Canada. *Healthcare Quarterly (Toronto, Ont.)* 12 Spec No Ontario, 91-95.
- NHS (2015) *Does my project require review by a research ethics committee*. England, UK: National Research Ethics Service.

- <http://www.hra.nhs.uk/resources/before-you-apply/is-nhs-rec-review-required/> Accessed 11/2011.
- NICE/SCIE (2007) *The National Institute for Health and Clinical Excellence-Social Care Institute for Excellence guideline on supporting people with dementia and their carers in health and social carer*. 42. England: The British Psychological Society and Gaskell. <http://www.nice.org.uk/nicemedia/live/10998/30320/30320.pdf> Accessed 12/12/2011.
- Nieboer, M. E., van Hoof, J., van Hout, A. M., Aarts, S. and Wouters, E. J. M. (2014) Professional values, technology and future health care: the view of health care professionals in The Netherlands. *Technology in Society* 39 (0), 10-17.
- Niès, J. and Pelayo, S. (2010) From users involvement to users' needs understanding: a case study. *International Journal of Medical Informatics* 79 (4), 76-82.
- Noel, P. H., Copeland, L. A., A, P. R., Lancaster, A. E., Pugh, M. J., Wang, C.-P. and Bollinger, M. J. (2010) VHA corporate data warehouse height and weight data: opportunities and challenges for health services research. *Journal of Rehabilitation Research and Development* 47 (8), 739-50.
- Nolte, E. (2010) *International benchmarking of healthcare quality: a review of the literature*. California, US: RAND Corporation.
- Orlikowski, W. J. and Baroudi, J. J. (1991) Studying information technology in organizations: Research approaches and assumptions. *Information Systems Research* 2, 1-28.
- Pace, S. (2004) A grounded theory of the flow experiences of web users. *International Journal of Human-Computer Studies* 60, 327-363.
- Packer, T. and Jeffries, M. (1997) Halopriedol, hips and toenails..... *Journal of Dementia Care* 5 (Nov/Dec), 22-24.
- Paim, F. R. S. and de Castro, J. F. B. (2003) DWARF: an approach for requirements definition and management of data warehouse systems. *Requirements Engineering Conference, 2003. Proceedings. 11th IEEE International*. 8-12 Sept. 2003.
- Parmanto, B., Scotch, M. and Ahmad, S. (2005) A framework for designing a healthcare outcome data warehouse. *Perspectives in Health Information Management / AHIMA, American Health Information Management Association* 2, 3.
- Patton, M. Q. (1987) *How to use qualitative methods in evaluation*. Program evaluation kit. Newbury Park; London: Sage.
- Patton, M. Q. (2002) *Qualitative research and evaluation methods*. 3rd edition. Thousand Oaks, Calif.: Sage Publications.
- Pearson, N., O'Brian, J., Thomas, H., Ewings, P., Gallier, L. and Bussey, A. (1996) Collecting morbidity data in general practice: the Somerset morbidity project. *British Medical Journal* 312, 1517-20.
- Personal Communications (2013) *DCM training*.
- Pidgeon, F. N., Turner, A. B. and Blockely, I. D. (1991) The use of grounded theory for conceptual analysis in knowledge elicitation. *International Journal of Man-Machine Studies* 35 (151-173), 151.

- Pitts, M. G. and Browne, G. J. (2007) Improving requirements elicitation: an empirical investigation of procedural prompts. *Information Systems Journal* 17 (1), 89-110.
- Polit, F. D., Beck, C. and Hungler, P. B. (2001) *Essentials of nursing research: methods, appraisal, and utilization*. 5th edition. Philadelphia, PA: Lippincott.
- Ponniah, P. (2001) *Data warehousing fundamentals: a comprehensive guide for IT professionals*. John Wiley and Sons, Inc.
- Post, A. R., Kurc, T., Cholleti, S., Gao, J., Lin, X., Bornstein, W., Cantrell, D., Levine, D., Hohmann, S. and Saltz, J. H. (2013) The analytic information warehouse (AIW): a platform for analytics using electronic health record data. *Journal of Biomedical Informatics* 46 (3), 410-424.
- Powell, A. E., Davies, H. T. O. and Thomson, R. G. (2003) Using routine comparative data to assess the quality of health care: understanding and avoiding common pitfalls. *Quality and Safety in Health Care* 12 (2), 122-128.
- Power, N. and Moynihan, T. (2003) A theory of requirements documentation situated in practice. In *Proceedings of the 21st annual international conference on Documentation (SIGDOC '03)*. ACM, New York, NY, USA. <http://doi.acm.org/10.1145/944868.944887>.
- Presser, L., Kruskova, M., Rowbottom, H. and Kancir, J. (2015) *Care.data and access to UK health records: patient privacy and public trust*. Technology Science. <http://techscience.org/a/2015081103/> Accessed 11/11.
- Prince, M., Wimo, A., Guerchet, M., Wu, Y. and Prina, M. (2015) *The global impact of dementia; an analysis of prevalence, incidence, cost and trends*. Alzheimer Disease International.
- Raab, L. T. (1998) *End user requirements gathering - Information Management Magazine Article*. <http://www.information-management.com/issues/19980901/432-1.html?zkPrintable=true> Accessed 01/03/2012.
- Raghupati, W. and Tan, J. (2002) Strategic IT applications in health care. *Communications of the ACM* 45 (12), 56-61.
- Raleigh, S. V. (2010) *Getting the measure of quality: Opportunities and challenges*. London, UK: The Kings Fund.
- Redman, T. C. (1998) The impact of poor data quality on the typical enterprise. *Communications of the Association for Computing Machinery (ACM)*. New York, 41(2), 79-82.
- ResearchOne (2012) *Research tools for free text anonymisation*. <http://www.researchone.org/research-tools-for-free-text-anonymisation/> Accessed 02/02.
- Richards, L., Johnston, L. H., Di Gregorio, S. and Bazeley, P. (2004) Evaluating the performance of NVivo users: what demonstrates skill with NVivo and why is it important to do so? *Strategies in Qualitative Research*. University of Durham, UK.
- Romero, O. and Abelló, A. (2010) A framework for multidimensional design of data warehouses from ontologies. *Data and Knowledge Engineering* 69 (11), 1138-1157.

- Rutty, J. E. (2010) *The meaning of being as a nurse involved in the work of death investigation: a North American view and its implications to practice in England*. PhD. Bradford, UK: University of Bradford.
- Safran, C., Bloomrosen, M., Hammond, W. E., Labkoff, S., Markel-Fox, S., Tang, P. C. and Detmer, D. E. (2007) Toward a national framework for the secondary use of Health Data: an American Medical Informatics Association White Paper. *Journal of the American Medical Informatics Association* 14 (1), 1-9.
- Sahama, R. T. and Croll, R. P. (2007) A data warehouse architecture for clinical data warehousing. *Research and Practice in Information Technology*. Ballarat, Australia, Vol. 68. Australian Computer Society Inc.
- Sandelowski, M. (1995) Sample size in qualitative research. *Research in Nursing and Health* 18 (179-183), 79.
- Sandra, C. and Garmon, B. (2007) Issues associated with secondary analysis of population health data. *Applied Nursing Research* 20 (2), 94-99.
- Sartipi, K., Yarmand, M. H. and Down, D. G. (2007) Mined knowledge and decision support services in electronic health. *International Conference of Software Engineering*. Proceedings of the International Workshop of Systems Developments in SOA Environment.
- Scandurra, I., Hägglund, M. and Koch, S. (2008) From user needs to system specifications: multi-disciplinary thematic seminars as a collaborative design method for development of health information systems. *Journal of Biomedical Informatics* 41 (4), 557-569.
- Schaefer, B. C., Tanrikulu, E. and Breiter, A. (2011) Eliciting user requirements when there is no organization: a mixed method for an educational data warehouse project. *Procedia - Social and Behavioral Sciences* 28 (0), 743-748.
- Schiefer, J., List, B. and Bruckner, R. (2002) A holistic approach for managing requirements of data warehouse systems. *AMCIS 2002 Proceedings*.
- Schuller, T., Wadsworth, M., Bynner, J. and Goldstein, H. (2012) *The measurement of well-being: the contribution of longitudinal studies*. UK Office for National Statistics (ONS).
- Schwandt, T. A. (1994) Constructivist, interpretivist approaches to human inquiry. In Denzin, N. K. and Lincoln, Y. S. (editors) *Handbook of Qualitative research*. Thousand Oaks: Sage Publications. 118-137
- Scobie, S., Basnett, I. and McCartney, P. (1995) Can general practice data be used for needs assessment and health care planning in an inner-London district? *Journal of Public Health* 17 (4), 475-483.
- Scott, W., Karen (2008) Clarifying analysis and interpretation in grounded theory: using a conditional relationship guide and reflective coding matrix. *International Journal of Qualitative Methods* 7 (2), 1-15.
- Seale, C. (1999) Quality in qualitative research. *Qualitative Inquiry* 5(4), 465-478. <http://qix.sagepub.com/content/5/4/465> Accessed 23/03/2012
- Sen, A. (2004) Metadata management: Past, present and future. *Decision Support Systems* 37 (1), 151-173.
- Servelec health and social care (2016) *Health and Social Care RiO*. England, UK:

<http://www.servelec-group.com/health-social-care/healthcare/products/rio/> Accessed 04/02/2016.

