A COMPUTATIONAL APPROACH TO PRIMARY HEALTHCARE INFORMATION QUALITY INDICATORS

A thesis presented in partial fulfilment of the requirements for the degree of
Doctor of Philosophy
in
Computer Science
at Massey University, Palmerston North,
New Zealand.

Stephen Richard Lean

2014

For my parents, Beverley and Glen. Gone much too soon, remembered everyday.

Abstract

In many countries around the world Information and Communication Technologies (ICT) are being leveraged to produce efficiency gains and cost reductions in healthcare by making health information more readily available in clinical contexts. This raises issues as to the use of health information in clinical decision making at point of care, as relying on poor quality information in this context can have serious consequences. This thesis investigates quality criteria that are used when assessing health information, with the objective of formalising those criteria for use with a prototype software system. Literature, as well as standards and currently used forms of electronic health records, were reviewed for what they offer for assessment of health information quality. A lack of criteria from these sources necessitated interviewing practicing General Practitioners (GPs) to determine criteria important to them, and how they assessed the information they want to use. Interviews were of a semi-structured type using vignettes, for clinical context. Recruitment used a Snowball methodology. Results were analysed and interpreted using Thematic Analysis and showed the GPs assessed information using criteria based on tacit knowledge, formed from community knowledge and past experience.

The Quality Criteria (QC), discovered to be integral to this process, were formalised using the developed Quality Criteria Model (QCM). A prototype system was developed to demonstrate that using a current health information standard, meta-data could be used to detect the presence of QC within health information and capture these via instantiation of the QCM. The results of successful detection of QC are then Health Information Quality Indicators (HIQI). Contributions for this thesis include the following: the set of discovered QC, thematic maps that capture the combination of criteria and the process used when applying them, the formalised model for QC (the QCM), determination that additional meta-data will be required to detect those QC categorised as being subjectively evaluated, and the demonstration that a software system can detect, and capture, QC found in health information. Implications are discussed such as that just having access to information is insufficient, and subjectively evaluated QC are problematic for implementation and use. Finally, conclusions are drawn and future work suggested such as user interface development for HIQI representation, alternative search algorithms for QC detection, and further development of the prototype toward a production system.

Acknowledgements

This thesis would not have been completed without the help of many people. I would like to take this opportunity to thank those who helped with this project.

Deep appreciation goes to my three supervisors: Prof. Hans W. Guesgen, Dr. Inga Hunter and Dr. Kudakwashe Dube. Many thanks for the guidance and expertise shown throughout. The complimentary skills that each brought to the project helped in immeasurable ways, along with the encouragement and patience. It was not only the help with the research that was appreciated, but also the support shown on a personal level.

My thanks to staff in the School of Engineering and Advanced Technology (SEAT), especially Michele Wagner, who helped in innumerable ways, and Prof. Stephen Marsland who supported the project when needed. I would also thank my fellow postgraduate students Gareth, Yuliya, Jawad and An, with who I shared an office. The help, encouragement and light relief, along with some lively discussions, provided along the way will always be remembered.

I would also like to thank Angela Baker for wise and supportive council, it really helped. Also, thanks to Janene Walter from Massey University's Graduate Research School for assistance with the process of completing a PhD.

Thanks to Massey University for financial support in the form of the Vice-Chancellor's Doctoral Scholarship, without which I would not have started this project. Also, to SEAT for part-time work opportunities and for the allocation of a budget for research expenses.

I am also grateful to those who participated in the interviews that formed part of this research. Without the interviews, this project would not have been possible.

Last, is thanks to family and friends who showed an interest and continued to ask how it was all going.

Contents

List of Figures

List of Tables

Listings

List of Abbreviations

ASTM	American Society for Testing and Materials International
CCR	Continuity of Care Record
CDA	Clinical Document Architecture
EHR	Electronic Health Record
ePA	Electronic Prescribing and Administration
eSCRV	Electronic Shared Care Record View
FHIR	Fast Healthcare Interoperability Resources
GP	General Practitioner
HISO	Health Information Standards Organisation
HIQI	Health Information Quality Indicators
HL7	Health Level 7
NZHS	New Zealand Health and Disability Sector
NZMA	New Zealand Medical Association
MOH	Ministry Of Health
OECD	Organisation for Economic Cooperation and Development
OO	Object Oriented
PMS	Practice Management System
POC	Proof Of Concept
QC	Quality Criteria
QCM	Quality Criteria Model
SCR	Summary Care Record
TDQM	Total Data Quality Management
XML	eXtensible Markup Language

Chapter 1

Introduction

General Practitioners (GPs) are part of a profession that is reflective in nature. It demands that each consultation with a patient be assessed individually, not as part of a routine. One of Aristotle's intellectual virtues, *phronesis*, or practical wisdom, has been studied as part of this reflective practice in clinical care settings (Frank, 2012). Use of health information for clinical decision making is a part of the profession. As such the same reflective nature is applied with health information. It is the approach to health information quality assessment, and a formalisation of key criteria used in that approach, that this thesis explores.

1.1 Background and Motivation

Worldwide, demand for healthcare services and healthcare costs, are increasing. Many factors contribute to this. In the western world contributing factors are ageing populations, increasing rates of obesity and diseases such as diabetes, and the rising costs of medical equipment and salaries for health professionals. New Zealand faces these issues, with the Organisation for Economic Co-operation and Development's (OECD) 2009 Economic Survey report for the country stating that the biggest threat to the long run fiscal sustainability of the country is rising healthcare costs (OECD, 2009, pp. 109-118).

Many countries have turned to leveraging electronic health information as one way of improving efficiencies, and reducing health expenditure. The New Zealand Health Sector (NZHS) too has adopted this approach and is in the process of formulating, and implementing, several strategies to realise efficiency gains through electronic health information. Internationally, and in New Zealand, Electronic Health Records[1] (EHR)

[1]There are many different terms for this, such as Electronic Health Record, Electronic Medical

have been seen as one tool integral to implementing these strategies.

Many examples, and uses for, EHRs exist, with a multitude of implementations and standards (Eichelberg et al., 2005). One use for an EHR is providing health information to health practitioners at point of care. Strategies for EHR implementation are moving beyond monolithic systems that collect and keep health information, and are not linked to other systems, to systems that link together health information from disparate sources into an EHR. Essentially, a conglomeration of patient information, brought together into a single view. Some examples of strategies implemented, or being implemented, in New Zealand include: GP2GP system for electronic transfer of complete patient records when patients transfer between GPs, electronic Prescribing and Administration (ePA) for electronic prescribing direct to the pharmacy, and electronic Shared Care Record View (eSCRV) which provides a view of a patient's relevant health information from a single log on and aggregated from multiple sources (NZ National Health IT Board, 2013).

Coupled with this increased access to health information for clinical purposes at point of care, is the stated requirement that the information be used safely (NZ National Health IT Board, 2010), with respect to the patient, in clinical situations. The information must therefore be assessed by health practitioners as being fit for purpose, or of good quality such that the information can be used.

However, with more information becoming available to health practitioners, the health practitioner becomes further removed from any personal experience of the information, losing knowledge that would be beneficial in judging the information fitness for purpose (Fisher et al., 2003). This makes assessment of the information fitness for purpose (quality), and judgements for the information safe use, problematic. The situation is further complicated by health practitioners, such as General Practitioners (GP), belonging to a profession constrained by ethics such as always acting in the best interest of the patient, and minimising harm from treatment decisions. With GPs, they will have access to more health information, as health sector strategies come to fruition, but are then faced with judgements as to the quality of the information, information that the are unable to accept (rely on), or reject (discard), at face value.

In light of the importance of having health information available at point of care for treatment decisions, this thesis will try to answer questions with the issue of assisting health practitioners assessment of health information, so as to aid in judgements of the information fitness for purpose.

Record, Personal Health Record and Computerised Patient Record to name a few. For this thesis it will be referred to as an Electronic Health Record.

1.2 Research Aims And Goals

The goal of this research was to determine Quality Criteria (QC) used in the assessment of quality of health information, and to model those identified QC. Further, to demonstrate, via a prototype, that a system could detect the presence of QC within health information, and capture the QC via the developed model. Realised QC, associated to health information, are Health Information Quality Indicators (HIQI). The research is a proof of concept study, to demonstrate the feasibility of providing HIQI, not that the method, or result, is the only, or best, way of achieving the result.

As more health information is made available for health practitioners for clinical purposes at point of care, issues of making the most of that information become important. This thesis seeks to illuminate and understand these issues. If it can be understood how the information is evaluated for use, as to its fitness for purpose, system design and implementation can be tailored to take this into account. Therefore, this research is aimed at assisting in making the most of health information, by determining what it is about the health information that allows the GP to safely use the information.

1.2.1 Research Questions

For safe use, GPs must have confidence in the quality of the information. The GP must evaluate the information quality for the information they are presented with. This led to the research question:

> What can a model based approach provide the health practitioner, when using electronic health information, within a clinical decision making context, such that the health practitioner is assisted in the judgement of the quality of the health information?

Providing Health Information Quality Indicators (HIQI) to assist in information judgement requires underlying factors for quality, and information use, be understood. These factors are posed as research questions:

1. What criteria do GPs use when evaluating quality of electronic health information?

2. What information (meta-data), that currently accompanies electronic health information, can determine these criteria?

3. What additional information (meta-data), that is not currently available with electronic health information, could determine these criteria?
4. If present, are the criteria able to be detected by a system?
5. If detected, are the quality criteria able to be modelled with the electronic health information and indicated to the GP?

1.3 Research Scope

The scope of research is limited to how a group of primary care health practitioners assess the quality of health information they wish to use. Specifically, a group of health practitioners (GPs) who, as well as working in private practice, are also rostered to supply services to an after hours emergency clinic.

Questions of access to, and privacy of, health information are not part of this research. The context for the research takes as a given that the information is available and relevant to the GPs current context of need. There is no claim as to how the health information should be presented to the GP, or how HIQI should be represented.

1.4 Research Design

Several methodologies are used at different stages throughout the thesis. A need to interview GPs was identified. A qualitative research approach was used for this stage. This used semi-structured interviews using vignettes, plus a quantitative rating exercise for interview participants to complete. Recruitment was carried out using a Snowball methodology. Interview analysis and interpretation was completed using Thematic Analysis, along with some statistical analysis used for the rating exercise.

Model development employed proven technology and techniques, such as XML schema, and type creation and specialisation. Demonstration of model and other required functionality was carried out using a prototyping methodology, as prototyping is well suited to the demonstration of concepts via software implementation.

1.5 Thesis Contributions

The thesis has provided several contributions, founded on the success of this research. The contributions are discussed in-depth in Chapter 6. They are listed here in summary, and are as follows:

- Identification of criteria used by a group of practising GPs for assessing quality of health information, that assists them in the determination of the health information fitness for purpose.
- Development of thematic maps that capture the combination of criteria, and the process used when applying the criteria.
- Identified that Quality Criteria (QC) are able to be formally defined, and categorised. The QC are able to be defined through three dimensions of the health information: entry, source for entry, or whole document.
- Development of a model for capturing the semantics of Quality Criteria, the Quality Criteria Model (QCM).
- Demonstrated that QC are able to be detected using meta-data, within a current health information model, the CCR, without the need to alter, or extend, the CCR model.
- Demonstrated that QC, once detected/realised, can be instantiated and incorporated into health information, via the QCM, and their presence indicated to the GP as Health Information Quality Indicators (HIQI).
- Determined that additional meta-data is required to detect QC, categorised as subjectively evaluated.
- A theory that providing scores for abstract quality dimensions, such as Accuracy and/or Completeness, and applying them to health information, has limited utility for GPs as professionals with a duty of care to limit adverse harm to patients in clinical decision making.

1.6 Thesis Plan and Organisation

The remainder of the thesis is structured as follows:

Chapter 2. The motivation and need for the research is explained, and the context of the research problem is given. Research questions are defined. Literature is reviewed, where studies such as why information quality is important, information quality assessment, and how information quality affects decision making are investigated to determine these studies efficacy in answering the research questions. Lastly, the need to go to stakeholders (GPs) for further information is explained.

Chapter 3. This chapter describes the interviews conducted with the group of GPs. Interview methodology, participant recruitment, question development, and the

criteria rating exercise, conducted with participants, are discussed. Interview results are presented, followed by data analysis, and data interpretation. Themes for GPs use, and assessment of, health information are discussed, and Quality Criteria (QC) selection, justification, and organisation are given.

Chapter 4. Health information models are discussed, and their features described, to give understanding as to the environment that the modelling of QC will involve. Common features to health information models, such as type specialisation, are discussed. Development of the Quality Criteria Model (QCM), including its components, structure, and features, for modelling QC is described. Examples for its use in processing health information are provided. The rationale for highlighting QC presence or absence within health information, as opposed to usage for a quality score of the health information, is given.

Chapter 5. Development of the prototype system used to demonstrate the proof of concept is given. The rationale for a prototype, a specific health information model for use with the prototype, and development environment, is described and justified. Elements from the specific health information model, used for detecting QC, are explained. Prototype components, and features, are highlighted and example operations provided. A look at how the QCM could be incorporated for use with another health information model completes the chapter.

Chapter 6. Reflection on answers to research questions, and contributions of the thesis, are discussed and justified. Implications for theory and practice are highlighted, along with recommendations to come from this research.

Chapter 7. Conclusions and lessons learnt are given. How successful the research was, and future work that has come from the results, are discussed.

The thesis also includes several appendices. These are:

Appendix A provides ethics approval documentation, Figure A.1, for interviews with the GPs.

Appendix B provides the documents used in the interview recruitments process, and the interviews themselves i.e. Interview Information Sheet, Participant Consent Form, Question List, and Rating Exercise Form.

Appendix C gives the XML-schema that formally describes the Quality Criteria Model. It also contains the resulting QCM instance, Figure 5.19, from example prototype operation from Chapter 5, Section 5.19.

Chapter 2

Background, Motivation And Research Problem

2.1 Chapter Overview

This chapter is organised into two parts, followed by conclusions. First, the motivation for this research is presented. This includes the domain in which the research was carried out, as defined in the Scope Of Research section, Chapter 1, Section 1.3. It describes issues being addressed within that domain and how they relate to this thesis. This includes the use of information as a resource, and specific issues with information within health.

Second, related research in the area of data and information quality is described in the Literature Review. The Literature Review is the main part of the chapter and starts by defining data quality, and why data quality is important. It begins broadly and continually narrows, based on the topic of this research, until specific work is highlighted that is close to the subject under investigation. The problem area is also described. This gives the reader a background understanding of approaches taken by others in the field of data quality. It is also used to illustrate the threads of research that are used to contribute to the thesis and to provide a context for where the thesis sits in the broad domain of data quality.

Last, conclusions are made and an overview of the components for this research is given. It describes why the researcher's own investigation was required. This highlights specific areas this research involves, and sets a platform for research direction.

2.2 Health Information as a Resource

There is a distinction that exists between what is data and what is information. The distinction is based on context. Data is raw facts, while information has meaning and usefulness to the user. An example demonstrates this. For a person viewing a bus timetable, but not needing to catch a bus, the timetable contains data. However, should the person wish to catch a bus, the timetable contains information. The timetable for the latter person has meaning and usefulness in the context of wanting to catch a bus. Another example, from the health domain, would be a practice nurse viewing a patient's health record and seeing the patient is allergic to penicillin. For the practice nurse the allergy status is data. For a GP using the same record in a consultation with the patient, and wishing to prescribe the patient penicillin, the allergy status is information.

This thesis deals with information, specifically electronic health information and its use by a GP, at point of care, in clinical decision making. It is this context for use, where the information has meaning and potential usefulness, in treatment of patients, that makes it information. Where the term data is used in this thesis it is in line with the above definition i.e. referencing data as a collection of facts, rather than with meaning for a specific context. The term data is also used in the literature review where the topic of many studies was data quality.

Information is a resource utilised widely within organisations for decision making tasks. This can be from strategic, long term goal planning e.g. research and development focus for a pharmaceutical company, to daily operational requirements e.g. pick lists for shipping product from a warehouse. This is no different within health, with the New Zealand Health and Disability Sector (NZHS) focused on making more, and better, use of health information. A core part of the NZHS IT strategy is that information should be collected once and used many times, with six defined uses for health information (NZ National Health IT Board, 2010). In no order, the six uses are:

1. support for patient self management
2. support for clinical intervention
3. governance of clinical processes i.e. maintenance of clinical standards, training, and risk management.
4. administration e.g. quality assurance
5. strategy and policy development

6. research

Health information is used in clinical decision making contexts (Kerr et al., 2007; Protti and Bowden, 2010; Blumenthal and Tavenner, 2010), as per use two given above. Several factors are relevant for health practitioners using health information in decision making. Within the scope for this thesis, the relevant factors are:

- the context for which the information is needed.
- the relevance of the available information to that context.
- the quality of the information to be relied upon.