- Seyff, N. and Graf, F., 2010. User-Driven Requirements Engineering for Mobile Social Software. In *Software Engineering (Workshops)*, 503-512.
- Shahbaz UI Haq, Q. (2016) Chapter 4 - Data models. *Data Mapping for Data Warehouse Design*. Boston: Morgan Kaufmann. 7-16
<http://www.sciencedirect.com/science/article/pii/B9780128051856000046>
- Shanks, G. and Corbitt, B. (1999) Understanding data quality: social and cultural aspects. *10th Australian Conference on Information Systems*.
- Shanks, G. and Darke, P. (1998) Understanding data quality in data warehousing: a semiotic approach. In *MIT conference on Information Quality*. Boston. 247-264.
- Sheikh, A., Cornford, T., Barber, N., Avery, A., Takian, A., Lichtner, V., Petrakaki, D., Crowe, S., Marsden, K., Robertson, A., Morrison, Z., Klecun, E., Prescott, R., Quinn, C., Jani, Y., Ficociello, M., Voutsina, K., Paton, J., Fernando, B., Jacklin, A. and Cresswell, K. (2011) Implementation and adoption of nationwide electronic health records in secondary care in England: final qualitative results from prospective national evaluation in "early adopter" hospitals. *BMJ* 343.
- Shenton, A. K. (2004) *Strategies for ensuring trustworthiness in qualitative research projects* IOS Press.
- Sheta, E. O. and Eldeen, N. A. (2013) The technology of using a data warehouse to support decision-making in healthcare. *International Journal of Database Management Systems (IJDMS)* 5 (3), 75-86.
- SHIP (2012) *The collation, management, dissemination and research analysis of anonymised Electronic Patient Records*. Scottish Health Informatics Program. <http://www.scot-ship.ac.uk/> Accessed 03/12/2011.
- Silver, M., Sakata T., Su, H. C., Herman, C., Dolins, S. B., and O'Shea, M. J. (2001) Case study: how to apply data mining techniques in a healthcare data warehouse. *Journal of Healthcare Information Management (JHIM)* 15 (2), 155-164.
- Simmhan, L. Y., Plale, B. and Gannon, D. (2005) A survey of data provenance in e-science. *SIGMOD Records* 34 (3), 31-36.
- Singh, R. and Singh, K. (2010) A descriptive classification of causes of data quality problems in data warehousing. *International Journal of Computer Science Issues* 7 (3), 41-50.
- Skoutas, D. and Simitsis, A. (2007) Ontology-based conceptual design of ETL processes for both structured and semi-structured data. *International Journal on Semantic Web and Information Systems (JSWIS)* 3 (4).
- Sloane, P. D., Brooker, D., Cohen, L., Douglass, C., Edelman, P., Fulton, B. R., Jarrott, S., Kasayka, R., Kuhn, D., Preisser, J. S., Williams, C. S. and Zimmerman, S. (2007) Dementia care mapping as a research tool. *International Journal of Geriatric Psychiatry* 22 (6), 580-589.
- Smith, R., Xu, J., Hima, S. and Johnson, O. (2013) GATEway to the cloud, a case study: a privacy-aware environment for Electronic Health Records research. *IEEE 7th International Symposium on Service Oriented System Engineering (SOSE)*. IEEE.

- Song, I.Y., Khare, R. and Dai, B., 2007, November. SAMSTAR: a semi-automated lexical method for generating star schemas from an entity-relationship diagram. In *Proceedings of the ACM tenth international workshop on Data warehousing and OLAP* (pp. 9-16). ACM. <http://dl.acm.org/citation.cfm?id=1317334>
- Stolba, N. (2007) *Towards a sustainable data warehouse approach for evidence-based healthcare*. Ph.D. Vienna: Vienna University of Technology.
- Stolba, N., Banek, M. and Tjoa, A. M. (2006) The security issue of federated data warehouses in the area of evidence-based medicine. In *Proceedings of the First International Conference on Availability, Reliability and Security (ARES '06)*. Washington, DC, USA. IEEE Computer Society.
- Strauss, A. L. (1978) *Negotiations : varieties, contexts, processes, and social order*. The Jossey-Bass social and behavioral science series. San Francisco; London: Jossey-Bass.
- Strauss, A. L. (1987) *Qualitative analysis for social scientists*. Cambridge: Cambridge University Press.
- Strauss, A. L. and Corbin, J. M. (1990) *Basics of qualitative research : Grounded theory procedures and techniques*. Newbury Park, CA.;London: Sage.
- Strauss, A. L. and Corbin, J. M. (1998) *Basics of qualitative research: techniques and procedures for developing grounded theory*. 2nd edition. Thousand Oaks, Calif. ; London: Sage.
- Surr, C., Hamilton, J., Firth, P., Jones, S., Negu, D., Khalid, S., Jones, S., Tammer, Q. and Fox, S. (2015) Monitoring and improvement using dementia care mapping. *Journal of Dementia Care* 23 (5), 16-17.
- Sutcliffe, A., Fickas, S. and Sohlberg, M. M. (2006) PC-RE: a method for personal and contextual requirements engineering with some experience. *Requirement Engineering*. Vol. 11. Springer-Verlag London Limited.
- SystemOne (2016) *What is systemOne?* UK: <http://www.tpp-uk.com/products/systmone> Accessed 02/02/2016.
- Tablan, V., Damlijanovic, D. and Bontchena, K. (2008) A natural language query interface to structured information. *5th European Semantic Web Conference (ESWC)*. Tenerief, Spain, Vol. 5021. Springer-Verlage Inc.
- Takecian, P. L., Oikawa, M. K., Braghetto, K. R., Rocha, P., Lucena, F., Kavounis, K., Schlumpf, K. S., Acker, S., Carneiro-Proietti, A. B. F., Sabino, E. C., Custer, B., Busch, M. P. and Ferreira, J. E. (2013) Methodological guidelines for reducing the complexity of data warehouse development for transactional blood bank systems. *Decision Support Systems* (0).
- Teixeira, L., Ferreira, C. and Santos, B. S. (2012) User-centered requirements engineering in health information systems: a study in the hemophilia field. *Computer Methods and Programs in Biomedicine* 106 (3), 160-174.
- Tejay, G., Dhillon, G. and Chin, A. (2006) Data quality dimensions for information systems security: a theoretical exposition (Invited Paper).

In Dowland, P. A. F., Steve and Thuraisingham, Bhavani and Wang, X. Sean (editor) *Security Management, Integrity, and Internal Control in*

- Information Systems*. IFIP International Federation for Information Processing. Vol. 193. Springer US. 21-39
http://dx.doi.org/10.1007/0-387-31167-X_2
- Thakur, G. and Gosain, A., 2011. DWEVOLVE: a requirement based framework for data warehouse evolution. *ACM SIGSOFT Software Engineering Notes*, 36(6), 1-8.
- The National Archives (2001) *Health and social care act 2001*. England, UK: <http://www.legislation.gov.uk/ukpga/2001/15/section/60> Accessed 15/01.
- Thornton, A., Hatton, C. and Tatham, A. (2004) Dementia Care Mapping reconsidered: exploring the reliability and validity of the observational tool. *International Journal of Geriatric Psychiatry* 19 (8), 718-726.
- Tung, F.-C., Chang, S.-C. and Chou, C.-M. (2008) An extension of trust and TAM model with IDT in the adoption of the electronic logistics information system in HIS in the medical industry. *International Journal of Medical Informatics* 77 (5), 324-335.
- UK Biobank (2007a) *UK Biobank ethics and governance framework*. <https://www.ukbiobank.ac.uk/wp-content/uploads/2011/05/EGF20082.pdf> Accessed 12/09/2015.
- UK Biobank (2007b) *UK Biobank: protocol for a large-scale prospective epidemiological resource*. Accessed 12/02/2014 UK Biobank (2007) *UK Biobank: protocol for a large-scale prospective epidemiological resource*. Accessed 12/02/2014 <http://www.ukbiobank.ac.uk/wp-content/uploads/2011/11/UK-Biobank-Protocol.pdf>
- UK Biobank (2012) *UK Biobank/ data management and sharing plan*. England, UK. <http://www.ukbiobank.ac.uk/wp-content/uploads/2013/10/ukbiobank-data-management.pdf> Accessed 14/09/2015.
- Urquhart, C. (2000) *Strategies for conversation and systems analysis in requirements gathering: a qualitative view of analyst-client communication*. The Qualitative Report 4 (1), 1-19.
- Uzuner, O., Luo, Y. and Szolovits, P. (2007) Evaluating the state-of-the-art in automatic de-identification. *Journal of the American Medical Informatics Association* 14 (5), 550-563.
- van Teijlingen, E. R. and Hundley, V. (2001) *The importance of pilot studies*. University of Surrey. Social Research Update 35.
- Verma, D., Tyagi, A. and Sharma, D. (2014) Data quality problems in data warehousing. *International Journal of Innovative Research in Technology* 1 (7), 265-269.
- Wager, K. A., Lee, F. W. and Glaser, J. p. (2013) *Managing health care information systems; a practical approach for health care management*. Third edition. Jossey-Bass,.
- Wand, Y. and Wang, R., Y (1996) *Anchoring data quality dimensions in ontological foundations*. New York, Association for Computing Machinery.
- Wang, F. (2014) Explaining the low utilization of government websites: using a grounded theory approach. *Government Information Quarterly* 31 (4), 610-621.

- Wang, R. Y. and Strong, D. M. (1996) Beyond accuracy: what data quality means to data consumers. *Journal of Management of Information Systems* 12 (4), 5-33.
- Weiskopf, N. G., Hripcsak, G., Swaminathan, S. and Weng, C. (2013) Defining and measuring completeness of electronic health records for secondary use. *Journal of Biomedical Informatics* 46 (5), 830-836.
- Weiskopf, N. G. and Weng, C. (2012) Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *Journal of American Medical Informatics Association* 20 (1), 144-51.
- Westbrook, J. L., McIntosh, C. J., Sheldrick, R., Surr, C. and Hare, D. J. (2013) Validity of Dementia Care Mapping on a neuro-rehabilitation ward: Q-methodology with staff and patients. *Disability and Rehabilitation* 35 (19), 1652-1659.
- Wiesenauer, M., Johnner, C. and Rohrig, R. (2012) Secondary use of clinical data in healthcare providers - an overview on research, regulatory and ethical requirements. *Studies in Health Technology and Informatics* 180, 614-618.
- Willemse, M. B., Smit, D., De Lange, J. and Pot, M. A. (2011) Nursing home care with people with dementia and residents' quality of life, quality of care and staff well-being: design of the living arrangements for people with dementia (LAD)-study. *BMC Geriatrics* 11 (11), 1-7.
- Winter, R. and Strauch, B., 2003, January. A method for demand-driven information requirements analysis in data warehousing projects. In *System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference* (pp. 9). IEEE. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1174602
- Wisniewski, M. F., Kieszowski, P., Zagorski, B. M., Trick, W. E., Sommers, M., Weinstein, R. A. and for the Chicago Antimicrobial Resistance, P. (2003) Development of a clinical data warehouse for hospital infection control. *Journal of the American Medical Informatics Association* 10 (5), 454-462.
- Woods, B. and Lintern, T. (2003) The reliability and validity of dementia care mapping. In Innes, A. (editor) *Dementia Care Mapping; applications across cultures*. Baltimore, Maryland: Health profession Press, Inc. 25-38
- Woolley, R. J., Young, J. B., Green, J. R. and Brooker, D. J. (2008) The feasibility of care mapping to improve care for physically ill older people in hospital. *Age and Ageing* 37 (4), 390-395.
- World Health Organization and Alzheimer's Disease International (2012) *Dementia: a public health priority*. http://www.who.int/mental_health/publications/dementia_report_2012/en/index.html Accessed 02/12/2012.
- Wu, I.-L. and Chen, J.-L. (2005) An extension of trust and TAM model with TPB in the initial adoption of on-line tax: an empirical study. *International Journal of Human-Computer Studies* 62 (6), 784-808.
- Yang, H. and Chen, Y.-P. P. (2015) Data mining in lung cancer pathologic staging diagnosis: Correlation between clinical and pathology information. *Expert Systems with Applications* 42 (15–16), 6168-6176.

- Yang, J.-J., Li, J., Mulder, J., Wang, Y., Chen, S., Wu, H., Wang, Q. and Pan, H. (2015) Emerging information technologies for enhanced healthcare. *Computers in Industry* 69 (0), 3-11.
- Yin, R. K. (2003) *Case study research: design and methods*. 3rd edition. Applied social research methods series. Newbury Park; London: Sage.
- Yin, R. K. (2009) *Case study research: design and methods*. 4th edition. Applied social research methods series. London: Sage.
- Younger, D. and Martin, G. W. (2000) Dementia Care Mapping: an approach to quality audit of services for people with dementia in two health districts. *Journal of Advanced Nursing* 32 (5), 1206-1212.
- Yusof, M. M. (2015) A case study evaluation of a critical care information system adoption using the socio-technical and fit approach. *International Journal of Medical Informatics* 84(7), 486-499.
- Zhang, N., Rector, A., Buchan, I., Shi, Q., Kalra, D., Rogers, J., Goble, C., Walker, S., Ingram, D. and Singleton, P. (2005) A linkable identity privacy algorithm for HealthGrid. *Studies in Health Technology and Informatics* 112, 234-45.
- Zhang, Z. (2007) Effective requirements development -- a comparison of requirements elicitation techniques. In *Software Quality in the Knowledge Society*. British Computer Society.
- Zhou, X., Chen, S., Liu, B., Zhang, R., Wang, Y., Li, P., Guo, Y., Zhang, H., Gao, Z. and Yan, X. (2010) Development of traditional Chinese medicine clinical data warehouse for medical knowledge discovery and decision support. *Artificial Intelligence in Medicine* 48 (2-3), 139-152.
- Zimmerman, S., Sloane, P. D., Williams, C. S., Reed, P. S., Preisser, J. S., Eckert, J. K., Boustani, M. and Dobbs, D. (2005) Dementia care and quality of life in assisted living and nursing homes. *The Gerontologist* 45 (suppl 1), 133-146.
- Zowghi, D. and Coulin, C. (2005) *Requirements elicitation: a survey of techniques, approaches and tools*. In *Engineering and Managing Software Requirements*. Springer Berlin Heidelberg, 19-46

Appendix 1: Consent form (general)



Shehla Khalid
School of Health Studies
Horton A, 4th Floor
Great Horton Road
Bradford BD7 1DP
T: 01274 236399
M: 07751800475
E: s.khalid3@student.bradford.ac.uk

Consent form

Title: A data management framework for Dementia Care Mapping (DCM) using a data warehouse approach to improve the quality of dementia care

Name of Investigator: Shehla Khalid

Please initial the boxes

1. I confirm that I have read and understood the information sheet for the above study and have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.

2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my legal rights being affected.

3. I understand that the conversation/ interviews with the researcher will be audio tape recorded. I give permission to the researcher to use an audio recorder and transcribe the data. I am sure the recorded data and the transcriptions will be stored in a safe place and destroyed after 5 years.