The factors can be summarised as having good quality information available, relevant to the current context. For this research it was the assessment, by the health practitioner, of the quality of the information to be used, in a clinical context, that was of interest. The health practitioners in this thesis being a group of practising General Practitioners (GPs).

For a given decision task there will be an outcome. Outcomes are severely impacted by the quality of the information used (Shankaranarayanan et al., 2006; Price et al., 2008). This fact is one of the primary motivations for this research; what assists a GP in the judgement of health information quality. GPs face further pressure when wanting to use health information. That is that they are are time poor for decisions, with consultations with patients in general practice being limited to fifteen minutes. As a GPs access to information increases, through implementation of national health IT strategies, this pressure will become more acute. The following section will discuss information within the context of the New Zealand Health Sector, including initiatives for making health information more available to health practitioners.

2.3 Health Sector Pressures

Sharing health information meaningfully and safely has long been a goal of health providers around the world, including New Zealand. For example New Zealand is in the process of implementing strategies from the *National Health IT Plan 2010*. Another example is the United States of America's Meaningful Use regulations that has as one objective to provide a summary of care record for transitioning, or referring, patients to other health providers (Blumenthal and Tavenner, 2010). This leads to using a patient's health information collected in one context e.g. a patient's primary health care

Table 2.1: New Zealand Health Sector information volumes and flows circa 2006.

17,000 Primary Care Organisations	21 District Health Boards
• 50,000 GP visits	• 1,200 hospital admissions
• 105,000 prescriptions	• 2,000 Emergency Department visits
• 40,000 lab tests	
• ? NGO interventions	• 4,000 outpatient visits

provider (GP), and using it in another context e.g. a hospital's Accident and Emergency department, by a health practitioner who did not originally collect the information. Sharing means that a health practitioner can access the health information of a person they are treating at the point of care. Sharing of health information also occurs with, and for, secondary use, but for this research, it was the health practitioner's perspective that was of interest.

The NZHS handles large volumes of information and complex information flows e.g. 1,300,000 referrals per annum from GPs alone. Access and effective use of this information is needed for efficiency gains in this sector as a lack of appropriate information is costly, both in financial terms and in adverse outcomes for patients. However, this information must be used safely, especially in clinical decision making. The effective, but safe use, has become a key driver for the sector.

It is helpful to look at the amount of data being collected on a daily basis throughout the NZHS. This is to gain an understanding of data volumes and flows, that systems within the Sector are currently collecting. Table 2.1 shows daily health interventions for New Zealand (circa 2006) divided into primary care and hospital interventions. More detailed data on information volumes and flows is given in Table 2.2. This represents General Practice information flows per annum, with "out" being information coming from General Practice and "in" being information going to General Practice.

Tables 2.1 and 2.2 represents an enormous amount of information, both in volume and differing types, collected and recorded about patient's health. At issue is not only the amount of data but also that the data is not readily shared, and stays isolated with the system it was collected in. This means it is not available for use by health practitioners who are outside of those systems. The situation leads to isolated "silos of information", problems with the quality of the data captured, and continued fragmentation of both data and systems (NZ National Health IT Board, 2010).

The need for efficiency gains that reduce costs in the NZHS is highlighted in the Organisation for Economic Co-operation and Development's (OECD) biennial economic survey for New Zealand 2009. The report states that increasing health sector costs are

Table 2.2: New Zealand general practice information volumes and flows circa 2006.

Out	In
• 340,000 cytology requests	• 331,000 cytology results
• 180,000 patient records	• 435,000 discharge advices
• 390,000 vaccination reminders	• 534,000 discharge letters
• 270,000 new patient registrations	• 185,000 mammography results
• 4,100,00 pathology requests	• 710,000 after hours reports
• 260,000 patient reports	• 2,163,000 outpatient letters
• 20,000,000 prescriptions	• 6,409,000 pathology results
• 1,300,000 referrals	• 426,000 radiology results
• 760,000 reminders	• 400,000 screening results

the largest threat to New Zealand's long term fiscal sustainability. Also, that spending in the health sector, has been the fastest growing aspect of public expenditure for many decades (twice the rate of GDP growth since 2001), and is the countries' most pressing financial challenge (OECD, 2009).

In New Zealand, efficiency gains through use of, and greater access to, health information are to be achieved through strategies and initiatives spelled out in the *New Zealand National Health IT Plan 2010* published by the New Zealand Ministry of Health. The Health IT Plan 2010, along with an update produced for 2013/2014, is the latest in a series of strategy documents that are being implemented, designed to realise the desired gains in the NZHS. The strategy documents are augmented with implementation standards produced by New Zealand's Health Information Standards Organisation (HISO).

Along with greater access to health information is a requirement that the information be used safely, without adverse outcomes for patients. An example that safe use is required comes from a report from the Institute of Medicine of the National Academies 1999. The report estimates that as many as 98,000 deaths were the result of medical errors in the United States for the year 1999 (Institute of Medicine, 1999). A proportion of these deaths is attributed to absent or faulty information regarding prescriptions, and treatments (Leitheiser, 2001). The implications of poor quality information in health is explored in detail in Section 2.7.3, and Chapter 3 Section 3.5.4.

2.3.1 Health Sector Direction - Tension for Change

For this research, aggregation of data from disparate sources into a single view at the point of care was the premise that the information use was based on. This focus

is in line with proposed initiatives for health information use within the NZHS (NZ National Health IT Board, 2010). Further, it was only interested in data for the current, required context of patient care. With the desire to make more, and better, use of health information, and information from disparate sources, at the point of care, factors regarding characteristics of the information, and its use, become important. These factors include:

- The quality of the information that is to be relied upon.
- Is there enough accompanying information (meta-data), or enough desired information, to allow for a judgement to be made as to the information quality?
- Ensuring patient safety is not compromised by a judgement to use the information (NZ National Health IT Board, 2010).
- Out-of-context use of the information, where determinations of quality are especially challenging due to GPs becoming more remote from information creation, and information sources (Fisher et al., 2003).

The last point above needs some explanation as to what is meant by out-of-context use of information. The following section provides a definition for this term.

2.4 Out-Of-Context Use of Information

With the imminent implementation of Summary Care Records and Shared Care Records within the New Zealand Health Sector, designed to be accessed by health practitioners from anywhere, the issue of out-of-context use of health information arises. For this research, out-of-context use of health information is defined as the use of a patient's health information, collected in one context e.g. a hospital's Accident and Emergency department, and used in another context e.g. a patient's primary health provider, or vice-versa. This would be information the health practitioner did not produced themselves, may not have seen before, and is being used in a clinical situation beyond what the information was originally collected for. This is especially likely, for example, in emergency and unscheduled care settings, such as an after hours clinic, or an Accident and Emergency department.

This functionality raises issues around the safe use (see Section 2.7.3 for definition) of information. Within the scope of this research project, what does a GP need to know about the health information they are accessing such that they can have confidence in its quality, and use it meaningfully for clinical decisions, without risking patient

safety? From another perspective, what is the process used by a GP to determine if the information they are wanting to use is sufficient in quality, such that they can use it safely?

An evaluation of the United Kingdom's Summary Care Record (SCR) Early Adopter Programme revealed this very concern regarding the use of the SCR when being shared by users in different contexts. The quote below highlights this:

> Most data quality standards in current use were designed for supporting individual clinicians treating individual patients. In contrast, the Summary Care Record (SCR) is intended to be shared between users in a variety of different contexts. Whilst much research has been undertaken on data quality, studies to date have not addressed the specific question of the problems that arise from interpreting inaccurate or incomplete data in cross-contextual use. (Greenhalgh et al., 2008b)

2.5 The Problem of Providing Health Information Quality Indicators

The factors from Section 2.3.1 become more prominent for GPs as National Health IT Plan strategies are implemented. GPs will be faced with access to more health information, health information they may wish to use in the clinical care context. It is not only information they wish to use, it may also be information they may wish to ignore e.g. information that contradicts what the patient is telling them. The point being that the GP can not simply accept, or reject, health information at face value. They must evaluate health information in a clinical context for its safe use. There must be justification for the acceptance, or rejection, of the information.

The issues highlighted above were the basis for the formation of the research questions given in Chapter 1, Section 1.2.1. The research questions formed the basis for the literature review. The literature review being conducted to determine if the literature provided answers to the research questions.

2.6 Literature Review Process

Information quality, and the determination of information quality, are at the core of this research. Information quality was investigated in the literature; specifically, research involved in determining factors important in information quality, and in the evaluation

of those factors. Although this was the primary focus, an overview of the information quality domain was also undertaken, to provide a rounded perspective.

To compute an information quality indicator for health information content, it requires that issues such as how users determine the quality of data be investigated. Of interest for this research, is what users want to know about the information they wish to use, that aids in their determination of quality.

A broad brush approach was taken in the first instance. This meant starting from a wide range of concepts regarding information use. Included were areas of research that helped give an understanding of what aspects are drawn into this area of study, and which helped to form a base of knowledge. This defined the scope of where this research sits, within a wider scientific context.

As the health information content must be made available at the point of care, current approaches to this issue were investigated i.e. Electronic Health Record (EHR). This was done to determine if these current approaches provide mechanisms for not only discerning the quality of that information, but also for including such indicators with the information itself. The review was important for determining what these communication mechanisms provide in the way of meta-data, that could be used to carry additional data (quality indicators) on the information content of an EHR. This included investigating relevant standards around the development of EHRs, and health information communication, as well as currently used forms of health information e.g. hospital discharge forms and general practice Practice Management Systems (PMS) record forms.

2.6.1 Terms and Definitions

A keyword search of online databases was conducted as part of the literature review.

These databases included:

- Science Direct
- Elsevier
- PubMed
- ACM Digital Library
- Google Scholar
- JSTOR

Table 2.3: Key Phrases/Key Word Searches

Key Words and Key Phrases	
Model Based Approach	**Judgement Of Quality**
• information quality evaluation • information quality dimensions • belief in information • information quality standards • information quality and electronic health records • information quality criteria • information quality computation • information quality framework • information quality and utility theory • formalisms of information quality • information quality model • information quality and information theory	• improving information quality • quality meta-data • scoring information quality • information quality metrics • calculating information quality • functions for information quality • antecedents of trust • objective information quality evaluation • subjective information quality evaluation • trust in information • confidence in information • antecedents of information quality • metric assessment methods
Clinical Decision Making Context	**Assistance To User**
• patient safety and information quality • information quality and decision making • human error and information quality • information quality and risk • risk analysis	• multi criteria decision analysis • health systems and information quality • information quality indicators • information evaluation framework • information quality and problem solving • improving information quality • operationalising information quality • scales evaluation

- Sage Journals Online
- National Center for Biotechnology Information (NCBI)
- Mendeley
- IEEE Xplore

A list of the most relevant key phrases/keywords used in these searches is given in Table 2.3. The search terms are categorised by concepts from the overall research question given in Section 1.2.1. Where the word information is used, data, health data, and health information was also substituted in the search. The search strategy involved following where the searches led until either the topic being followed moved beyond the scope of this research, or references kept returning to the same area of interest i.e. the search became circular.

2.7 Literature Review

The purpose of the review was to determine whether the research questions had been addressed. To demonstrate that the research questions were not already answered. The

review showed that there was a gap in the current literature, that required the researcher to discover further information.

The review is structured in the following way. First, an overview of the data quality concept is given. This looks at what data quality means, and how it has been defined. Following this is an explanation of why data quality is important and of value. Next, data and its place in the NZHS is examined. This highlights data volumes and flows within the NZHS. Following this is an examination of why it is important to have good quality information available within the NZHS. Next, factors that influence data quality, and others' approaches to investigating data quality, are looked at. This examines some of the major influences on research into data quality. Next, data quality, and its role in decision making, is highlighted, along with the importance of the context the data is needed for, and the decision is made in. Lastly, approaches used to determine the quality of data are explored, and conclusions are drawn regarding the rationale for conducting this research.

The discovery of further information, via an interview process with stakeholders (a group of practising General Practitioners), was required. This was because for information quality evaluation of health information, at point of care, for primary healthcare providers, there was little found in the literature. For this research topic it was required to understand what it is about health information that the GPs want to know, that leads to a sense of confidence in the quality of that information. It is important for GPs to have confidence in the quality of the health information they wish to use, as they need to be able to evaluate whether the information is safe to use. Safe in terms that should they accept and use the health information, the well-being of the patient is not compromised.

The research project utilised literature from two main domains. These were the domains of information/data quality and electronic health information communication. Additionally, literature from domains such as psychology, marketing, manufacturing, and decision analysis was also incorporated in the review, as this was where the search led when looking at the topic of data quality. Data quality is the term used here as this was the term used for these studies. This thesis however investigated health information and health information quality assessment. Please refer to Section 2.2 for a definition of the difference between the terms data and information. The review incorporates abstract concepts from one further area of research. This being around concepts of trust, confidence and belief. These concepts were seen as being an aspect of how users evaluate information. Do they trust, or have confidence, or belief in the information? The three terms, trust, confidence, and belief, are defined in this context as the positive assessment of the quality of the data, such that it is felt the data is fit for purpose,

and safe to rely on. Specifically, what is it about the data that leads to this positive assessment?

2.7.1 Defining Data Quality

In order to analyse and study data quality, it is necessary to first understand what is meant by data quality (Wang et al., 1993). As noted by other researchers, no one single definition for data quality exists, that is universally accepted by those researchers in the domain (Klein and Rossin, 1999). The user oriented view of data quality (users, in this context, use organisational data for decision making) is that of quality data being "data that are fit for use" (Wang and Strong, 1996; Loshin, 2001).

The context for use of data is important. Users define what is good quality data for each intended use of the data, dependent on the context for that desired use (Strong et al., 1997; Pringle et al., 2002; Even and Shankaranarayanan, 2007). Put another way, the quality of the information is dependent on its intended use (Wand and Wang, 1996). Users of information can use the same information, but for different tasks. Conversely, the same information can be used within an organisation by different users for different tasks (Kerr, 2006). What is considered quality data for one context may not be sufficient for another context (Wand and Wang, 1996; Even and Shankaranarayanan, 2007). For example when a customer of a bank checks their balance, data to tens or hundreds of dollars may be sufficient. For the bank itself however, precision to the cent is needed.

From this it can be seen that data quality is a multidimensional concept (Wand and Wang, 1996; Klein and Rossin, 1999), and should be studied as such. This is not least of all due to the data themselves being multidimensional (Juran and Godfrey, 1999).

For this research, the following definition for data quality, as clearly stated by Redman (2001), in turn based on Juran and Godfrey (1999), was adopted:

> Data are of high quality if they are fit for their intended uses in operations, decision making, and planning. Data are fit for use if they are free of defects and possess desired features. (Redman, 2001, Pg. 73)

This is a widely adopted definition (Strong et al., 1997). The "possess desired features" part of the "fit for use" definition was of special interest, as one of the core components of this research was determining what the desired features (criteria) for discerning quality health information were.

2.7.2 Why Data Quality Is Important

Within information systems, data quality is of critical importance (Even et al., 2006). Poor data quality can have a deleterious effect on an organisations effectiveness (Wand and Wang, 1996), with information of good quality required for effective operations and decision making (Price and Shanks, 2005). This is true within the health domain. For example, significant issues were found when looking at quality of data in EHRs, for the purpose of deriving EHR based quality of care measurements. Issues with the data included data accuracy, completeness, and comparability (Chan et al., 2010). Tang et al (2007) looked into discrepancies in diagnosis of diabetes in the United States. The discrepancies in diagnosis were found when differing types of information, for the same EHR, were examined. When administrative data (billing) was used, only 75% of diabetics were identified. When coded information (clinical data) from within the EHR was used, 97% of diabetics were identified (Tang et al., 2007). This illustrates the impact data quality can have when using health information for clinical purposes other than the original consultation.

Data quality has been identified as important within other domains. Perceived Information Quality (PIQ) was investigated as a factor in the success of inter-organisational data exchanges, and intention to use the resulting information (Nicolaou and McKnight, 2006). Within the software development domain it is proposed that data quality should be an early activity, and not something additional to the system. Data quality should be a feature of the system, not only in development, but also in production (Buccella et al., 2008). The study by Buccella et al (2008), also highlighted that data quality is strongly related to the use of data. These last two statements are core assumptions for this thesis.