4. I understand that anonymised quotes from the interview may be used in the researcher's thesis and subsequent publications.

5. I understand that my personal details and other identifiable information will not be used in any documentation or publications

6. I agree to take part in the above study.

Name of participant

Date

Signature

Name of researcher

Date

Signature

When completed: 1 copy for participant; 1 copy for researcher/ site file.

Appendix 2: Consent to take part in a focus group

Shehla Khalid
School of Health Studies
Horton A, 4th Floor
Great Horton Road
Bradford BD7 1DP
T: 01274 236399
M: 07751800475
E: s.khalid3@student.bradford.ac.uk



Consent to take part in focus group

Title: A data management framework for Dementia Care Mapping (DCM) using a data warehouse approach to improve the quality of dementia care

Name of Investigator: Shehla Khalid

Please initial the box

1. I agree to take part in the focus group, arranged within my organisation.
2. I understand that the focus group conversation with the researcher and other participants from my organisation will be audio recorded.
3. I give permission for the researcher to use an audio recorder and transcribe the data.
4. I understand anonymised quotes from the focus group may be used in the researcher's thesis and subsequent publications.
5. I understand that the recorded data and the transcriptions will be stored in a safe place and destroyed after 5 years.
6. I understand I am free to withdraw from the study at any time

Name of participant

Date

Signature

Name of researcher

Date

Signature

When completed: 1 copy for participant; 1 copy for researcher/ site file.

Appendix 3: Information sheet for health and social care organisations

Shehla Khalid
School of Health Studies
Horton A, 4th Floor
Great Horton Road
Bradford BD7 1DP
T: 01274 236399
M: 07751800475
E: s.khalid3@student.bradford.ac.uk



Information Sheet: members from health and social care organisations

Study Title: A data management framework for Dementia Care Mapping (DCM) using a data warehouse approach to improve the quality of dementia care

I would like to invite you to take part in my research study. Before you decide whether or not to take part, it is important to understand why the research is being conducted, what it will involve and what your role will be as a participant. Please take time to read the following information carefully.

What is the project all about?

Dementia Care Mapping (DCM) is an observational process and a practice development tool to improve the quality of care and quality of life of people with dementia within formal dementia care settings.

Each year DCM users produce large amounts of data. However, the use of such important, rich data is only limited at a local level, which means the individual mappers or organisations collecting the data uses it within their own settings or research projects, for the purpose it was originally collected. The DCM data collected from various mappers and organisations can be re-analysed for secondary purposes for producing new knowledge that can inform future improvements in the DCM method/tool and provide suggestions for providing quality dementia care within formal dementia-care settings. However, one of the potential reasons of the lack of data sharing among DCM users is the lack of IT system for such purposes. Current electronic data analysis and storage systems have limited functionality and do not permit efficient and integrated storage and analysis of data over time and its dissemination to various users.

This study aims to design a data repository, called a data warehouse, for managing the DCM data for future uses. Within this data warehouse the DCM data will be arranged/presented according to the users' requirements and needs of data storage and retrieval for various analytical, comparison, benchmarking and research purposes. The user-identified structure of the data warehouse will inform the effective, efficient, integrated and systematic solution of managing the DCM data by variety of DCM users from health care organisations, social care organisations, care monitoring/regulation organisations, charities and research backgrounds.

What is the purpose of the study?

This study aims to gather the potential users' requirements for storing and accessing DCM data within a data warehouse. The intentions are to design a user-identified structure of a data warehouse informing a data management framework for DCM to enable variety of DCM users/stakeholders to have a structured and systematic way of accessing data for multidimensional analysis, benchmarking, pattern recognition, monitoring care quality and DCM research.

What will the study involve?

This study will involve conducting semi-structured interviews or focus groups with the staff working in key roles (care delivery, management or decision making) within the health care/social care organisation, care monitoring/improvement/inspection organisations, charities, statutory bodies, researchers and DCM expert users, who are using DCM for various purposes or have intentions to use it in the future. The researcher will ask the participants the relevant questions (using semi-structured interview techniques) to explore their experience of using DCM data for analysis purposes and their requirements to use it for further retrieval purposes

Who is the researcher?

This project is being carried out by a postgraduate researcher from the University of Bradford (Shehla Khalid) under the supervision of Dr Claire Surr (Bradford Dementia Group, University of Bradford) and Dr Daniel Neagu (School of Computing, University of Bradford).

What will you as a participant have to do?

If you agree to take part in the study, you will be contacted by the researcher to arrange a face-to-face interview or a focus group (if there are more than four participants from your organisation agree to take part in the study) at your convenient date or time. During the

interview or focus group, you will be asked about your experience of using DCM data for analysis purposes within your organisation and your future requirements of analysing the

DCM data for other purposes. During the interview or focus group, an informal conversation will take place between you and the researcher which will take maximum 30-45 minutes of your time.

When will the research take place?

The research will start in March 2012 and I would like to have conducted an interview or focus group with participants by Oct/Nov 2012. It is anticipated that data collection and initial analysis will have been completed by the end of the year 2012.

What are the benefits of taking part?

Your contributions as a participant will help in identifying the possible structure of a data warehouse for managing the DCM data. DCM data will be arranged within the warehouse according to your requirements and needs of data storage and access. It is hoped that you as a care monitoring/providing organisation, will have an opportunity to access the quality, timely and complete DCM data from the DCM data warehouse for comparisons, multidimensional analysis, decision and bench marking purposes.

What are the possible risks of taking part?

No risks have been identified for being taking part in this study. If you do not agree to take part there will be no implications of this research on you or your organisation.

Do you have to take part?

It is your decision to take part. If you decide not to take part this does not affect your employment in any way. You can agree now, but can revoke your decision at any time. You are not obliged to give any reasons.

Will the interview data including my personal information be anonymised?

Wherever possible interview data will be anonymised at the point of collection e.g. research diaries, questionnaires. The interviews, where possible, will be conducted as anonymously as possible using identification codes at the start of the recording and in allocating file names. Transcription of the audio files will ensure full anonymisation through use of research participant codes. When writing up the research findings pseudonyms will be used throughout and any details that might lead to identification of a participant will be changed to ensure their anonymity.

All direct quotations from respondents will be anonymised. Personal contact details will not be held at any point during or after the project, except e-mail addresses or a postal address or telephone number. This will be kept in a separate file on a password protected computer, from any research interview data.

What will happen to the recorded interview or focus group data?

The recorded interviews and focus groups will be taken off the audio recorder and placed on computers. The recordings will be transcribed by the researcher. The transcriptions will be stored on University computer accessed by password. The transcriptions will be analysed to find the relevant information. Once analysis of the data is completed, the audio recordings of the DCM users meetings, focus groups and interviews will be stored on a CD kept in a locked filing cabinet for a period of no more than 5 years after completion of the project, for the purposes of further analysis and write up. After the 5 years it will be destroyed. The filing cabinet will be accessed only by the researcher and the supervisors.

Who has reviewed the study?

This study has been approved by the University ethics panel within Bradford University. They asked us to remind you that, as with anything else, the research will be covered by normal insurance policies and if you are unhappy about anything that takes place throughout the project, you have the right to make a formal complaint.

Where can I get more information?

If you have any concern or questions, please contact the researcher (Shehla Khalid) at 01274 236423 or 07751800475, s.khalid3@student.bradford.ac.uk.

Thank you very much for considering taking part in this research.

Appendix 4: Information sheet for DCM researchers

Shehla Khalid
School of Health Studies
Horton A, 4th Floor
Great Horton Road
Bradford BD7 1DP
T: 01274 236399
M: 07751800475
E: s.khalid3@student.bradford.ac.uk



Information Sheet: DCM researchers

Study Title: A data management framework for Dementia Care Mapping (DCM) using a data warehouse approach to improve the quality of dementia care

I would like to invite you to take part in my research study. Before you decide whether or not to take part, it is important to understand why the research is being conducted, what it will involve and what your role will involve as a participant. Please take time to read the following information carefully.

What is the project all about?

Dementia Care Mapping (DCM) is an observational process and a practice development tool to improve the quality of care and quality of life of people with dementia within formal dementia care settings.