Further research in the domain of business looked at the currency (time currency) of data in relation to sales and Customer Relationship Management (CRM) campaigns. It was found that sales volume was insufficient as a customer selection criterion, and that the currency of the data was important, as customers could be selected that were no longer eligible to accept offers (Heinrich et al., 2009). Hill, in his PhD thesis (2009), focused on the quality of customer information and developing a framework that supported business functions such as marketing campaigns, credit scoring, customer service, and management decision making. (Hill, 2009).

Klein and Lehner (2009) looked at data capture in sensor data streaming environments, where the issue of efficiently providing applications with data quality information, derived from sensor data streams, was at issue (Klein and Lehner, 2009). They proposed a model for the processing and propagation of data quality, to evaluate the impact of data quality, to reduce incorrect business decisions.

These studies show data quality is an issue in many domains, including health. For a generalised example, on the database field level, for a database where quality is not being addressed, the error rate is 1% to 5% (Redman, 1997). Within the health domain, examples of accuracy errors are seen in a study of cardiac risk assessment where 28% of 500 patient charts identified errors in patient age, gender, blood pressure, HDL cholesterol medications or smoking status (Persell et al., 2009). A study by Powell et al, (2006), found that of 31 patients, 6 patients confirmed errors in dates, or clinical information in their records (Powell et al., 2006).

Another example showed discrepancies with accuracy within individual records; where of 97 patient encounters, 55% had active pain documented in the problem list, or in free text, but "no pain" documented in the data entry template (Saigh et al., 2006). While these last few studies into data quality have been in the domain of health, their focus has not been on judgements of data quality. They looked into data quality for use in producing quality of care measures, or into quality aspects such as accuracy or completeness of the data. For example, a study by Hasan and Padman (2006), where they proposed a methodology for assessing a system's (decision support registries) vulnerability to data quality issues (Hasan and Padman, 2006).

2.7.3 Data Quality And Health

Access to relevant health information by health practitioners at the point of care has long been a goal of health providers around the world, including New Zealand (Kirk, 1967; NZ National Health IT Board, 2010). This is making health information regarding a patient available, electronically, to a clinician outside of the context in which the information was gathered, so that the clinician can use that information when seeing the patient. Health practitioners in an out of hours clinic in the UK feel that assessing patients without being able to refer to patient's medical records is akin to "stabbing in the dark" and risked patient's well being and safety (Greenhalgh et al., 2008a).

An Electronic Health Record (EHR) is a longitudinal view of data collected from several sources (Moller and Vosegaard, 2008), with the idea that this can then be used by a clinician at point of care for assessing/treating patients. This is a broad definition and deliberately so, as for this research it is undesirable to specify an actual record type, but rather to refer to electronically collected and stored health information about a person.

The perceived usefulness of an EHR, that of communicating health information, has long been recognised, as can be seen from the quote below from Florence Nightingale. Although in Florence's case it would just be Health Record.

> In an attempt to arrive at the truth I have applied everywhere for information but in scarcely an instance have I been able to obtain hospital records fit for any purpose of comparison. Florence Nightingale, 1863.

More recently in his 1965 paper, James Kirk highlights this requirement as one of 20 recommendations (Kirk, 1967). The health arm of the National Academy of Sciences, The Institute Of Medicine, stated in 1991 that an EHR, or similar, is vital technology for healthcare (Dick et al., 1997). Some attempts made in producing EHRs, and architectures and standards for creating EHRs, are as follows:

- **ISO 18308**: the International Organisation for Standardisation's (ISO) 18308, Requirement for an Electronic Health Record Architecture, is a standard developed to provide a universal set of technical and clinical requirements that support using and exchanging Electronic Health Records (ISO, 2008b).
- **ASTM Continuity of Care Record (CCR)**: a standard developed by the American Society for Testing and Materials (ASTM) International. The standard specifies a core set of data objects that captures the most relevant clinical information, allows it to be aggregated, and forwarded to another practitioner or setting, for supporting the continuity of care (ASTM, 2005).
- **openEHR**: is a set of specification for implementing a full EHR. This specification is based on, and fully conforms with, ISO 18308. At its core is a formal constraint based content model the defines content as Archtypes. These archtypes are used to describe instances of clinical data, captured in an openEHR record (openEHR Foundation, 2013b).
- **HL7 CDA/CCD**: Health Level 7 (HL7) is an American National Standards Institute (ANSI) approved standards development organisation. HL7 first produced health information messaging standards only. The Clinical Document Architecture (CDA) standard was produced later, and defines the "structure and semantics" for a clinical document (EHR) (Health Level 7 International, 2013c). The Continuity of Care Document (CCD) standard, a collaboration between HL7 and ASTM, is a type of CDA document, that is constrained as to its contents by ASTM's Continuity of Care Record (CCR) standard (Atalag et al., 2009).
- **NHS Summary Care Record (SCR)**: the United Kingdom's National Health Service (NHS) Summary Care Record (SCR) is a centrally stored summary of a person's health, created from the person's general practitioner (GP) record. Three areas of a person's health are included, these being medication details, allergies,

and adverse reactions. The SCR is intended to be used when "other records are unavailable or incomplete" (Greenhalgh et al., 2010), such as in emergency and unscheduled care situations. It has had a controversial implementation, and its future is in doubt.

- **New Zealand Shared Care Record (SCR)**: is a summary of personal health details such as current medications, current diagnosis, problem list, alerts and allergies (NZ National Health IT Board, 2010). It is an upcoming initiative, planned for the whole New Zealand Health Sector.
- **Electronic Shared Care Record View (eSCRV)**: was developed out of a need identified with the 2011 Christchurch (Canterbury region), New Zealand, earthquake. It allows health practitioners in the Canterbury region to access patient information from a single sign on. Implementation is under way.

These examples were chosen as they all deal with electronically collecting and transmitting health information. Some are more focused on health information messaging e.g. HL7, while others are more focused on clinical (data) content e.g. ASTM's CCR. They are introduced here to give some idea of what EHRs are, and how attempts have been made to implement them. All examples given are able to be used as Electronic Health Records, or specify requirements that an Electronic Health Record should fulfil. They also all have the goal of semantic interoperability, where information can be exchanged freely between systems for clinical purposes, with faithful reproduction, and without loss of information. The important point for this thesis is that electronically storing health data for a patient, and having the ability to exchange that data with other health practitioners for the purpose of treating the patient, has become a driving force in health care. This includes the NZHS.

This functionality raises issues around the safe use of the information. By safe use it is meant that by using the information in clinical treatment, a patient's safety is not compromised. This fits well with a commonly accepted definition for data quality. This being that data is of high quality if it is fit for its intended use. This definition is discussed in detail in Section 2.7.1. A clinician needs to have confidence in the quality of the information, so decisions made based on the information are safe i.e. do not result in an adverse outcome for the patient, due to the information being faulty in some way. This reflects the New Zealand Health IT Plan (2010) desired outcome, that there be some form of accessible electronic version of a person's health information that would serve as the "go to" point for information on a person's health. This is referred to as the "One source of truth".

For meaningful, safe use of health information at the point of care, GPs must determine

if the information is fit for its intended purpose. Assessing the quality of the information GPs wish to use is part of this process. Dimensions for data quality has been one avenue for investigating the assessment of data quality, and is explored in the following section.

2.7.4 Data Quality Dimensions

Data quality has been identified as being composed of many dimensions (Raghunathan, 1999; Fisher et al., 2003) e.g. completeness, accuracy, consistency, timeliness, relevance (Ballou and Pazer, 1995; Wang and Strong, 1996). While the fact that data quality can be composed of multiple dimensions has been established, there is no general agreement, or accepted set of definitions, to describe these dimensions (Wand and Wang, 1996).

Wang et al (1996) began the approach of looking at data quality from the data consumer's perspective (Wang and Strong, 1996). They used survey instruments to determine from data users what dimensions the users considered to be important for data quality. They used these dimensions to develop a framework of data quality. Once dimensions are identified, attempts can be made to measure data quality along those dimensions (Watts et al., 2009).

The measurement (evaluation) of quality dimensions can be an objective or subjective exercise (Naumann and Rolker, 2000; Lee et al., 2002; Batini et al., 2009; Watts et al., 2009). For example, accuracy and completeness are more objective measurements (Watts et al., 2009) where the dimensions are intrinsic to the data itself, rather than to the context it is needed for. While a measure of relevance may vary with usage i.e. subjective measure.

Research has shown that methodologies for investigating quality dimensions fall into two categories: general-purpose and specialised (Batini et al., 2008). General-purpose approaches deal with a range of quality dimensions and are domain independent. Specialised approaches are focused on a particular data quality task, for a specific type of data, and within a particular application domain.

Using abstract quality dimensions to score information quality was looked at as an approach for this thesis. However, this was not pursued as it was determined GPs would not, and could not, just accept a score for quality for health information. This decision was arrived at after the interviews conducted with the GPs. The reasoning and justification for not applying this approach is explained in detail in Chapter 4, Section 4.4.1.

In Tables 2.4, and 2.5, examples of dimensions for quality found in the literature are given. The vertical axis list dimensions alphabetically while the horizontal axis identifies the researchers and research papers that study the dimensions.

Table 2.4: Data quality dimensions found in literature - Part 1

Quality Dimension	Boehm and In 1996	Boxwala et al 2004	Brown and Sonksen 2000	Burgess et al 2002	Dondio and Barrett 2007	Funk et al 1998	Gil and Artz 2007	Jang et al 1995 (reasoner)	Jarke and Vassiliou 1997	Lee et al 2002	Luhmann 2000	Mayer et al 1995	Nicolaou and McKnight 2006	Pipino et al 2002	Pirzada and McDonald 2004	Strong et al 1997	Vassiliadis et al 2000	Wang and Strong 1996	Wang 1998	Woudstra and van den Hoof 2008
Ability												X								
Access security																		X		
Accessibility				X					X	X				X		X		X	X	
Accuracy		X	X	X					X				X			X	X	X		
Aesthetics				X																
Age of data							X													
Appropriate Amount of Info						X				X				X		X		X	X	
Authority							X													
Availability	X			X																X
Believability									X	X				X		X		X	X	
Benevolence												X								
Cognitive effort																				X
Coherence																	X			
Compatability				X																
Completeness		X	X	X					X	X			X	X		X	X	X	X	
Comprehensiveness	X																			
Concise Representation						X				X				X		X		X	X	
Consistency		X	X	X					X				X				X			
Consistent Representation						X				X				X		X			X	
Context & Criticality							X													
Control Transparency							X						X							
Convenience				X																
Correctness	X																X			
Coverage				X																
Credibility								X	X								X			
Currency (temporal)				X					X											
Data association							X													
Data forwarded															X					
Data received															X					
Delivery Time				X																
Durability				X																
Ease of Manipulation/Operation						X				X				X					X	
Ease of understanding														X		X		X		
Ease of use				X																
Efficiency				X																
Evolvability	X			X																
Familiarity											X									
Features				X																
Fit for intended use				X																
Flexibility	X																			

Table 2.5: Data quality dimensions found in literature - Part 2

Quality Dimension	Boehm and In 1996	Boxwala et al 2004	Brown and Sonksen 2000	Burgess et al 2002	Dondio and Barrett 2007	Funk et al 1998	Gil and Artz 2007	Jang et al 1995 (reasoner)	Jarke and Vassiliou 1997	Lee et al 2002	Luhmann 2000	Mayer et al 1995	Nicolaou and McKnight 2006	Pipino et al 2002	Pirzada and McDonald 2004	Strong et al 1997	Vassiliadis et al 2000	Wang and Strong 1996	Wang 1998	Woudstra and van den Hoof 2008
Free from Error						X				X				X					X	
Integrity	X											X								
Interoperability	X																			
Interpretability						X		X	X	X				X		X	X	X	X	
Leadership					X															
Likelihood							X													
Maintainability	X																			
Maintenance				X																
Modifiability	X																			
Neutrality					X															
Objectivity				X		X				X				X		X		X	X	
Origin									X											
Performance	X			X																
Perspective																				X
Physical proximity																				X
Presentation				X																
Provenance							X													
Quality Assurance				X																
Quantity				X																
Recency							X													
Recommendation							X													
Relevancy				X		X			X	X				X		X	X	X	X	
Reliability	X			X																X
Representational consistencey																		X		
Reputation				X						X				X		X		X	X	
Reusability	X																			
Safety	X																			
Saves time																				X
Scalability	X																			
Security	X					X				X				X		X			X	
Stability					X															
Timeliness		X	X			X		X	X	X			X	X		X		X	X	
Topic of knowledge																				X
Understandability						X				X									X	
Updating				X																
Up to dateness																				X
Usability				X																
Usefulness									X								X			
Value Added						X								X		X		X	X	
Verifiability	X																			

Table 2.6: Example research utilising data quality frameworks.

	Knowledge based approach (stakeholder involvement)	Identifies/uses quality dimensions	Identifies quality metrics	Computational reasoning over data quality parameters	Stakeholder input required to work	Model or framework defined	In health domain	Quality link to trust and/or confidence
Boehm and In	yes	yes	yes		yes	yes		
Burgess et al	yes	yes		yes	yes	yes		
Funk et al	yes	yes			yes			
Hertzum	yes	yes	yes					yes
Jadad and Gagliardi		yes					yes	
Jarke and Vassiliou	yes	yes*	yes			yes		
Lee et al	yes	yes	yes		yes	yes		
Vassiliadis et al	yes	yes	yes	yes	yes	yes		
Wang	yes	yes	yes	yes	yes	yes		
Ware and Gandek								
Wilson	yes				yes		yes	
Woudstra and van den Hoof	yes	yes						

2.7.5 Data Quality Frameworks

The past two decades have seen data quality research progress significantly (Madnick et al., 2009). Many efforts have produced data quality frameworks that describe data quality requirements for the investigated contexts/domains. These frameworks have been used to understand and manage data quality issues in real world environments (Wang and Strong, 1996). One of the most prominent is the Total Data Quality Management (TDQM) frameworks developed at the Massachusetts Institute of Technology (MIT) Sloane School of Management in the 1980s. TDQM has become a programme of study in itself at MIT (Massachusetts Institute of Technology, 2011).

Earlier frameworks were tied to a specific context, or domain, where data quality was an issue that needed addressing. In more recent efforts, attempts have been made to construct general frameworks capable of being applied to many different settings (Stvilia et al., 2007). From this have been the creation of a hybrid framework where general dimensions of data quality are first applied, followed by additional dimensions from the specific area of investigation (Burgess et al., 2007).

Table 2.6 shows research found during the literature review that produced data quality frameworks (highlighted column). The table also shows other methods incorporated with the frameworks, from requiring stakeholder input to function, to identifying metrics for dimensions.

Producing data quality frameworks, to help manage data quality in differing environments, is a successful and active strategy. Frameworks have been used to identify data quality requirements and formulate ways of applying strategies to understand, manage, and improve data quality. For example, the Canadian Institute for Health Information (CIHI) developed, and successfully implemented, an integrated framework to

measure data quality in health (Orfanidis et al., 2004). The CIHI framework assesses a database along five quality dimensions: Accuracy, Timeliness, Comparability, Usability, and Relevance. Each dimension is broken down into related characteristics, and each characteristic is divided into several criteria. A manual data assessment tool is used to produce a Data Quality Assessment Report, that can be used to develop an action plan to address quality issues found in the database (CIHI, 2009).

Specialised frameworks are not transferable to the domain of study for this thesis, as they are developed for specific tasks, in a specific context. General frameworks, such as CIHI, are focused on data quality of collections, or for managing acquisition of data that is of good quality. The frameworks were not developed for the purpose of assisting the data user in terms of assessing the quality of data at time of use.

2.7.6 Data Quality and Decision Making

Effective decision making is influenced by the quality of data used (Fisher et al., 2003; Poeppelmann and Schultewolter, 2012) i.e. poor quality data can have an adverse impact on decision making (Price and Shanks, 2008). To make the decision, an assessment by the decision maker as to the quality of the data to be used, is an important task. This is in-line with the definition given in Section 2.7.1, with data of high quality being data that is fit for its intended use, based on a specific context. It is in the hands of the decision maker to decide whether the information is fit for use, by evaluating this based on the context it is needed for (Even et al., 2006).

When using information in decision making, it is important that the decision maker is assured of the quality of the information they wish to use, as well as providing for the decision maker to be able to assess the quality of the information (Shankaranarayanan and Cai, 2006). In healthcare, this is especially important as the need for effective decision making is high (Kerr et al., 2007). As more information is made available to GPs, supporting GPs with assessment of health information quality becomes desirable. Indeed, with the requirement for health information safe use it is a vital requirement.

2.7.7 Assisting Information Quality Determination

Users of information are charged with assessing the quality of the information and whether the information is fit for use for the current task/context. Quality has become an important factor for users in deciding which information to use, and which not to use (Burgess et al., 2004).

Information that can assist the assessment of the quality of data can be stored with

information as metadata (Chengalur-Smith et al., 1999; Fisher et al., 2003; Even and Shankaranarayanan, 2007). Metadata is abstracted data about data (Shankaranarayanan et al., 2006; Kerr, 2006). Metadata of this type is referred to in the literature as quality metadata (Shankaranarayanan et al., 2006) and can be defined as quality measurements, linked to information used by decision makers (Watts et al., 2009). This linking process has been referred to as data tagging.

Problems occur where users are unaware of the context of data collection and assumptions are made about the quality of data without accompanying, or insufficient, data quality information (Kerr, 2006). This places emphasis on the use of meta-data elements for making decisions on data, specifically how standardised meta-data, and an explicit meta-data model, can increase the usefulness of the data it describes. This is especially important when users are further removed from any personal experience with data creation where knowledge of benefit in discerning the appropriateness of the data has been lost (Fisher et al., 2003).

The use of meta-data, to improve interpretation of the underlying information, has relevance in this research as it is the meta-data elements that will be used to describe quality indicators in the accompanying information. More specifically, the quality meta-data elements could be used to indicate a quality criterion found in the health information.

The computational approach to determining data quality levels is not well represented in the literature. An example found was the implementation of a first order data quality reasoner that utilises a reduction algorithm over a set of quality parameters (dimensions) that have been organised into a dominance hierarchy of relationships (Jang et al., 1995). The data consumer queries a data item wanting to know the value for a particular quality dimension e.g. believability. Other parameters are known, and attached to the data item e.g. credibility of the source of the data item has a value of "high". These quality parameters are arranged into a dominance hierarchy. The algorithm essentially walks down the hierarchy of dominance relationships, subsuming those dominated parameters, until they can not be reduced further, to produce an overall quality value that is then assigned to believability. Further development of this model was not found, and following research Wang et al. (2003), a co-author of the original 1995 paper, exhibits the same first order data quality reasoner.

Another computational approach was devised for assessing the quality of data entering the United Kingdom's General Practice Research Database (GPRD) (Tate et al., 2011). As the GPRD is used variously by governments, academics, and the pharmaceutical industry, it is critical that the data is of high quality. In this study they assessed data from individual general practices using statistical pattern recognition techniques to determine quality scores for each practice. It was the general practice that was scored

for producing quality data, rather than the data itself.

Much of the data quality research investigated has had this goal, that of measuring, or scoring data, and assuring that data is of high quality before being stored e.g. (Tate et al., 2011). Alongside this has also been the goal of managing data collections to ensure that data being held remains of high quality e.g. Kerr (2006). However, since the acceptance that data quality should be approached from the data consumer's perspective, efforts have been made to assist the data consumer in determining the quality of the data they wish to use (Wang and Strong, 1996).

2.7.8 Conclusion

With Wang and Strong (1996), attention began to focus on the data consumer as the arbiter of what was quality data. The data consumer would decide, based on the given context, as to the fitness for use of the data. This meant that the data consumer would determine the quality of data, based on features of the data important to the user.

Understanding what data consumers deem important to know (quality criteria) about information when discerning its quality, and therefore fitness for use, has become more important as information systems allow data consumers to access increasing quantities of information from more and more diverse sources (Prat and Madnick, 2008). Health practitioners are no different in this respect, as health organisations strive to make better use of health information and to increase connectedness of health information systems. This is driving the creation and implementation of health standards, and the desire to have electronic health information available for use at the point of care. This may be achieved in New Zealand via efforts such as Summary Care Record, Shared Care Record, Electronic Shared Care Record View, or some other implementation of an Electronic Health Record.

With this approach, criteria for quality, important to the GP, for the current context, can then be explored. With these criteria identified, this then allows for identifying and highlighting these to the GP. The discovery and highlighting of quality criteria deemed important to know by GPs is at the heart of this research.

The literature was examined to determine if prior work by others contained enough detail regarding what criteria clinicians use to assist them in determinations of quality. This was done to determine if the research questions posed in Section 1.2.1 had been addressed. The result of the literature review showed that the research questions had not been addressed and they would need to be answered by moving to a new phase of this research project.

This had two consequences. First, the criteria for quality from outside the General

Practice domain, needed to be investigated to determine if they were applicable for use within the General Practice domain. Second, GPs would need to be approached to determine what criteria for quality were specifically important to them when determining information fitness for use i.e. the information quality. This required a move to a new phase for the research, where these two points would be addressed. This phase would fill the gaps found after conducting the literature review.

2.8 Chapter Summary

In this chapter the rationale for the research has been given, along with a review of relevant literature. The background and areas of research that were important for addressing the problem under investigation were highlighted and discussed. This provided a crystallisation of the problem and identified a way that the research questions could be addressed.

Chapter 3

Theoretical Framework

3.1 Chapter Overview

Further investigation of research questions was needed. The following chapter deals with how this task was approached, and completed. Analysis and interpretation are given for interviews carried out with a group of practicing GPs. Interviews were conducted using semi-structured interviews, incorporating vignettes, with individual GPs from the group. At the end of each interview the GP would also be asked to rate a list of identified criteria for their importance in knowing, for providing an indication as to the quality of health information. What emerged from the interview process was that the GPs interviewed used tacit knowledge e.g. experience, and community knowledge, to evaluate the quality of the information available to them.

It was found that there were additional criteria that could be made available with the health information that would enhance the GPs ability to discern the quality of the health information. Criteria elicited from the interview process, and from the rating exercise, were then mapped to Quality Dimensions for information quality. The criteria, with their associated dimensions, were then placed into categories to develop a theoretical framework. The categories for the framework were the class of information the criteria are used with e.g. the criterion "data age" was placed in the category Contextual, as the class of information this criterion is concerned with is placing the health information within a temporal context.

3.2 Introduction

Quality criteria for health information are defined in this research as a general practitioner's (GPs) preference for what they want to know about health information they wish to use, that enhances the ability of the GP to determine the quality of the information. This can be described as the "importance in knowing" these criteria when assessing health information. This is for out-of-context use of the information i.e. information the GP did not create themselves and that they have not seen before. In determining the quality of the information, the GP develops a sense confidence as to if they can use the information safely i.e. the information is fit for purpose. These criteria needed to be identified, and a determination of how important each was to the GP for determining the quality of health information.

Some quality criteria specific to health information were found in the literature. However, the quality criteria found were for quality of data collections, the production of quality data, and producing standard documentation for data collections e.g. the Canadian Institute for Health Information Data Quality Framework (CIHI, 2009). This is different from the focus of this thesis, where the interest was in quality criteria used for assessing the quality of health information at point of care, without the GP knowing how the information was produced, stored, or communicated to them. When referring to quality criteria (QC) throughout this thesis, this is what is meant i.e. criteria used by a GP in determining the quality of information.

3.3 Candidate Quality Criteria

Existing standards pertaining to what an Electronic Health Record (EHR) should consist of were investigated. They were investigated for their specific inclusion, or not, of requirements that would indicate to the GP the quality of the health information i.e. inclusion of quality criteria. The investigation revealed requirements that must be met for an EHR to be faithful to the needs of health care delivery, valid and reliable in a clinical sense, ethically sound, and meet any medico legal requirements. Also required are faithful reproduction upon communication, while maintaining confidentiality requirements (ISO, 2008b,a, 2005; ASTM, 2005). The standards organisation Health Level 7 International (HL7) has produced several standards for health information capture, presentation and messaging e.g. HL7 v2 and v3 for messaging, and HL7 Clinical Document Architecture (CDA) for clinical documents (Dolin et al., 2001, 2006). These standards have been used to implement Electronic Health Records e.g. openEHR is an Electronic Health Record specification based on the requirements set out in ISO 18308

and conforms to this standard (openEHR Foundation, 2013b). The standards have also been used to implement parts of an Electronic Health Record, or for Electronic Health Record transmission (messaging) e.g. the HL7 v2 messaging standard is in widespread use (Atalag et al., 2009).

Currently used Electronic Health Records (EHRs) were investigated. These included examples such as Microsoft's HealthVault and Google Health (now discontinued (Google, 2013)). These are Personal Health Records (PHRs) designed to allow users themselves to store health data about themselves online, via a health web portal. The researcher also visited his own general practitioner (GP) and had his medical record printed. In this case the record was produced from a Practice Management System (PMS), MedTech32 from MedTech Global. In New Zealand, MedTech32 has >80% market share for GP PMS's (MedTech, 2012). As with the standards, these examples were investigated for their absence or presence of information quality criteria requirements.

Literature from domains other than healthcare were also investigated. The focus was on studies into the use of information from the perspective of what users of that information use to decide the quality of the information. The domains for these studies were diverse and many, with the main ones listed here:

- Data/Information quality
- Psychology
- Marketing
- Decision Making
- Multi-Criteria Decision Analysis (MCDA)
- Risk Management
- Trust/Confidence/Belief in information
- Information Science

Little was found regarding criteria used for judgement of quality of information contained within an EHR. It was this lack of quality criteria that necessitated a study to discover criteria from the GPs themselves. Criteria that were identified, that were thought could be important to know when judging quality, were developed from sources such as standards for EHRs (both New Zealand and international), in-use health records (e.g. hospital discharge summaries and records from GP Practice Management Systems), and from the literature.

The developed criteria were labelled candidate quality criteria and compiled into a list. The list was used for a rating exercise, conducted during the interview phase of the research, to determine if a GP would find them important when judging quality. Sources used in developing the candidate quality criteria are summarised in Figure 3.1.

The criteria were developed by looking at the content and meta-data used for representing health information from these sources. For example, a standard that required an entry in a record be attested to was developed into "Attestation of entries recorded", that for this research, meant that would knowing that attestation had been recorded, be important to a GP when judging quality of the health information.

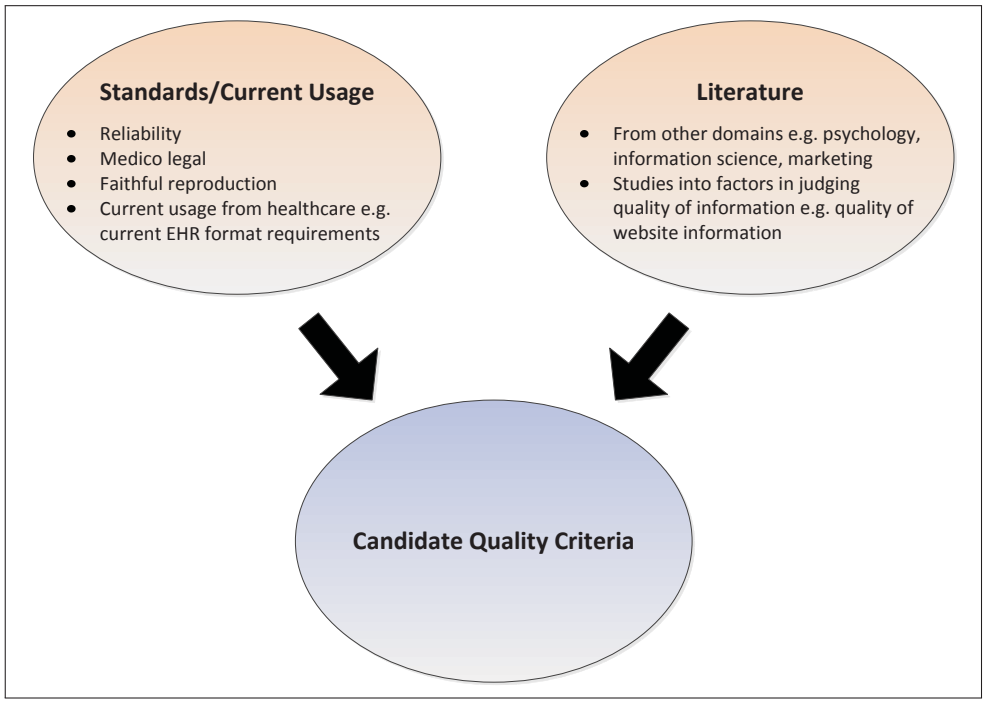


Figure 3.1: Candidate Quality Criteria

3.4 Formalised Research Question

The contextual setting for this research is the GP-patient encounter where the GP wishes to use patient health information. This is information they themselves did not collect, away from where the information was originally gathered, used and recorded. In this situation, how can the GP decide that the information they wish to base clinical decisions on is safe i.e. fit for purpose?

The consequences of a health practitioner basing a judgement on unreliable information can and does contribute to adverse outcomes, and deaths, for patients (Institute

of Medicine, 1999; Leitheiser, 2001; Fichman et al., 2011). It also contributes to unnecessary medical procedures (doubling up), extra/longer stays in hospital, and increased costs (Classen et al., 1997). This is wasteful of resources, and creates inefficiencies.

The overarching question that encompasses this topic is therefore:

> What can a model based approach provide the health practitioner, when using electronic health information, within a clinical decision making context, such that the health practitioner is assisted in the judgement of the quality of the health information?

This in turn can be broken down into research questions that are more direct and tangible. These questions are:

1. What criteria do GPs use when evaluating quality of electronic health information?
2. What information (meta-data), that currently accompanies electronic health information, can determine these criteria?
3. What additional information (meta-data), that is not currently available with electronic health information, could determine these criteria?
4. If present, are the criteria able to be detected by a system?
5. If detected, are the quality criteria able to be modelled with the electronic health information and indicated to the clinician?

The interview process was designed to answer questions 1 through 3.

3.5 Determination of Health Information Quality Criteria

Gaps in the literature, currently used forms of EHRs, and standards for producing EHRs, required that GPs be interviewed. This was to determine if quality criteria found from outside the domain of General Practice were applicable for GPs and to determine those quality criteria that were specifically important to GPs. The interviews, combined with the two sources shown in Figure 3.1 provide a triangulation of source for quality criteria.

3.5.1 Interview: Aims and Objectives

The objectives for the interviews were fourfold. First, what criteria did the GPs apply when looking at health information, and how did the GPs use those criteria to evaluate the quality of information when determining fitness for purpose. Second, what was currently available with the health information that helped them to discern these criteria. Third, what additional information, if any, that could be made available to them would help discern these criteria. Last, to obtain ratings of importance for the candidate criteria put to them in the rating exercise portion of the interview.

The objective was to identify those criteria which helped the GP to discern the quality of the information they wished to use. The aim of the rating exercise was to have a GP rate the candidate criteria to determine those most important when used in determining quality of health information. The rating exercise would be used to aid in the later use of criteria i.e. in the development of a prototype system that would produce quality indicators from quality criteria.

3.5.2 Interview: Methodology

An interview process with stakeholders was decided on as the best method to fill the gaps in information found after the literature review. Interviewing is a proven qualitative technique for looking in-depth into a subject, to gain more detailed knowledge about it, or for investigating the subject beyond surface appearances (Wengraf, 2001). Qualitative interviews have been loosely differentiated into three categories: unstructured, semi-structured, and structured (Miller and Crabtree, 1999). For this research the semi-structured interview format was adopted, as this type of individual in-depth interview is widely used in healthcare research and is able to inform a broad variety of research questions e.g. “What are general practitioner’s attitudes towards patients who are smokers that impact on quality of care?” (DiCicco-Bloom and Crabtree, 2006). This being a very similar type of question to those in this thesis. For example, from Section 3.4 above, question 1: “What criteria do clinicians use when evaluating quality of health information?”.

As the interviews investigated quality evaluation of health information at point of care, clinical contexts were required to place the GPs in those situations. The requirement for context was addressed by the use of vignettes. Finch (1987), defines vignettes as:

> “short stories about hypothetical characters in specified circumstances, to whose situation the interviewee is invited to respond.” (Finch, 1987)

Vignettes can be a valid measure of clinician response to actual clinical encounters with patients (Peabody et al., 2004a). Vignettes assist in overcoming two major issues in interviewing general practitioners to do with clinical situations, and their response to them. First, using vignettes overcomes the practical issue of obtaining actual medical records to use for example situations (Peabody et al., 2004b). Second, vignettes overcome the ethical issues that may restrict GP responses, should actual clinical situations be used, or where sensitive topics are being addressed e.g. schizophrenia. Vignettes also offer the advantage of providing a uniform collection of scenarios that provide a focused response base (Hughes and Huby, 2002).