Each year DCM users produce large amounts of data. However, the use of such important, rich data is only limited at a local level, which means the individual mappers or organisations collecting the data uses it within their own settings or research projects, for the purpose it was originally collected. The DCM data collected from various mappers and organisations can be re-analysed for secondary purposes for producing new knowledge that can inform future improvements in the DCM method/tool and provide suggestions for providing quality dementia care within formal dementia-care settings. However, one of the potential reasons of the lack of data sharing among DCM users is the lack of IT system for such purposes. Current electronic data analysis and storage systems have limited functionality and do not permit efficient and integrated storage and analysis of data over time and its dissemination to various users.

This study aims to design a data repository, called a data warehouse, for managing the DCM data for future uses. Within this data warehouse the DCM data will be arranged/presented according to the users' requirements and needs of data storage and retrieval for various analytical, comparison, benchmarking and research purposes. The user-identified structure of the data warehouse will inform the effective, efficient, integrated and systematic solution of managing the DCM data by variety of DCM users from health care organisations, social care organisations, care monitoring/regulation organisations, charities and research backgrounds.

What is the purpose of the study?

This study aims to gather the potential users' requirements for storing and accessing DCM data within a data warehouse. The intentions are to design a user-identified structure of a data warehouse informing a data management framework for DCM to enable variety of DCM users/stakeholders to have a structured and systematic way of accessing data for multidimensional analysis, benchmarking, pattern recognition, monitoring care quality and DCM research.

What will the study involve?

This study will involve conducting semi-structured interviews or focus groups with the staff working in key roles (care delivery, management or decision making) within the health care/social care organisation, care monitoring/improvement/inspection organisations, charities, statutory bodies, researchers and DCM expert users, who are using DCM for various purposes or have intentions to use it in the future. The researcher will ask the participants the relevant questions (using semi-structured interview techniques) to explore their experience of using DCM data for any kind of analysis purposes and requirements to use it for further retrieval purposes

Who is the researcher?

This project is being carried out by a postgraduate researcher from the University of Bradford (Shehla Khalid) under the supervision of Dr Claire Surr (Bradford Dementia Group, University of Bradford) and Dr Daniel Neagu (School of Computing, University of Bradford).

What will you as a participant have to do?

If you agree to take part in the study, you will be contacted by the researcher to arrange an interview either face-to-face or via Skype at your convenient date and time. You will have an informal conversation with the researcher about your experiences of using DCM data

within your research studies and your future requirements of getting access to integrated DCM data

(data that will be collected taken from various individual mappers and organisations) for potential secondary uses.. The interview will take 30-45 minutes of your time and will be audio tape-recorded.

When will the interviews take place?

The research will start in March 2012 and I would like to have conducted an interview with participants by May/June 2012. It is anticipated that data collection and initial analysis will have been completed by the end of the year 2012.

What are the benefits of taking part?

Your contributions as a participant will help in identifying the possible structure of a data warehouse for managing the DCM data. DCM data will be arranged within the warehouse according to your requirements and needs of data storage and access. It is hoped that you as a DCM researcher, will have an opportunity to access the quality, timely, anonymized and complete DCM data from the DCM data warehouse for research purposes i.e. benchmarking, data mining, analysis etc.

What are the possible risks of taking part?

No risks have been identified for being observed in this way. If you do not agree to take part there will be no implications of this research on you or your organisation.

Do you have to take part?

It is your decision to take part. If you decide not to take part this does not affect your employment in any way. You can agree now, but can revoke your decision at any time. You are not obliged to give any reasons.

Will the interview data including my personal information be anonymised?

Wherever possible interview data will be anonymised at the point of collection e.g. research diaries, questionnaires. The interviews, where possible, will be conducted as anonymously as possible using identification codes at the start of the recording and in allocating file names. Transcription of the audio files will ensure full anonymisation through use of research participant codes. When writing up the research findings pseudonyms will be used throughout and any details that might lead to identification of a participant will be changed to ensure their anonymity.

All direct quotations from respondents will be anonymised. Personal contact details will not be held at any point during or after the project, except e-mail addresses or a postal address or telephone number. This will be kept in a separate file on a password protected computer, from any research interview data.

What will happen to the recorded interview data?

The recorded interviews will be taken off the audio recorder and placed on computers. The recordings will be transcribed by the researcher. The transcriptions will be stored on University computer accessed by password. The transcriptions will be analysed to find the relevant information. Once analysis of the data is completed, the audio recordings of the DCM users meetings, focus groups and interviews will be stored on a CD kept in a locked filing cabinet for a period of no more than 5 years after completion of the project, for the purposes of further analysis and write up. After the 5 years it will be destroyed. The filing cabinet will be accessed only by the researcher and the supervisors.

Who has reviewed the study?

This study has been approved by the University ethics panel within Bradford University. They asked us to remind you that, as with anything else, the research will be covered by normal insurance policies and if you are unhappy about anything that takes place throughout the project, you have the right to make a formal complaint.

Where can I get more information?

If you have any concern or questions, please contact the researcher (Shehla Khalid) at 01274 236399 or 07751800475, s.khalid3@student.bradford.ac.uk.

Thank you very much for considering taking part in this research.

Appendix 5: Study participants' details

No	Participant code	DCM Status	Professional role	Affiliations
1	DCM Trainer 1	DCM Trainer	DCM consultant and trainer	University (UK)
2	DCM Practitioner 2	Basic mapper	Senior clinical psychologist	Healthcare setting (UK)
3	DCM Practitioner 3	Basic mapper	Consultant clinical Neuropsychologist	Healthcare setting (UK)
4	DCM Researcher 4	Basic mapper	Research associate	Research organisation (non-UK)
5	DCM Practitioner 5	Advanced mapper	Change manager	Healthcare setting (UK)
6	DCM Practitioner 6	Basic mapper	Activities Co-Ordinator	Healthcare setting (UK)
7	DCM Practitioner 7	Basic mapper	Deputy director of nursing	Healthcare setting (UK)
8	DCM Researcher Practitioner 8	Advanced mapper	Research associate	University (UK)
9	DCM Researcher Practitioner 9	Basic mapper	Research associate	University (UK)
10	DCM Practitioner Trainer 10	DCM trainer	Psychologist	Research organisation and local dementia charity (non-UK)
11	DCM Researcher Trainer 11	DCM trainer	Nurse/PhD student	Research organisation (non-UK)
12	DCM Practitioner 12	Basic mapper	Lead nurse	Hospital setting (UK)
13	DCM Researcher 13	Mapper	Researcher	University (non-UK)
14	DCM Practitioner Trainer 14	Mapper	Clinical psychologist	Social care organisation (UK)
15	DCM researcher 15	Basic mapper	Researcher	University (UK)
16	DCM Researcher Practitioner	DCM trainer	Research nurse	NHS trust (UK)

	Trainer 16			
17	DCM Researcher Practitioner Trainer 17	DCM trainer	Research nurse	NHS trust (UK)
18	DCM Trainer 18	DCM trainer	DCM consultant and trainer	University (UK)
19	DCM Trainer 19	DCM trainer	Dementia consultant and trainer	University (UK)
20	DCM Researcher 20	Basic mapper	No defined role	Dementia charity (UK)
21	DCM Trainer and Practitioner 21	DCM trainer	No defined role	UK
22	DCM practitioner trainer 22	DCM trainer	No defined role	UK
23	DCM practitioner 23	Basic mapper	Dementia pioneer	UK based dementia charity
24	DCM Trainer Practitioner 24	DCM trainer	DCM consultant and trainer	University (UK)
25	DCM Researcher 25	Basic mapper	Research student	University (non-UK)
26	DCM Trainer 26	DCM trainer	No defined role	UK
27	DCM Researcher 27	Basic mapper	Researcher	UK
28	DCM Researcher 28	Basic mapper	Researcher	Dementia research organisation (Non- UK)
29	DCM Trainer 29	DCM trainer	Strategic lead of DCM	Dementia care organisation (Non- UK)

Appendix 6: General interview guide for all study participants

Topic 1:

Current uses of DCM and DCM data

Topic 2:

Issues and concerns related to the current uses of DCM data

Topic 3:

Potential uses of DCM data for secondary purposes

Topic 4:

Concerns and issues related to the secondary use of DCM data

Topic 5:

Data management issues

Appendix 7: Semi-structured interview guide (for researchers)

Introduction

Introduce myself; talk about aims of the project; concept of a data warehouse and secondary use of DCM data; confidentiality of the interview data.