Vignettes have been used in the area of clinical decision making, where decisions are characterised by uncertainty and complexity (Skaner et al., 2004). Skaner et al (2004), used vignettes, as well as actual cases, in a study of Clinical Judgement Analysis (CJA) of how general practitioners diagnose heart failure. While this is not judgement of health information quality, it does show the usefulness of vignettes for placing general practitioners into a clinical situation for interviewing purposes.

3.5.3 Interview: Participants

Participants were selected from a pool of General Practitioners (GPs) that are rostered in a medical clinic that provides after hours services. Participants were contacted via letter inviting them to participate. However, due to a low response rate, a snowball sampling methodology was employed to garner further participants. The snowball sampling methodology consists of using those who have agreed to participate, to refer the researcher on to others who could participate. It has advantages with certain sampling populations such as those socially isolated, elites, or other hard to reach populations (Atkinson and Flint, 2001). A total of six practising GPs were recruited in this way.

The pool of GPs were targeted as they deal with emergency and unscheduled care situations, where they are likely to have to use information that they have not collected themselves and have not seen before i.e. out-of-context use of information. See Chapter 2 section 2.4, for definition of out-of-context use.

Demographic data for each GP was collected before each interview. A form was filled in with data on age, gender, last completed year of training, number of years in practice, and whether the GP had completed several years additional training to become a vocationally registered GP. Two other pieces of data were collected. This was GP comfort with using Information and Communication Technology (ICT), and their Competence with using ICT. Each of these were self assessed, using a scale from 1 to 10. A rating of 10 indicating the GP felt extremely comfortable using ICT, and a 1 indicating they

Table 3.1: Interviewed GP's demographic data.

Interview Participants	Demographic Data					
	Dr. A	Dr. B	Dr. C	Dr. D	Dr. E	Dr. F
Age Bracket	60 - 69	40 - 49	50 - 59	50 - 59	50 - 59	50 - 59
Gender	M	M	M	F	F	M
Latest Completed Training Year	1976	2010	1979	1985	1983	1984
Number Of Years In Practice	25	23	31	28	20	26
Comfort With ICT	9	10	8	10	7	10
Competence With ICT	9	10	7	10	5	7
Vocationally Registered	Yes	Yes	Yes	Yes	Yes	No

were not comfortable at all. The demographic data is shown in Table 3.1.

3.5.4 Interview: Instrument Design

The lack of criteria for quality in health information from the literature necessitated an investigation into what criteria were preferred by general practitioners (GPs) when discerning the quality of health information. There was a need to discover if those candidate criteria taken from standards, in use health information, and other domains within the literature, were important to GPs in aiding judgements of quality. An interview process, using vignettes, and conducted with stakeholders (general practitioners), was decided on as the appropriate way to fill this gap in understanding.

The type of health information, and the context for its use formed a major focus of the research. For this reason the interview questions, and associated vignettes, were focused on usage situations that reflected this out-of-context use. It was important to place the GP into a specific usage situation, and then provide questions that were open ended and designed to elicit the way they used health information in that situation. This was to place the GP in the context defined for this research, that being use of health information at point of care.

To satisfy these requirements, three categories of health information known to commonly form parts of any potential Electronic Health Record (EHR) were used. These areas were:

1. Adverse events
2. Medical history
3. Medicines list

These areas were chosen for the following reasons:

- a GP uses information from these areas of an Electronic Health Record regularly.
- this type of health information is collected on a regular basis.
- these types of health information are presented regularly to a GP, outside of the context of their original use.
- for portable, highly accessible forms of Electronic Health Records, such as the United Kingdom's National Health Services Summary Care Record and New Zealand's Shared Care Record and Summary Care Records, this is the type of information they will contain.

An example vignette and associated questions, best illustrates this approach. For Vignette 1 the clinical scenario was simple, but common to GPs, and was taken from the first category, Adverse Events. Vignette 1:

> "A patient presents with symptoms indicating an ear infection. His medical record shows an allergy to penicillin."

An allergy to a drug is an adverse event (adverse drug reaction) and is an effect deemed as noxious, injurious, or harmful to the patient, and should result in withdrawal of the drug from the patient (Lazarou et al., 1998). For example anaphylaxis where the patient can stop breathing due to throat swelling. The same study by Lazarou et al (1998) found that adverse drug reactions were responsible for 106,000 (95% confidence interval, 76,000-137,000) deaths in hospital patients in the United States in 1994. The study stated that this made adverse drug reactions the fourth to sixth leading cause of death. An indication of an allergy to penicillin on a patient's Electronic Health Record must be considered carefully.

Question 1 for this vignette was:

> "How do you use this information when deciding on a potential treatment?"

This was followed by a variation to the vignette. The variation was:

> "When questioned regarding the allergic reaction, he describes symptoms that, while a reaction, make you think that it was not an allergic reaction i.e. more likely a side affect."

Followed by the question:

> "How do you consider this information?"

A side effect is different from an adverse event. Examples of a side effect are the patient felt nauseas, or had a rash. The variation was designed to make the GP think on the Electronic Health Record they are looking at, and what would affect their judgement of that information e.g. what do they want to know about the entry "allergy to penicillin".

In summary, this approach is similar to the seminal work into data quality of Wang and Strong (1996), where they followed methods developed in marketing research for determining the quality criteria for data, by rationalising that users of data are identifying attributes (criteria) of data quality that are important to data users (Wang and Strong, 1996).

3.5.5 Interview: Data Collection

Ethics approval was sought, and gained, from the Central Regional Ethics Committee. The application included the Interview Information Sheet, Participant Consent Form, Question List, and Rating Exercise Form. Copies of these can be found in the Appendix A and Appendix B.

Interviews were conducted with six GPs at locations suggested by the GPs and averaged one hour in duration. Interviews were audio recorded with the interviewer taking additional notes. The researcher was the only interviewer, which meant less risk of inconsistencies across interviews.

The first interview conducted was used as a pilot for following interviews. The pilot interview allowed the interviewer to gain a feel for the interview process. It meant the interviewer could critique and refine their interviewing technique and discover which questions were most important and which questions needed refinement. From this pilot minor changes were made.

These minor changes were refining interviewing technique and minor changes to the the content of some vignettes and questions e.g. instead of using an actual commercially available drug name when stating a patient had been prescribed an anti-depressant, the vignette just stated that the patient was prescribed an anti-depressant. This change was made as it was felt using a specific drug name would distract the GP into thinking about what they would specifically prescribe rather than thinking about the information around prescribing an anti-depressant.

The interview format consisted of the interviewer introducing a vignette that described a clinical situation to the GP. The interviewer would then ask questions on health information usage related to the scenario described in the vignette. The interviewer used prompts and probes during responses to ensure:

1. the GP remained focused on answering the question.
2. clarification of any points made by the GP such that it was understood what the GP meant.
3. that there was coverage of the question asked i.e. the question was answered and to encourage the GP to say more about a certain topic if the interviewer felt it was relevant.

At the end of the questions phase of the interview, the GPs were asked to perform a rating exercise. Here the GP was given a table of criteria, derived from the literature review, and asked to rate them. These candidate criteria were data (metadata) that may accompany any health information e.g. criterion that indicated contradictory entries in the information. Using the criterion just given as an example, the rating exercise was performed by evaluating the criterion using the question "How important is it in knowing that the information contains contradictory entries when deciding how confident the GP is that the information is of a quality that it can be safely relied upon when making a clinical judgement?" Each criteria was evaluated by the GP and given a rating on a scale of 1 to 4 where 4 is very important in knowing, 3 is somewhat important, 2 is slightly important and 1 is not important.

3.5.6 Interview: Data Analysis

Interviews were transcribed from the audio recordings. Thematic analysis of each interview was conducted using NVivo software. Thematic analysis is a qualitative data analysis technique where the researcher identifies patterns within the collected data and codes them as themes (coding categories). Boyatziz 1998 defines a theme as:

> "a pattern found in information that at a minimum describes and organises the possible observations and at a maximum interprets aspects of the phenomenon" (Boyatzis, 1998)

Thematic analysis moves through three stages of inquiry: recognising a pattern within data (seeing), encoding (categorising, or describing) the pattern (seeing as something),

which then enables the third stage, interpretation (Boyatzis, 1998). Identification of patterns, or themes, is conducted in one of two primary ways; deductive coding or inductive coding (Marks and Yardley, 2004; Braun and Clarke, 2006). Deductive coding (top down approach) where coding is based on a priori theoretical ideas the researcher brings to the data, while inductive coding (bottom up), is a data driven approach where themes emerge directly from the raw data itself (Boyatzis, 1998; Braun and Clarke, 2006; Fereday and Muir-Cochrane, 2008).

While thematic analysis is usually seen as a process to be used as part of other qualitative methodologies (Boyatzis, 1998), Braun and Clarke, (2006), make a strong argument that thematic analysis should be considered a method in its own right for its ability to provide a rich, detailed, and complex account of data (Braun and Clarke, 2006). Finally, thematic analysis is not like formal data mining techniques where systems use rule based, or machine learning algorithms, to automatically search data collections for patterns. With thematic analysis the researcher plays an active role in identifying patterns/themes and selects those that are of interest (Taylor and Ussher, 2001).

Candidate Criteria Rating Exercise

For the rating exercise, some statistical analysis was employed. This was to determine the following:

- average ratings for the criteria rated.
- total score for each criterion.
- counts for each criterion rated e.g. how many scores of 4 were allocated to a particular criterion.

A listing of the 21 candidate quality criteria, used in the rating exercise, are shown in Figure 3.2. This list also represents the order they were presented to the GPs for the rating exercise conducted at the end of each interview.

The rating scale used was from 1 to 4 where 4 was very important to know, 3 was somewhat important to know, 2 was slightly important to know and 1 was not important to know. The results of this exercise are shown in Figure 3.3.

The candidate quality criteria have been ranked using the average score parameter. Using the average values, a steep drop off is noted beyond criterion 15 (C15). The drop off at this point is also indicated in the count of very important and somewhat important scores given to those criteria beyond C15. The rating exercise allowed a determination of criteria that were valued over others and for a ranking of the criteria.

Candidate Quality Criteria
Knowing the qualifications of the clinician involved in healthcare event
Knowing the qualifications of the recorder of the information
The years of experience of the clinician
The years of experience of the recorder of the information
The position held by the clinician
The position held by the recorder of the information
Attestation of entries recorded
Qualifications, experience and position of person attesting to information
Where the information was captured
Who has been responsible for storing the information
Knowing information was produced at an organisation known for its expertise in that field
Knowing the information was produced at an organisation that has a good reputation
Age of the data
Indication of contradictory entries
Knowing length of time from information captured to information recorded
Knowing testing methodology for test results
Positive/negative personal experience with provider
Contextual narrative for event for entry
Reason for change of entry e.g. change of diagnosis or change in prescribed medication
Indication of where to go for further information
Internal consistency/cohesion of record

Figure 3.2: List of 21 candidate quality criteria

	Quality Criteria	Interviewees						Average	Total	Count			
		Dr.A	Dr.B	Dr.C	Dr.D	Dr.E	Dr.F			1	2	3	4
C1	Indication of contradictory entries	4	4	4	4	4	4	**4**	24				**6**
C2	Knowing information was produced at an organisation known for its expertise in that field	3	4	4	4	4	2	**3.5**	21		**1**	**1**	**4**
C3	Knowing the information was produced at an organisation that has a good reputation	2	4	4	4	4	3	**3.5**	21		**1**	**1**	**4**
C4	Attestation of entries recorded	2	4	3	4	4	4	**3.5**	21		**1**	**1**	**4**
C5	Internal consistency/cohesion of record	4	3	3	3	4	4	**3.5**	21			**3**	**3**
C6	Positive/negative personal experience with provider	4	4	3	4	2	3	**3.33**	20		**1**	**2**	**3**
C7	Qualifications, experience and position of person attesting to information	2	4	4	4	2	4	**3.33**	20		**2**		**4**
C8	Contextual narrative for event for entry	4	2	3	4	3	4	**3.33**	20		**1**	**2**	**3**
C9	Indication of where to go for further information	4	1	3	4	4	4	**3.33**	20	**1**		**1**	**4**
C10	Knowing the qualifications of the clinician involved in healthcare event	4	1	4	3	4	3	**3.17**	19	**1**		**2**	**3**
C11	The years of experience of the clinician	3	3	3	4	3	3	**3.17**	19			**5**	**1**
C12	Age of the data	4	2	2	4	3	4	**3.17**	19		**2**	**1**	**3**
C13	Reason for change of entry e.g. change of diagnosis or change in prescribed medication	4	2	2	4	3	4	**3.17**	19		**2**	**1**	**3**
C14	Where the information was captured	2	4	2	4	4	2	**3**	18		**3**		**3**
C15	The position held by the clinician	1	1	4	4	4	3	**2.83**	17	**2**		**1**	**3**
C16	Knowing testing methodology for test results	4	3	1	2	4	1	**2.5**	15	**2**	**1**	**1**	**2**
C17	Knowing length of time from information captured to information recorded	2	4	1	2	4	2	**2.5**	15	**1**	**3**		**2**
C18	Knowing the qualifications of the recorder of the information	1	1	1	4	2	4	**2.17**	13	**3**	**1**		**2**
C19	Who has been responsible for storing the information	1	1	2	3	2	1	**1.67**	10	**3**	**2**	**1**	
C20	The years of experience of the recorder of the information	1	3	1	1	1	3	**1.67**	10	**4**		**2**	
C21	The position held by the recorder of the information	1	1	1	2	1	3	**1.5**	9	**4**	**1**	**1**	

Figure 3.3: List of 21 rated candidate quality criteria

Deductive Thematic Coding Of Transcripts

Before the thematic analysis began a thorough read through of all interview transcripts was conducted to check for transcription errors and ensure readability. The analysis used both the deductive and inductive approaches to thematic analysis. The deductive (top down) approach was applied first. This utilised the candidate quality criteria (QC) (see Figures 3.2 for listing and Section 3.3 for source) as an a priori template of codes (themes).

The deductive technique was used first as it provided an introduction to the process and a starting point. Interview transcript responses were coded at the candidate criteria i.e. when the GP's response referred to a theme of one of the candidate QC, the response was coded at that theme. For example, Dr. E's response to a question (Vignette 2, Question 1) regarding what they would want to know about a diagnostic entry from a record was:

> "You'd want to know if they saw a specialist. Or whether they saw the house surgeon. Because that would make a difference to the accuracy of the diagnosis. You would want to know if they'd had any subsequent problems with their symptoms that they were obviously displaying at the time."

This was coded at the candidate QC "Contextual narrative for event for entry", or Contextual narrative for short. Contextual narrative means there was an accompanying explanation with the entry explaining what led to the entry being in the record. Table 3.2 represents the number of times the particular criteria were discussed, ordered by the total number of times referenced in the interviews. Table 3.3 shows the candidate criteria ranked by the results of the rating exercise carried out at the end of each interview.

This coding by criteria shows no responses from candidate QC 16 onwards. This correlates with the drop off in scores for the rating exercise at the same criterion.

Inductive Thematic Coding Of Transcripts

The inductive (bottom up) technique was then employed. Table 3.4 represents the results of the inductive coding, displayed similarly to the deductive analysis. While the term emergent themes is used to describe these themes, the researcher is active in the process of discovering these themes (Braun and Clarke, 2006). These themes have been ranked by total number of references across all GPs.

Table 3.2: Coded Responses for candidate QC nodes by GPs ranked by total number of references

Candidate Quality Criteria	Interviewees						
	Dr.A	Dr.B	Dr.C	Dr.D	Dr.E	Dr.F	Total References
1 : Contextual narrative for event for entry	7	3	1	13	9	9	42
2 : Positive or negative personal experience with provider	1	15	1	5	4	9	35
3 : Knowing the qualifications of the clinician involved in healthcare event	2	2	3	8	1	1	17
4 : Indication of where to go for further information	1	2	0	4	4	5	16
5 : Knowing information was produced at an organisation that has a good reputation	2	5	0	1	0	6	14
6 : Years of experience of the clinician	0	4	0	6	0	1	11
7 : Indication of contradictory entries	1	2	1	1	2	3	10
8 : Knowing information was produced at an organisation known for its expertise in that field	1	3	1	3	0	2	10
9 : Internal consistancy or cohesion of record	2	3	1	2	1	1	10
10 : Reason for change of entry	1	0	0	4	0	3	8
11 : Age of the data	0	0	0	6	0	1	7
12 : Position held by the clinician	1	1	0	3	0	1	6
13 : Qualifications, experience and position of person attesting to information	1	0	0	2	0	0	3
14 : Attestation of entries recorded	0	0	0	1	0	0	1
14 : Where the information was captured	0	0	0	1	0	0	1
16 : Knowing testing methodology for test results	0	0	0	0	0	0	0
17 : Knowing the length of time from information captured to information recorded	0	0	0	0	0	0	0
18 : Knowing the qualifications of the recorder of the information	0	0	0	0	0	0	0
19 : Who has been responsible for storing the information	0	0	0	0	0	0	0
20 : Years of experience of the recorder of the information	0	0	0	0	0	0	0
21 : Position held by the recorder of the information	0	0	0	0	0	0	0

Table 3.3: Coded Responses for candidate QC nodes by GPs ranked by rating exercise

Candidate Quality Criteria	Interviewees					
	Dr.A	Dr.B	Dr.C	Dr.D	Dr.E	Dr.F
1 : Indication of contradictory entries	1	2	1	1	2	3
2 : Knowing information was produced at an organisation known for its expertise in that field	1	3	1	3	0	2
3 : Knowing information was produced at an organisation that has a good reputation	2	5	0	1	0	6
4 : Attestation of entries recorded	0	0	0	1	0	0
5 : Internal consistancy or cohesion of record	2	3	1	2	1	1
6 : Positive or negative personal experience with provider	1	15	1	5	4	9
7 : Qualifications, experience and position of person attesting to information	1	0	0	2	0	0
8 : Contextual narrative for event for entry	7	3	1	13	9	9
9 : Indication of where to go for further information	1	2	0	4	4	5
10 : Knowing the qualifications of the clinician involved in healthcare event	2	2	3	8	1	1
11 : Years of experience of the clinician	0	4	0	6	0	1
12 : Age of the data	0	0	0	6	0	1
13 : Reason for change of entry	1	0	0	4	0	3
14 : Where the information was captured	0	0	0	1	0	0
15 : Position held by the clinician	1	1	0	3	0	1
16 : Knowing testing methodology for test results	0	0	0	0	0	0
17 : Knowing the length of time from information captured to information recorded	0	0	0	0	0	0
18 : Knowing the qualifications of the recorder of the information	0	0	0	0	0	0
19 : Who has been responsible for storing the information	0	0	0	0	0	0
20 : Years of experience of the recorder of the information	0	0	0	0	0	0
21 : Position held by the recorder of the information	0	0	0	0	0	0

Table 3.4: Coded Responses for Emergent Themes

Emergent Themes	Interviewees						
	Dr.A	Dr.B	Dr.C	Dr.D	Dr.E	Dr.F	Total References
1 : Question EHR information	14	19	5	10	8	28	84
2 : Consulting EHR	12	11	6	27	7	11	74
3 : Finding information	3	9	4	36	6	12	70
4 : Accompanying information	11	5	3	26	4	11	60
5 : Consulting patient	10	11	6	17	7	4	55
6 : How information is generated	2	1	1	11	8	31	54
7 : Community relationships	0	29	5	2	5	7	48
8 : Past personal experience	1	26	3	3	4	8	45
9 : Perceived quality of information	0	12	6	5	3	16	42
10 : Approach to inconsistant information	5	11	7	8	6	4	41
11 : Approach to lack of information	1	8	1	24	1	6	41
12 : Their competence	0	8	7	6	2	5	28
13 : Approach to contradictory entries	4	2	5	9	2	4	26
14 : Their qualifications	4	3	5	9	2	2	25
15 : Organisational entity	2	9	3	1	3	6	24
16 : Weight of evidence	1	2	6	12	1	1	23
17 : Who generated the information.	0	1	1	11	4	6	23
18 : Specific processes	0	0	0	0	0	17	17
19 : Internal coherance	2	6	1	6	1	1	17
20 : Their experience	0	4	3	5	0	0	12
21 : Identifiable where it has come from	0	4	0	2	1	3	10
22 : Misc	0	1	0	5	0	0	6
23 : Sifting	0	0	0	3	0	0	3

The previous, deductive coding, used a template of the candidate Quality Criteria (QC). Purposefully this template did not have relationships between the candidate QC predetermined as it was important to investigate each independently first, against the interview data. The inductive approach allowed for greater flexibility. This meant that hierarchies between the emergent themes could be developed. The developed hierarchies are shown in Figure 3.4. Three themes have been removed as they only had references from one or two GPs. These were Sifting (23), Misc (22) and Specific processes (18)

The organisation and relationships of these hierarchies can best be illustrated with an example. The Finding Information theme is ranked third by references count. Figure 3.4 shows it branching down to the Consulting EHR theme, that is ranked second. This in turn branches to the Question EHR Information theme ranked first. Finding information was very much a recurrent theme in the interviews. The hierarchy indicates the relationship that to find information the GPs would consult the EHR. This would then lead to questioning the EHR information should an entry in the EHR raise questions for them. For example, if the GP had some past personal experience with the entry provider, this could make them question the entry. Dr. B's response to a question on what it is about recorded information that would provide confidence that it was safe to rely on (Vignette 1, Question 7) illustrates this. The response was:

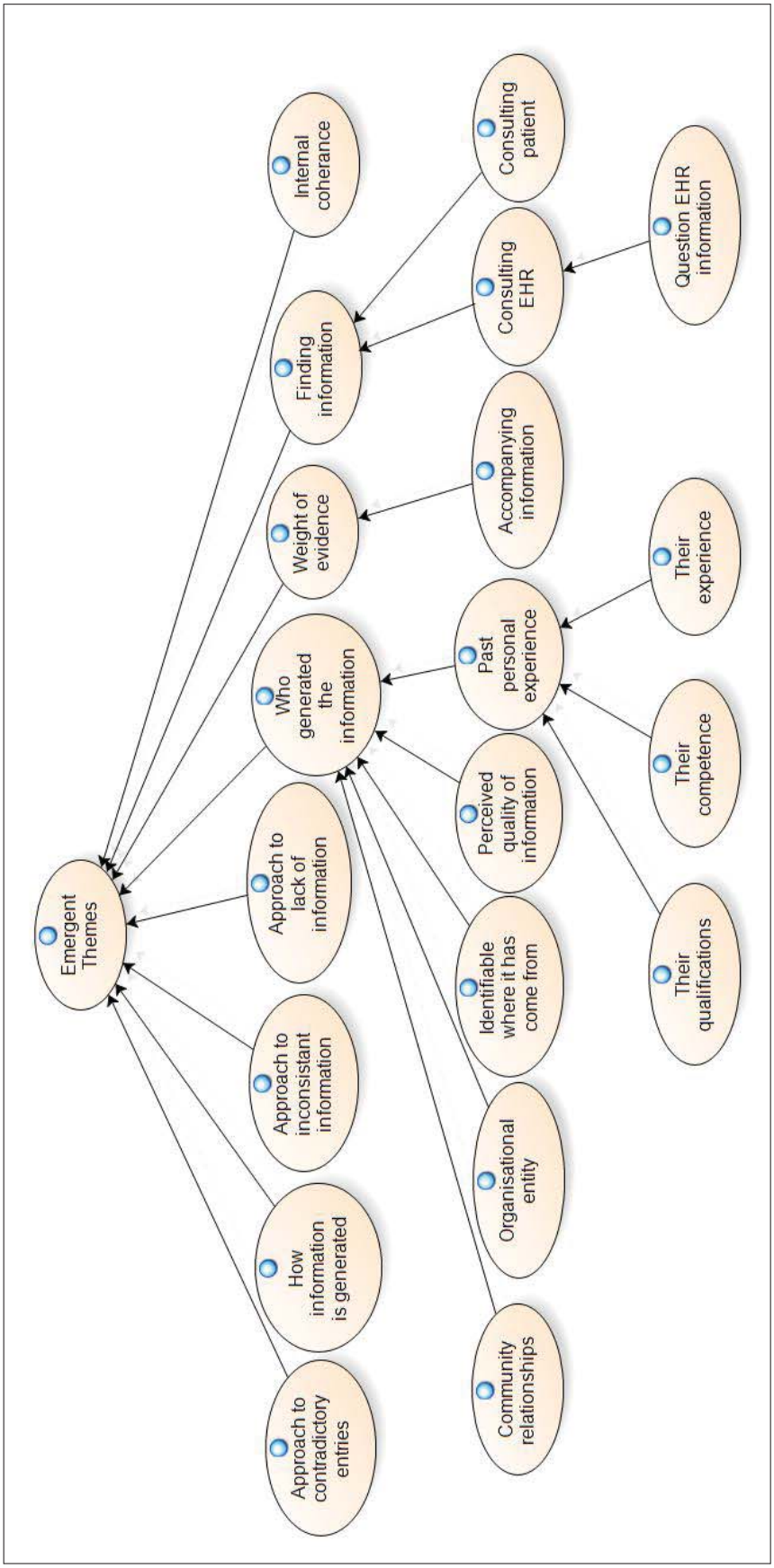


Figure 3.4: Hierarchy of Emergent Themes

> "Who recorded it, and what happened. And I think the third point would be consistency"

On probing further regarding the person contributing the entry, the response was:

> "The quality of the information, yeah. I mean if you produce generally rubbish notes, then do I believe the answers you put there? Ah, no. But on the other hand, if somebody obviously puts a great deal of effort into the patients, and their care, then I'm going to take your word with a huge amount more credibility"

The relationship is then; needing to find enough information, looking at some information, in this instance the EHR, and then if something is known regarding the source, questioning that information. This extends to organisations as well, should they be the source. For example Dr. F's response to a question regarding how they would use an entry stating an allergy to a drug (Vignette 1, Question 1), their response was:

> "Depends on the scenario in which I'm in. In one site I would have some considerable confidence that in fact the information presented to me was true and valid. In another site I would have to go off into a routine of actually justifying that information because I don't have the confidence that it's true."

The inductive coding process revealed themes that recurred across all the GPs. This can be seen by the coded responses in Table 3.4. Those themes with higher coding responses being ranked above those with lower responses. From these themes, a large proportion of responses are regarding how information is found (finding enough information), activities around questioning/verifying information, and what is known regarding the source of the information. There is a definite desire to know more, when the information is not apparent or the information, to use a software engineering term, has a "bad smell". By this it is meant that something alerts the general practitioner (GP) that they don't feel comfortable relying on that information, and need to verify, or know more about it. This is to confirm it, or discard it.

In fact the GPs interviewed show total distrust of the quality of information under certain circumstances. For example, Dr. F's response to a question (Vignette 2, Question 3) about how they would treat dosage information for an anti-depressant, that they may consider high for an initial prescription, was:

> "If it's my own, I tend to trust it quite a lot. If I'm sitting there, and I won't say the GPs name, but Blue comes up next to the patient at *** Street, I inherently don't trust anything that's in the medicines dispensed and stuff. Because I know it's probably ... there's a high chance that it's wrong, sorry to say."

This indicates that having a past personal experience with the source, in this case a negative experience, has a large effect on the GPs evaluation of the quality of the information. This also extends to organisations. This is seen in Dr. B's response to a question (General Questions, Question 8) on whether knowing the organisation that produced the information would affect their use of the information. Dr. B's response was:

> "Yes. Very much so. Because it's the individuals ... Which is a cluster of individuals ... is absolutely critical."

On probing further, and discussing if the information had come from an organisation with a poor reputation, the response was:

> "You would assume it was bullshit. And if they got it right, you'd still assume it was bullshit and they'd just got lucky."

These examples clearly show that knowing something about the source of the information is weighted heavily by the GP when evaluating its quality. The QC C6: Past personal experience falls into this category. However, it is past personal experience with the health information produced from the source, rather than personal knowledge of the individual as the source. This is illustrated by a response during the interview with Dr. F. Dr. F stated:

> "I can actually trust them even though I've never ever met them because I'm meeting them through their expression on the record. They might be a complete and utter *** for all I know, but I might still actually trust that information, even though I don't know them."

Conversely, even when something personal is known for the source, it is still the experience of the information received from that source that drives assessment of the information. Dr. F's further statement supports this. Dr. F stated:

> "I mean there was a GP here was completely unhinged. He was a bit strange. But then I got some notes from him which I thought were some of the exceptionally best notes I have ever seen in primary healthcare. And I trusted them, even though I thought he was a looney."

Not having enough information, or not being able to easily find information is also highly ranked within the emergent themes. The Finding Information theme is ranked third. An example of issues caused by this is shown in Dr. D's response to a question (General Questions, Question 7) regarding instances where the GP had wanted to rely on information but felt they couldn't because there just wasn't enough there, and what would have "put them over the line" into relying on it. Dr. D's response was:

> "Yeah, well it was the ability to dig deeper. To give me more information on which to base my decision. Otherwise I was faced with, did she require ongoing management, on very nebulus, fluffy information. I mean I really did not have enough information."

Another example is in Dr. A's response to a question (General Questions, Question 2) on how happy they were in trusting that the information they wished to rely on was fit for the purpose they wanted to use it for. Dr. A's response referred to the diagnosis of a patient with schizophrenia used in Vignette 2 and was:

> "So for example that psychiatrist question you asked me, if I've got just a thing - diagnosed with schizophrenia 2007, that doesn't make as much difference, nowadays I'd put that in one area but I'd have the letter and the reasons why that person came to that conclusion in somewhere else. Because he would justify it when he wrote and I would go and read that again. So the whole sort of stuff around it, the more information there is about the decision, will help me more to make a judgement about that information."

The follow up question to this was "Is that always an easy thing to do, to find?" The response was:

> "No, not at all."

In this exchange it is not only finding enough relevant information, but having more information about the decision to diagnose schizophrenia that would aid in the judgement of the information Dr. A wished to use.

The examples given here are indicative of the type of references used to justify the creation of emergent themes discovered during the inductive coding phase of the analysis. As can be seen in Table 3.4 there are many references made by the GPs to justify each theme. As stated previously the themes fall into groupings around finding information, or finding enough information, and what is known about where it has come from i.e. the source. The groupings can be generalised as themes grouped on the information itself and themes grouped on the source of the information.

One other grouping of emergent themes, not yet mentioned, was also discovered. The grouping is regarding processes used by the GP when dealing with information that they question. The grouping includes themes dealing with the questioning of information. The themes are:

- Approach to contradictory information
- Approach to inconsistent information
- Internal coherence (of information)
- Perceived quality of information
- Weight of evidence

This grouping is important as it shows that the GPs are not willing to accept certain pieces of information, or lack thereof, and feel they must investigate further to evaluate if information is fit for purpose i.e. safe to base a clinical judgement on.

Across Themes Matrix Query

The result of the coding process were two sets of themes with one obtained deductively, and the other inductively. The themes, and the developed thematic maps, were checked against each other and against the transcript data to ensure internal coherence, consistency, and distinctiveness (Braun and Clarke, 2006). In effect, to ensure strength of evidence for supporting themes. From the candidate quality criteria, themes from criterion sixteen to twenty-one were not used. This was because these themes had no coded references from the transcripts and they had a marked drop off in average score from the rating exercise. From the emergent theme set, themes Sifting (23), Misc (22) and Specific processes (18) were excluded as what few coding references there were, were only from one to two general practitioners.

The two sets of themes were used in one last form of analysis. This used the NVivo Matrix Query facility. A matrix query checks all references from transcripts for one set of

themes against all references for a second set of themes. For this analysis the candidate criteria themes from the deductive analysis were cross referenced with the emergent themes from the inductive analysis. References from the transcripts were correlated between themes in this way. This showed where references from GPs for the emergent themes supported the candidate criteria template of themes, further corroborating the criteria's importance. Table 3.5 and Table 3.6 display the results of this query.

For these results the quality criteria are ranked on the vertical left of the table, using the total number of references for ranking purposes. While they have been ranked by references, as in Table 3.2, the numbers have been retained from their ranking by the rating exercise. The horizontal top of the table shows the emergent criteria, also ranked by total references from left to right. Total references was used, rather than the ranking for candidate criteria from the rating exercise, so as to be comparing "apples with apples" within the query.