1. Could you please tell me your experience of using DCM data within your research studies?

Prompts

Data collection

Data storage

Data analysis

Data management

2. Could you please tell me how do you use DCM data within your research studies?

Prompts

For what purposes?

How often?

3. Would you please give a brief description of your requirements of using DCM data?

Prompts

Within your research studies

For future research studies

For what purposes

4. Could you please identify what kind of data do you collect and you think is important to collect for your research or analysis?

Prompts

How often these data attributes?

Any new attributes in the future?

5. What are the expected goals of your current research project? What are you working on accomplishing?

Prompts

Important of DCM data analysis

Importance specific data attributes

6. What data dimensions are important to your analysis? (location, time, wellbeing, participants etc)

Added questions:

1. What type of data can help you in your potential secondary analysis of DCM

2. data?
 3. Why do you think this particular data type can help you in your potential secondary analysis of DCM data?
7. How accurate does the data have to be?

Prompt

Quality issues and concerns

Added questions:

1. Some study participants have mentioned their concerns regarding the quality of DCM data for secondary uses. What are your views about it?
 2. What issues/concerns do you see with quality of DCM data for secondary uses?
 3. Who do you think is responsible for ensuring the quality of DCM data?
 4. What do you mean by reliability of DCM data for secondary uses?
8. How often do you use the DCM data for research purposes?

Prompts

Primary data or secondary data

Timely data

9. What is your vision for the future use of the DCM data for further research studies?
10. What would be your reason to access the data from a DCM data warehouse?
11. Is there anything else about the way you analyse the DCM data or future requirement for analysing it that you would like to share?

Well that's all the questions I needed to ask. Thank you very much for taking the time to share your experiences with me. Goodbye.

Appendix 8: Semi-structured interview guide (for practitioners and trainers)

Introduction

Introduce myself; talk about aims of the project; concept of a data warehouse and secondary use of DCM data; confidentiality of the interview data.

1. Could you please tell me your role within this organisation?
2. Could you please tell me what is your role in dealing with Dementia Care Mapping (DCM) data within your organisation?

Prompts

Care delivery

Information management

Decision-making

3. How are you using DCM data within your organisation and for what purposes?

Prompts

Kind of analysis

What part of DCM data is retrieved?

How often?

For what purposes?

4. Would you give a brief description of the type of analysis you conduct on DCM data and why?

Prompts

Within the organisation

For what purposes

One time or an on-going requirement

- If on-going requirement, will it be on a regularly scheduled basis or as requested
 - What will be the frequency?
5. Who will be receiving the analysis results (besides yourself)?
 6. Are there any security data issues that must be considered?
 7. Could you please identify what kind of attributes from DCM data are important for your retrieval purposes?

Prompts

How often these attributes?

Any new attributes in the future?

Once collected, how do you manage your DCM data for analysis purposes?

Prompts

Any particular system do you use

How do you find this system?

8. Do you plan to run the query/report yourself or will you expect other group such as IT to run it?

Prompts

What kind of queries?

What kind of reports?

How do you see the results as a visual (graphs, pie charts etc)

9. How accurate does the data have to be?

Prompt

Quality issues

Added questions:

Who do you think is responsible for data quality?

How do you maintain the quality of DCM data?

10. What is your vision for potential secondary uses of DCM data for further analysis?
11. What would be your reason to access the data from a DCM data warehouse?
12. Is there anything else about the way you analyse the DCM data or future requirement for analysing it that you would like to share?

Well that's all the questions I needed to ask. Thank you very much for taking the time to share your experiences with me. Goodbye.

Appendix 9: Individual interview summary memo

Summary of interview with Practitioner 2 (Clinical psychologist, a healthcare organisation. Basic user and DCM lead within this organisation)
My understanding of this interview data Here this person is familiar with her/his needs for DCM data use and has an idea how they (organisation) want to use the data in the future. She/he is also familiar with the limitations of current data analysis system (the Excel system) and mentioned that this system just provides data based on single mapping per spreadsheet. Bringing raw data together based on per person is not possible. Organisation is supporting of DCM and regular mapping. Yet, mappers struggle to find time to map.
The points need to explore further Potential use of DCM data ‘learning from others’ (what other potential uses?) Organisation’s role in supporting mappers, mapper’s role in data quality, requirements for a system for in-house analysis such as ‘looking at data overtime’ (what other uses?)
Where to look for new data (source) Need to conduct more interviews and re-analyse the previous interviews
Participant’s main concerns: Limitations of the Excel programme

Appendix 10: Memo example








Memo: 17

From the very first few interviews I realised that 'data quality' and 'trusting others data' both categories are coming together. As both are pointing to a phenomenon of trustworthiness of DCM data. The main actors within this phenomenon are Individual mappers and organisations. These both have affiliation as well where mappers are part of the organisations. These two directly hold responsibility of the quality of data. For example, care settings conduct DCM to understand what is going on within their organisation. They conduct the process for developing practice or their staff, to monitor the quality of care and to find out how they can improve the quality of care and quality of life of people with dementia within their care settings. For this purpose, the main aim is to carry out DCM for specific purposes and feedback the results to staff to see the issues or areas where they can improve. (Ask practitioners if they think their data is biased).

Researchers show concern that the data collected is for practice development and might not be of research quality. This needs to be explored further to learn why is this the case.

The provenance (history of data set that authenticate the credibility of data by establishing its original source, the process of its creation) of DCM data can be traced back to the process of observation, who did that observation and what was the purpose of that observation. Further, the processing of data is also part of provenance of data. As interview data reveals participants' requirement of having trustworthiness data that will go into the data warehouse and users will need to know provenance of DCM data (which I called additional data in my categories) to judge the credibility of DCM data. This is important usually in data warehouses where data is integrated taken from various sources and used for a different purpose. The literature asserts that data quality can be credited or controlled by providing provenance of data to let users be judge of the quality and usability of data for a specific purpose.