The results for the higher ranked candidate criteria and the higher ranked emergent themes show generally a good correlation. Looking at specific pairs of cross referenced themes also supports the overarching themes identified. Themes of, finding enough information, activities around finding/verifying information, and how what is known about the source of the information, will affect their judgement of the information fitness for purpose.

For example Dr. D's response to a question (part of response to Vignette 3, Question 5) regarding whether to trust a medication change on a record for a patient and whether knowing the provider , or person, that had changed the medication was at the intersection of the two themes "Community relationships" and "Positive or negative personal experience with provider" (seventeen references across five GPs at this intersection). The response was:

> "Not necessarily, as long as I believed they knew what they were doing. Then you get the question of "how do I know that they know what they're doing". I mean having a personal relationship with another provider basically means that I know things like ... I can make some judgement calls, correctly or not, on their competence. Because I've got some information about them. Such as their age, their level of experience, their level of exposure to that kind of stuff."

On probing further as to would they consider it better to know, or not know, the response was:

> "I don't know. I would just have to consider that decision on its merits at the

Table 3.5: Accross Themes Matrix Query Results - Part 1

Quality Criteria	Emergent Themes										
	Question EHR information	Consulting EHR	Finding information	Accompanying information	Consulting patient	How information is generated	Community relationships	Past personal experience	Perceived quality of information	Approach to inconsistant information	Approach to lack of information
8 : Contextual narrative for event for entry	2	5	9	24	3	7	0	3	3	6	3
6 : Positive or negative personal experience with provider	10	1	0	4	1	2	17	20	10	4	0
10 : Knowing the qualifications of the clinician involved in healthcare event	1	1	2	6	3	2	0	3	2	3	0
9 : Indication of where to go for further information	3	1	12	2	0	2	0	0	1	2	3
3 : Knowing information was produced at an organisation that has a good reputation	7	1	0	1	0	1	4	7	4	0	0
11 : Years of experience of the clinician	1	0	1	2	1	0	1	4	2	2	0
1 : Indication of contradictory entries	3	0	1	0	1	0	2	2	0	4	0
2 : Knowing information was produced at an organisation known for its expertise in that field	2	1	0	1	0	0	3	5	3	0	0
5 : Internal consistency or cohesion of record	5	1	1	2	2	0	2	3	1	6	0
13 : Reason for change of entry	0	0	2	6	0	1	0	0	0	1	0
12 : Age of the data	1	2	2	3	1	1	0	0	0	0	2
15 : Position held by the clinician	1	0	1	3	0	1	0	1	0	1	0
7 : Qualifications, experience and position of person attesting to information	0	1	1	1	0	0	0	0	0	1	0
4 : Attestation of entries recorded	0	0	0	1	0	0	0	0	0	0	0
14 : Where the information was captured	0	0	0	1	0	0	0	1	1	0	0

Table 3.6: Accross Themes Matrix Query Results - Part 2

Quality Criteria	Emergent Themes									
	Their competence	Approach to contradictory entries	Their qualifications	Organisational entity	Weight of evidence	Who generated the information	Specific processes	Internal coherance	Their experience	Identifiable where it has come from
8 : Contextual narrative for event for entry	3	5	7	0	5	7	2	2	3	0
6 : Positive or negative personal experience with provider	8	1	6	6	2	5	2	1	5	3
10 : Knowing the qualifications of the clinician involved in healthcare event	6	1	12	0	5	4	0	0	5	2
9 : Indication of where to go for further information	1	0	1	0	0	1	1	0	0	0
3 : Knowing information was produced at an organisation that has a good reputation	0	0	1	7	2	0	2	0	1	1
11 : Years of experience of the clinician	4	1	8	0	3	2	0	0	4	1
1 : Indication of contradictory entries	1	4	0	0	0	0	0	2	0	0
2 : Knowing information was produced at an organisation known for its expertise in that field	2	0	3	4	3	0	0	0	1	2
5 : Internal consistency or cohesion of record	2	2	0	0	1	1	0	4	0	0
13 : Reason for change of entry	0	4	1	0	0	3	1	0	0	0
12 : Age of the data	0	0	0	0	2	1	1	0	0	0
15 : Position held by the clinician	1	0	3	0	0	2	0	0	1	1
7 : Qualifications, experience and position of person attesting to information	0	1	1	0	1	2	0	0	0	0
4 : Attestation of entries recorded	0	0	0	0	0	0	0	0	0	0
14 : Where the information was captured	0	0	1	0	1	0	0	0	1	0

> time. I mean certainly, if you know somebody, you are able to trust their ... you have an element of trust built around a whole lot of information around that person, some of which is objective, some of which is subjective, and it's based on past experience. In other words if I've had dealings with that person, and they've made lots of correct decisions, and helpful decisions, I'm more likely to assume that this one will equally be the same."

This shows that the two themes key-in the general practitioner, to some detail of the information source, that affects their judgement of that information. On questioning/verifying information, this is illustrated by Dr. F's response to a question (part of response to General Questions, Question 8) that was referenced at the intersection of the two themes "Question EHR information" and "Internal consistency or cohesion of record". The response was:

> "If I see some nice and neat, not necessarily concise, but clear, logically sequenced notes with a good build up to why the diagnosis was made and stuff. I can actually trust them even though I've never ever met them because I'm meeting them through their expression on the record. They might be a complete and utter *** for all I know, but I might still actually trust that information, even though I don't know them."

An example that covers both the finding enough information and the knowledge, or lack thereof, about the source is at the intersection of the "Finding information" and the "Indication of where to go for further information" themes. Here, Dr. E's response to a question (part of response to Vignette 2, Question 6) regarding any information that could accompany a medicines dosage entry that would provide trust in the information is illustrative. The response was:

> "And I mean you could write to the previous doctor and say "why have you done this?" and they might have a perfectly good reason for it. You've got to find out more information."

A further example from Dr. E also highlights the questioning of information, and the need to find out more. This was also from the same intersection of themes. Dr. E had indicated that if there was a diagnosis for the patient of schizophrenia, a serious diagnosis, with no real detail e.g. who the patient had seen, they would seriously dispute the diagnosis, ignore the entry in the record, and start all over again with the patient. The response was:

> "Have I got the time, or has my nurse got the time, to find out actually what had actually gone on. I would probably check it up a bit further. I would probably still put them on antidepressants but I'd probably try to get some definite record of what they ... you know like a piece of paper in my hand. I'm not very trusting."

Some intersections from the query show correlations that were expected. For example, for the two themes "Past personal experience" and "Positive or negative personal experience with provider". However the examples given above encompass the overarching themes of finding information, questioning/verifying information, and what is known about the source of the information. All with the aim of judging the information fitness for purpose.

Thematic Analysis - Summary

The thematic analysis required approximately twelve to thirteen readings of each interview transcript. The analysis used deductive and inductive thematic coding techniques. The deductive coding used a template of candidate quality criteria to code against. With the inductive technique, the researcher derived themes from the data itself. The derived themes were labelled emergent themes.

The analysis revealed that the general practitioners (GPs) interviewed desired to know more about information when information was not apparent, or the information led to more questions. This issue is seen in responses that can be summed up as "who recorded it, what happened, and why is it in the record", as well as "do I know the source, what do I know about them". The GPs were often not satisfied that they had enough detail to judge the information they wished to use as fit for purpose i.e. was of good enough quality. This would lead to querying the information and further investigation.

The overarching themes can be visualised using thematic maps. The thematic map for themes grouped on the information itself is shown in Figure 3.5. This map represents those themes identified as important in finding and verifying information. The thematic map for themes grouped on information sources is shown in Figure 3.6. This map represents those themes identified as important when looking to the source of information.

The sets of themes (a priori candidate quality criteria, and emergent), and the thematic maps were checked against each other and checked against the data set (interview transcripts). This process was used to corroborate the themes and confirm the findings (Fereday and Muir-Cochrane, 2008).

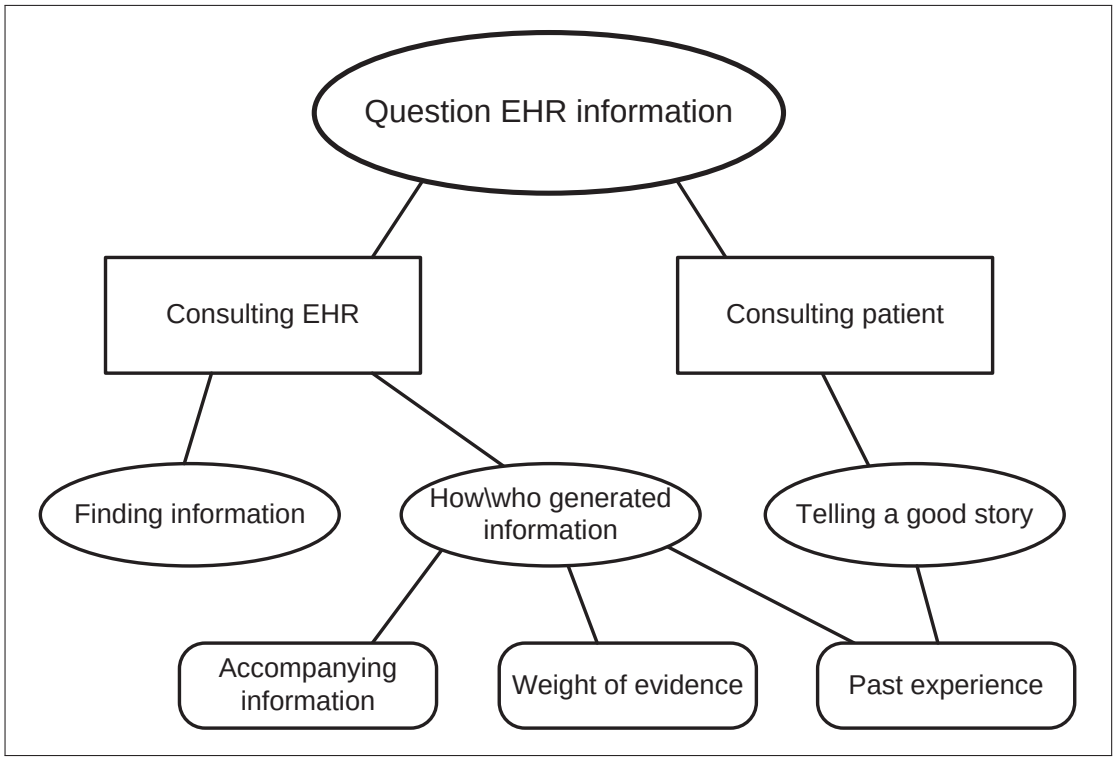


Figure 3.5: Thematic Map for themes on information

3.5.7 Interview: Data Interpretation

Most of the group interviewed had something to say on most of the themes identified, providing a rich discourse on the topic. Data analysis revealed that the group of general practitioners (GPs) interviewed need a very close relationship with the information they use. A recurring theme across all GPs interviewed when dealing with entries from Electronic Health Records (EHRs) was "Who made the entry?", "Why?", and "How did it get into the EHR?". The GPs often lacked enough information to answer these questions. Or, enough detail to judge entries when questions were raised as to the safety of them relying on a piece of information. If an entry struck the GP as being fine, made sense, and ticked all the boxes in their personal internal framework used to judge information, all well and good. If, however, flags were raised in going through the GPs personal check-list for judging the information fitness for purpose, then often they would be "stuck" as to the information safety, and need to go off into a routine of trying to verify what they were looking at. As some stated e.g. Dr D, information can not simply be ignored should it seem not right, it needs to be taken notice of. This routine was often difficult and time consuming, with lack of desired details on the information proving problematic. A case of the devil is in the details, or lack thereof.

At the beginning of this research these were not anticipated findings. It was thought

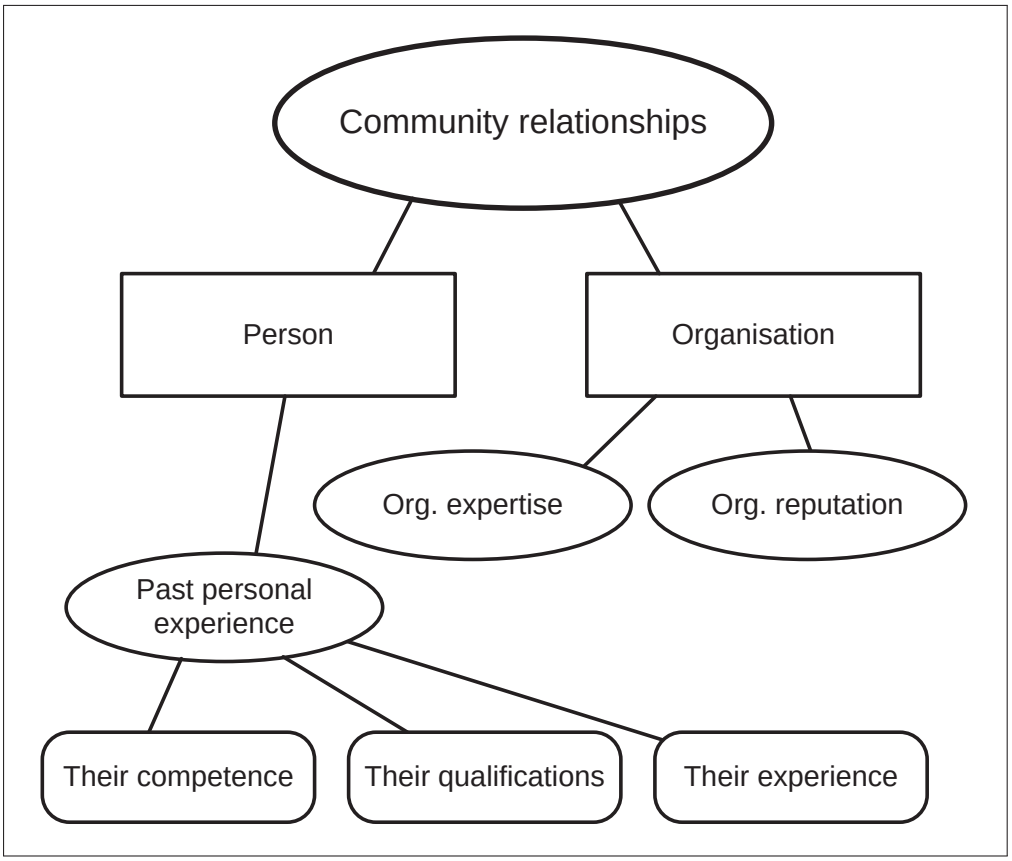


Figure 3.6: Thematic Map for themes on information source

that mapping quality criteria to data quality dimensions and using those dimensions to abstract a computational measure of quality for health information would be an acceptable approach. This would be similar to approaches taken by those investigated in the literature in Chapter 2. However, through the interviews, and later data analysis, the approach just described came into question. Questions were raised in the researchers mind regarding the efficacy of this approach. This was because of the nature of health information, how the group of GPs approached that information, and the context of its use.

Health information is highly complex, with many components. For example, a medication entry can have any of the following:

- Product name e.g. Amoxicillin
- Date/Time prescribed
- Type e.g. intravenous fluid
- Description e.g. Amoxicillin

- Status e.g. no longer active
- Brand name e.g. Amoxil
- Manufacturer
- Strength e.g. 250mg
- Form e.g. capsule
- Quantity
- Directions e.g. Insulin sliding scale
- Delivery method e.g. chew
- Frequency
- Source e.g. a healthcare provider

These are elements taken from the available entries provided by the American Society for Testing and Materials International's (ASTM) Standard Specification for Continuity of Care Record (CCR) standard. It is not a complete list of available elements for a medication entry. The value of scoring/computing a medication entry along quality dimensions such as Accuracy and/or Completeness, in effect distilling the medication entry to a single value, is debatable.

The GPs interviewed approach to such an entry is "Does that entry make sense against my internal framework, or check-list, for what I would expect of a medication prescription of that type, for the context it was prescribed?" Unless it is possible to state the entry is 100% Accurate and/or Complete, what is the value of stating it is 78% Accurate and/or Complete? Accuracy and completeness are used as examples here as much of the research into data quality in healthcare has focused on these two dimensions (Hasan and Padman, 2006). Accuracy and completeness of medical data is an important issue, as Wagner and Hogan (1996) report that accuracy in medical registries ranged from 67% - 100%, while completeness ranged from 30.7% - 100% (Wagner and Hogan, 1996).

The GPs would generally consider an entry as Accurate if it were attested to i.e. had been reviewed, and signed off. Presence of attestation could be used to say it was 100% accurate and/or complete, but if not attested to, the issue remains. Computation using quality dimensions is explored further in Chapter 4, Section 4.4.1.