Appendix 11: A list of some codes

Type	Name	Memo Link	Sources	References	Created On	Created By	Modified On	Modified By	
	Accessing the system		2	2	30/08/2012 12:15	SK	05/09/2012 10:45	SK	
	Aggregated data use		1	1	29/08/2012 15:48	SK	05/09/2012 10:39	SK	
	Anonymizing the data		4	7	29/08/2012 15:36	SK	05/09/2012 10:39	SK	
	Awareness		7	11	29/08/2012 15:41	SK	09/09/2012 13:44	SK	
	Benchmarking data		2	3	30/08/2012 23:07	SK	05/09/2012 10:40	SK	
	Benefits of current analysis system		2	2	29/08/2012 15:57	SK	05/09/2012 10:45	SK	
	Careful data entry and integration		2	2	29/08/2012 15:49	SK	05/09/2012 10:45	SK	
	Change overtime		0	0	29/08/2012 15:55	SK	29/08/2012 15:55	SK	
	Common required DCM data		1	1	30/08/2012 12:21	SK	05/09/2012 10:45	SK	
	Comparing data overtime		0	0	29/08/2012 15:42	SK	29/08/2012 15:42	SK	
	Comparing data overtime (2)		0	0	30/08/2012 11:43	SK	30/08/2012 11:43	SK	
	Comparing services		4	4	29/08/2012 15:56	SK	05/09/2012 10:45	SK	
	Complete data		0	0	29/08/2012 15:59	SK	29/08/2012 15:59	SK	
	concerns		2	3	29/08/2012 15:36	SK	05/09/2012 10:40	SK	
	Conditions		2	2	29/08/2012 15:43	SK	05/09/2012 10:34	SK	
	Confidence		2	3	29/08/2012 15:34	SK	05/09/2012 10:45	SK	
	Data Accuracy		1	1	30/08/2012 12:13	SK	30/08/2012 13:50	SK	
	Data accuracy (2)		1	1	30/08/2012 12:11	SK	05/09/2012 10:45	SK	
	Data confidentiality		1	1	29/08/2012 15:47	SK	05/09/2012 10:39	SK	
	Data outcomes		0	0	29/08/2012 15:46	SK	29/08/2012 15:46	SK	
	Data ownership		1	2	29/08/2012 15:47	SK	05/09/2012 10:39	SK	
	Data security		2	5	29/08/2012 15:36	SK	05/09/2012 10:39	SK	
	Data sharing		2	2	29/08/2012 15:54	SK	05/09/2012 10:39	SK	
	DCM as non judgemental tool		1	1	29/08/2012 15:52	SK	05/09/2012 10:45	SK	
	DCM for Evaluation		1	1	30/08/2012 22:24	SK	30/08/2012 22:26	SK	
	DCM KPI		2	3	29/08/2012 15:53	SK	05/09/2012 10:23	SK	
	DCM use		5	27	29/08/2012 15:18	SK	05/09/2012 10:45	SK	
	Desire		1	2	29/08/2012 15:53	SK	05/09/2012 10:23	SK	
	Desire (2)		1	2	31/08/2012 09:14	SK	05/09/2012 10:23	SK	
	Diagnostic work		0	0	31/08/2012 09:11	SK	31/08/2012 09:11	SK	
	Dissemination of results		1	1	29/08/2012 15:39	SK	05/09/2012 10:33	SK	
	Evidence		4	5	29/08/2012 15:50	SK	05/09/2012 10:45	SK	
	Expectation		2	5	29/08/2012 15:28	SK	05/09/2012 10:33	SK	
	Experience		1	1	29/08/2012 15:48	SK	05/09/2012 10:17	SK	
	Flexibility in accessing the system and data		2	3	30/08/2012 12:20	SK	05/09/2012 10:39	SK	
	Flexibility		2	3	29/08/2012 15:51	SK	05/09/2012 10:45	SK	
	Gathering extra information using other methods		1	1	30/08/2012 14:02	SK	05/09/2012 10:33	SK	
	Getting help or support		1	2	30/08/2012 16:21	SK	05/09/2012 10:34	SK	
	Giving context		1	1	29/08/2012 15:51	SK	05/09/2012 10:17	SK	
	In house tools		1	2	31/08/2012 13:55	SK	05/09/2012 10:23	SK	
	inability		4	7	30/08/2012 11:50	SK	05/09/2012 10:45	SK	
	Inability (2)		0	0	30/08/2012 11:47	SK	30/08/2012 11:47	SK	
	Influencing the culture		1	1	29/08/2012 15:58	SK	05/09/2012 10:17	SK	
	Information required from the data		4	8	29/08/2012 15:38	SK	05/09/2012 10:45	SK	
	IRR check		2	2	30/08/2012 13:51	SK	05/09/2012 10:40	SK	
	Keeping track of patients		2	2	29/08/2012 15:35	SK	05/09/2012 10:39	SK	
	Laboreous		2	2	29/08/2012 15:49	SK	05/09/2012 10:45	SK	
	Lack of confidence		3	3	29/08/2012 15:39	SK	05/09/2012 10:45	SK	
	Limitations of current analysis system		5	18	29/08/2012 15:57	SK	05/09/2012 10:45	SK	
	Making judgement		0	0	29/08/2012 15:48	SK	29/08/2012 15:48	SK	
	Managing data quality		2	3	29/08/2012 15:56	SK	05/09/2012 10:23	SK	
	Mapping duration		3	3	29/08/2012 15:45	SK	05/09/2012 10:17	SK	
	Mapping frequency		4	8	29/08/2012 15:18	SK	05/09/2012 10:23	SK	
	Mapping time period		1	1	29/08/2012 15:45	SK	05/09/2012 10:17	SK	
	Meeting needs		0	0	29/08/2012 15:46	SK	29/08/2012 15:46	SK	
	Need		0	0	29/08/2012 15:29	SK	29/08/2012 15:29	SK	
	Non-DCM data		4	11	29/08/2012 15:24	SK	05/09/2012 10:45	SK	
	Not want to compare services		0	0	30/08/2012 11:36	SK	30/08/2012 11:36	SK	
	Novice users		0	0	31/08/2012 17:26	SK	31/08/2012 17:26	SK	

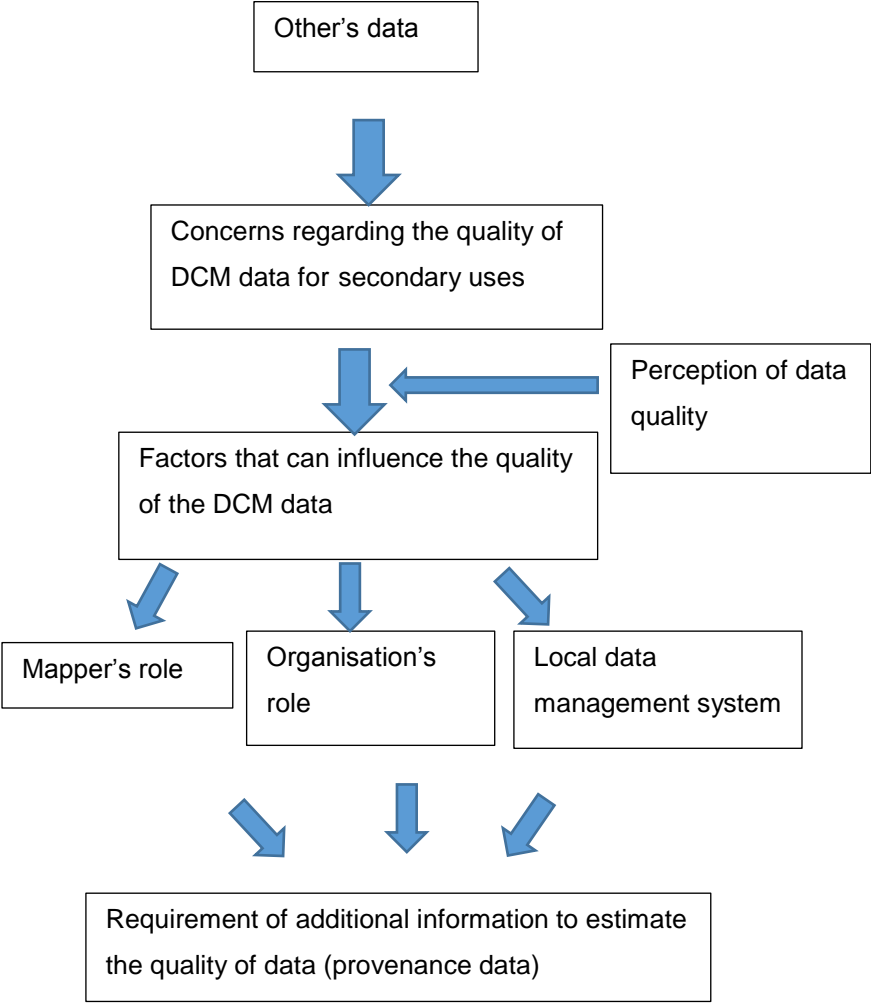
Appendix 12: A list of some codes and categories

codes	categories
<p>Interpreting data (giving meaning to the DCM raw data) Understanding DCM for non-mappers Interpreted data goes to staff or organisation Mappers role in data interpretation Interpreting while inputting the data Non-mappers will not be able to interpret the data Non-mappers access to the processed data</p>	<p>Primary use_data interpretation (local data management systems' requirements)</p>
<p>Having access to data according to individual maps Wanting to able to pull out individual data from time dimension, Wanting data arranged per patient, Wanting to analyse DCM data with other data Wanting to bring data in one system (electronically connected) Wanting to explore DCM data further Wanting to join up DCM data with other data Wanting to know individuals overtime over a number of maps Wanting to look at data from various angles (points) to get a complete picture Wanting to look at the service level data Wanting to look at ward data for localised changes Wanting to manipulate the data based on their needs Wanting data arranged per patient Wanting to re-arrange my data Wanting to look at the data overtime Wanting the opportunity to re-arrange the data Data conversion from one format to other Being able to link the data items Being able to arrange data in different ways Wanting to have the facility for comparing data Wanting to compare data side by side</p>	<p>Primary Use_flexibility in accessing data</p> <ul style="list-style-type: none"> • Re-arranging data • Integrating data • Comparing data • Categorising or organising data • Manipulating data <p>Shared access</p>

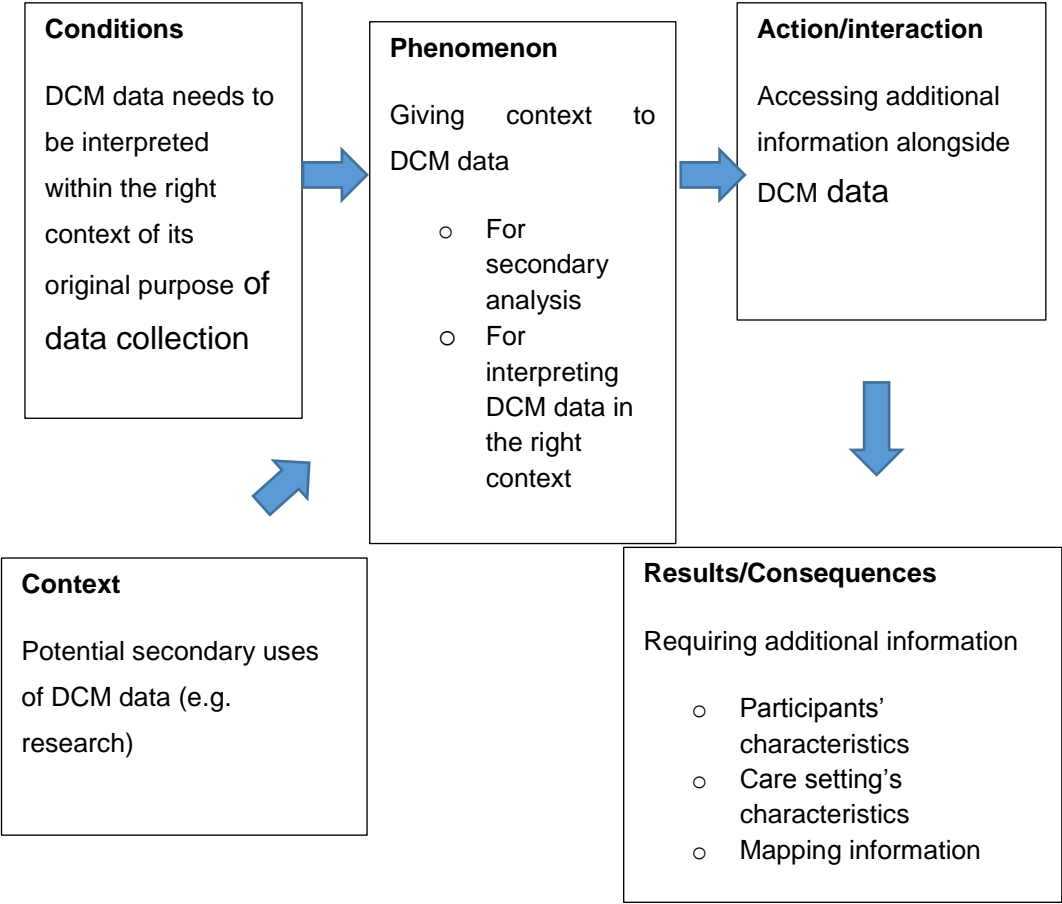
<p>Wanting flexibility in type of accessing information or data</p> <p>Wanting a system that could store all data at one place</p> <p>Looking at the information differently in different times</p>	
<p>Looking at whole group data for the whole of the service,</p> <p>Looking at high potential activities,</p> <p>Looking at collated maps from three data points and from the period of three months,</p> <p>Looking at activities from time perspective</p> <p>Wanting to look at the individual changes over time</p> <p>Looking at staff interaction that triggered the period of well-being</p> <p>Being able to compare individuals' wellbeing overtime in a setting</p> <p>Wanting to see development in care overtime</p> <p>Wanting to look at the collated data overtime</p> <p>Wanting to look at the change at organisational level</p> <p>Wanting to look at staff training needs</p> <p>Using DCM for developing practices</p> <p>Using DCM data for general research purposes</p>	<p>Secondary use_ requirements (the re-use of the data is required within the organisations mostly)</p> <ul style="list-style-type: none"> • Looking at service level data from time perspective • Seeing development of care overtime • Setting benchmarks for practice • Looking at change overtime at organisational level • Looking at change over time at individual level
<p>Not collecting other data with maps</p> <p>Seeing potential of analysing DCM data within the context of other data</p> <p>Time consuming activity for collecting more data?</p> <p>Cultural and organisational difference in collecting data</p> <p>Collecting contextual data</p> <p>Using contextual data (participants' age, gender related data) for analysis</p> <p>Wanting to know the contextual data</p> <p>Recording additional data to give extra depth to observation</p> <p>Collecting additional data (modifying DCM data)</p> <p>Collecting additional data</p>	<p>Secondary use_contextual data</p> <ul style="list-style-type: none"> • Giving meaning • Enhancing use of DCM data • Contextual analysis • Various types
<p>Reporting at individual ward level,</p> <p>Reporting data,</p> <p>Reporting data at organisation level</p>	<p>Primary use_reporting data</p> <ul style="list-style-type: none"> • Individual level • Organisational level

<p>Reporting data in descriptive way Using visual representation of data for feedback Graphs help non-mappers to understand the data Wanting to set up IT system for automated report generation</p>	<ul style="list-style-type: none"> • Visual representation • Descriptive presentation <p>Automated report generation</p>
<p>We don't do enough maps, Wanting to do longer and more mapping, Want to do more mapping, Working with others to look for solution Wanting to know how the PD's and PE's are related to the wellbeing Only can compare like for like data Not using DCM data for comparing wards or settings Issues with PD's and PE's Requiring more guidance on detractors and enhancers Setting up a dedicated time for mapping Visualising the future use of DCM Recognising usefulness of comparing data overtime Feeling the potential of re-using the DCM data Enhancing the use of DCM to make the re-use of data</p>	<p>Requirements related to the DCM method</p> <ul style="list-style-type: none"> • Managing the use of DCM • Enhancing the use of DCM • Exploring the use of data
<p>Wanting a system that could store all data at one place Requiring a system that provide integrated view of the data Reporting all DCM data together Others' data Needing mapper's reliability score Needing to know the type of care setting Needing mapper's experience Needing mapper's training Advanced mappers have reliable data Consulting DCM manual</p>	<p>Primary use_integrated view of data (having all the data at one place to view and use it for various purposes)</p> <ul style="list-style-type: none"> • Stored in one system • Stored within the organisation • All data together <p>Secondary use_data trustworthiness</p> <ul style="list-style-type: none"> • Mapper's reliability • Data trustworthiness

Appendix 13: A flow diagram showing the sequence and flow of emergence of category 'provenance data'.



Appendix 14: A model for describing category ‘contextual data’



Appendix 15: Publications and conference papers

1. Khalid S., Surr C., and Neagu D., (2014) 'Designing a Data Warehouse for Dementia Care Mapping (DCM) for Monitoring and Improving Quality of Dementia Care'. **ACM-W Europe womENCourage** 2014 International Conference; celebration of women in computing. University of Manchester, UK (short paper and poster presentation)
2. Khalid S., Surr C., and Neagu D., (2014) 'Designing a Data Warehouse for Dementia Care Mapping (DCM) for Monitoring and Improving Quality of Dementia Care' **British Society of Gerontology** 43rd Annual conference 1st -3rd Sep 2014 in University of Southampton, UK (Oral presentation)
3. Khalid S., Surr C., and Neagu., D (2013) 'Designing a Data Warehouse for Dementia Care Mapping (DCM); a data sharing platform to improve the quality of dementia care' **Alzheimer's Research UK (ARUK)** Yorkshire Network Annual Scientific Meeting 2nd July in University of Leeds, UK (Poster presentation).
4. Khalid S., Surr C., and Neagu., D (2013) 'Designing a Data Warehouse for Dementia Care Mapping (DCM); Exploring users' data needs to inform conceptual data modelling' The 8th **UK Dementia Congress** 2013 in Novemeber in Nottingham, UK (Poster presentation)