Lastly, the context the information is needed for is in the treatment of patients. The GPs know the consequence of relying on information that is labelled 78% accurate could adversely affect a patient, and hence they would not rely on it. In fact the GPs

know that they can't just trust information at face value and there is a need to evaluate it. Providing the GPs with indicators, based on realised quality criteria valued by the GPs, is seen as a way of assisting the evaluation of health information.

3.5.8 Interview: Limitations

There are limitations to the approach taken for the interviews. First, is the small sample size. With six GPs it is not possible to claim that the results, analysis, and conclusions from the interviews are generalisable. Second, the recruitment process for participants could be seen as a limitation.

Addressing the first concern of sample size, this thesis is a proof of concept study. Therefore it is not attempting to demonstrate that these are the definitive set of quality criteria used by all general practitioners (GPs) to determine the quality of health information. Rather, it is demonstrating that for a given set of criteria, this group of GPs identified them as important in knowing, when evaluating the quality of health information they are presented with.

The second possible limitation is the Snowball Methodology used for recruitment of participants. It could be argued that this does not provide a representative (randomised) sample of GPs. Again, as this is a proof of concept, the thesis is not trying to prove that this is the only way of achieving the goal, but rather this is one way of achieving it.

3.5.9 Interview: Conclusions

There is a gap in providing GPs with accompanying data with health information that would assist them in their interpretation as to the quality of the information they are presented with. This gap is both in the type of information that accompanies health information, and also in the amount of information provided for some data items within the health information.

The GP must discern if information is safe to use. Those GPs interviewed use an informal, ad hoc process in this determination of the quality of the information. The GP utilised the tacit knowledge they have gained from experience, and community knowledge gained from any past experience with the provider of the health information. This amounted to an evaluation against a GP's internal framework of evaluating the information, justifying their acceptance or rejection of the information. When they felt the need for this justification they would go into a routine that used this internal framework. They were also shown to use judgement based on the contents of the record

itself e.g. judgement of the quality of the health information was lowered if they felt the record was not consistent, or didn't hold up in a cohesive fashion.

Through the interview process it was found that those GPs interviewed felt that there were criteria, that if made available with the health information, would positively assist them in their decision making process. The decision making process being one of deciding if the information available to them is of sufficient quality that it can be relied on and would be safe to use. The GPs need a very close relationship with the information. This is understandable with the context of the information use i.e. in the treatment of patients. This meant that extra details were required when they felt a need for them, to satisfy themselves that they could accept, or reject, the information.

It was found that indicating quality criteria for the health information to GPs at the point of care could be beneficial. This conclusion is based on the fact that the GPs felt that having available with any health information, those criteria, both discussed indirectly via the interview process and from the rating exercise, would aid them in being able to judge their confidence in the quality of the health information. In effect, assisting the GP in their routine of justifying the acceptance, or rejection, of the information.

3.6 Theoretical Framework

Highlighted in the Data Interpretation section, Section 3.5.7, were insights into the usefulness of providing computational scores for data items along data quality dimensions. Although this questioned the data quality dimension approach, a mapping of Quality Criteria to quality dimensions was conducted to further explore the QC and the quality dimensions.

Through the interview process, data analysis, and data interpretation, a theoretical framework was developed. As stated in Chapter 2, Section 2.7.5, frameworks have been used to capture data quality requirements in specific contexts and to allow a structured approached to data quality issues. Before moving to the theoretical framework for quality criteria, some abbreviated names are given for the quality criteria to make them easier to manage and discuss. Table 3.7 shows these alternate names.

The developed theoretical framework represents quality criteria, with their associated mappings to quality dimensions, that describes what features of health information the interviewed group of General Practitioners look for, or use, when discerning the quality of health information. The framework is a mapping to the framework of Wang and Strong (1996). The Wang and Strong (1996) framework was chosen as they utilised

Table 3.7: Quality Criteria abbreviated names

Quality Criteria (QC) - Full Name	QC - Short Name	QC - Label
1 : Indication of contradictory entries	C1 : Contradictory entries	C1
2 : Knowing information was produced at an organisation known for its expertise in that field	C2 : Organisational expertise	C2
3 : Knowing information was produced at an organisation that has a good reputation	C3 : Organisational reputation	C3
4 : Attestation of entries recorded	C4 : Attestation	C4
5 : Internal consistancy or cohesion of record	C5 : Consistency/cohesion	C5
6 : Positive or negative personal experience with provider	C6 : Personal experience	C6
7 : Qualifications, experience and position of person attesting to information	C7 : Attestor role	C7
8 : Contextual narrative for event for entry	C8 : Contextual narrative	C8
9 : Indication of where to go for further information	C9 : Further information	C9
10 : Knowing the qualifications of the clinician involved in healthcare event	C10 : Clinician qualifications	C10
11 : Years of experience of the clinician	C11 : Clinician experience	C11
12 : Age of the data	C12 : Data age	C12
13 : Reason for change of entry	C13 : Change narrative	C13
14 : Where the information was captured	C14 : Information captured	C14
15 : Position held by the clinician	C15 : Clinician role	C15

a similar methodology to that used to formulate the theoretical framework developed here. The methodology involved going to data users and questioning them as to what they considered important regarding data quality. The theoretical framework is shown in Figure 3.7.

The mapping of the framework includes the addition of two categories:

- Organisational
- Personal

The categories were added to take into account the findings from the interviews with the general practitioners (GPs). The GPs identified organisational origins of information, and personal traits for people originating information, as important, in discerning data quality in the context the information is used. These categories express the areas the GPs take interest in, or want to know, when going through a routine of justifying their acceptance, or rejection, of information.

Two categories from Wang and Strong's (1996) framework were not included. They are Representational Data Quality and Accessibility Data Quality. Representational Data Quality is interested in quality dimensions such as Interpretability and Representational Consistency while Accessibility Data Quality is interested in quality dimensions such as Accessibility and Access Security. Quality dimensions from these two categories were outside the scope of this thesis.

The theoretical framework allows for an organisation of the identified quality criteria. This applies structure and understanding of how the criteria relate to the domain of investigation and what quality dimensions the criteria can be categorised as. The following section further categorises Quality Criteria (QC) to understand what they represent, and to move toward using the QC.

3.6.1 Categorising Quality Criteria

Categorising the QC aids understanding of what the QC represent. Further analysis of the interviews and framework, and based on initial categorisation within the theoretical framework, it was deduced that what QC indicate, at a conceptual level, are relationships that the GP think important to know. Three types of relationship were found. These relationships can be described as:

- between entries e.g. Criterion 1, contradictions between entries.

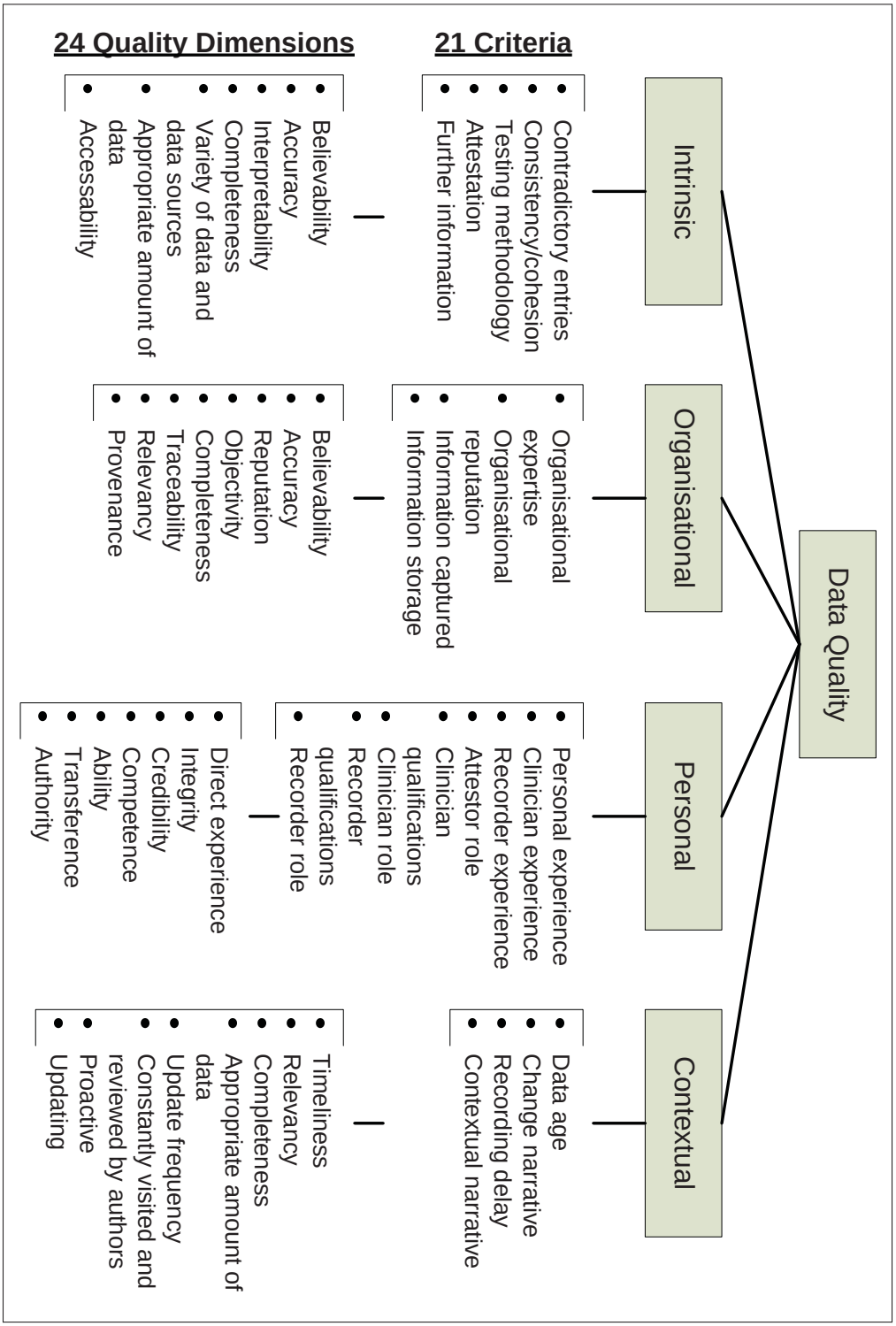


Figure 3.7: Theoretical Framework

- between actor and entries e.g. Criterion 4, entries are attested to.
- between actor and actor e.g. Criterion 6, positive/negative personal experience with entry provider.

For the purposes of these relationships an actor could be either an individual, or an organisation. An example of a relationship between actor and entry can be seen in some responses to interview questions from GPs given below. The question asked was (Vignette 1, Question 7):

> "What is it about recorded information on a patient's drug allergy status that would provide you with confidence that the information can be safely relied upon and used in clinical decisions?"

Dr. B responded:

> "Who recorded it, and what happened."

and Dr. D responded:

> "What makes me trust that information? The detail, who did it, who wrote it, whether I believe that other person knew what they were talking about."

This illustrates that QC highlight features of health information that describe relationships that the GP wants to know when deciding whether the information is fit for purpose. Put another way, QC add information to EHR regarding identified features of these relationships, that a General Practitioner (GP) think important to know. Figure 3.8 shows QC categorised by the relationship they feature.

3.6.2 Evaluation Based Categorisation Of Quality Criteria

A further way of categorising QC is by the type of process used to evaluate their outcome i.e. does the QC require an objective evaluation or a subjective evaluation. For example, C:1, Contradictory Entries is an objective evaluation i.e. are there entries that contradict other entries, while C:3, Organisational Reputation is a subjective evaluation i.e. the reputation of an organisation is subjective and may differ between GPs. Figure 3.9 shows QC categorised by whether they are based on an objective or subjective evaluation.

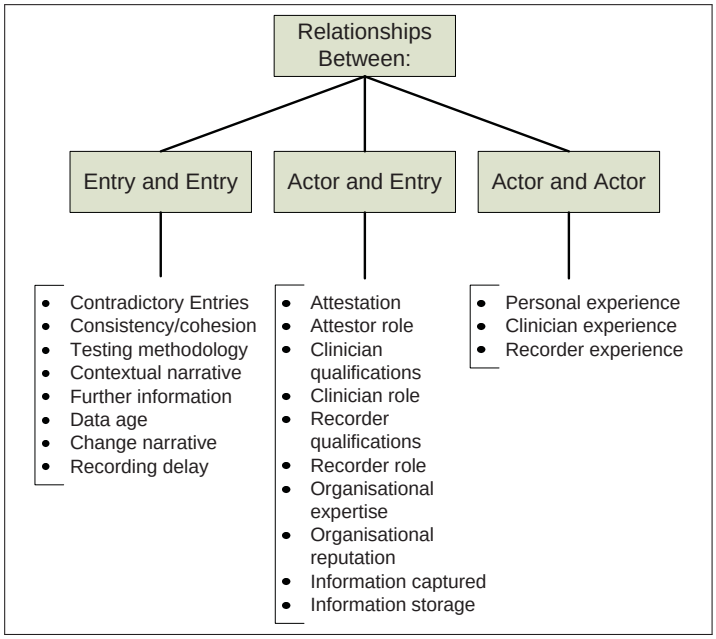


Figure 3.8: Quality Criteria Categorised by relationship type.

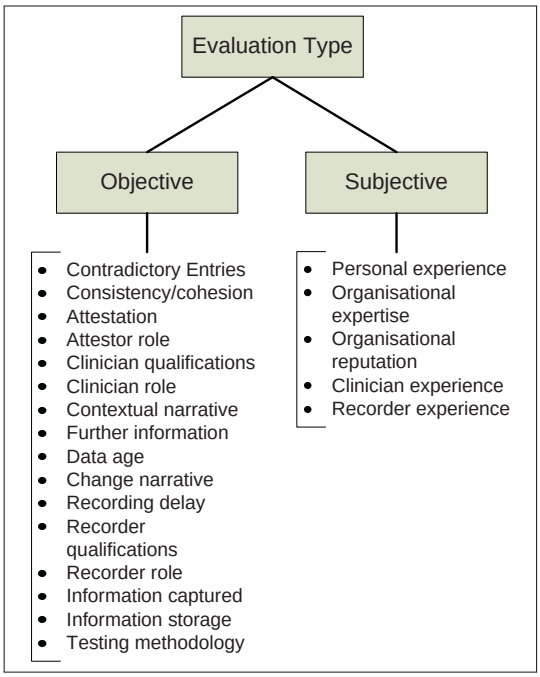


Figure 3.9: Quality Criteria categorised as Objective or Subjective

Categorising QC by the relationship within health information they belong to is at a conceptual level. Categorisation into objective and subjective evaluation types moves closer to usage of QC and gives direction as to how they could be further developed and implemented.

3.7 Chapter Summary

Interviews provided good support for study of QC for potential use in health information. Strong themes were common across interviews. Analysis showed support for concepts represented by QC. Further work showed flexibility of QC, with different categorisations. The flexibility would give guidance to formalisation of QC with development of the Quality Criteria Model, in the following chapter. What is clear is that the GPs wanted a very direct relationship with the information they are presented with, in order to make use of the information.

Chapter 4

Quality Criteria Model

4.1 Chapter Overview

This chapter is organised in three sections. First, a brief review of health information models is presented. This has a focus of identifying the structures used in these models. Second, QC are further extended, and a model for QC is presented, the Quality Criteria Model (QCM). The QCM is used to represent QC in a structured, formal way, and describes QC semantics and representative structure. The criteria are not modelled in isolation, but rather with the context of health information use and health information quality assessment in mind i.e. the context for which they were identified. Last, application of the formalised model is described using an example. Two models are developed, a static model for capturing QC, the QCM, and a dynamic model used for processing and producing results.

4.2 Health Information Model

The need for an Electronic Health Record (EHR) was identified and discussed in Chapter 2, with some examples introduced. In the Introduction of Chapter 3, requirements for an EHR, and the role an EHR needs to fulfil, were highlighted e.g. an EHR should meet any medico-legal requirements and provide for faithful reproduction upon communication (ISO, 2008b). In this section the information models of EHRs are investigated. This is the structure and content model used within an EHR. A general discussion incorporating generalised features of EHRs is given, as all those investigated share common properties. The focus is on the clinical information being stored rather than administrative or patient demographic information. Individual EHRs, without patient data, will be used to provide examples for points made. These examples range from specifically

implemented EHRs that form part of a patient's whole record e.g. New Zealand's proposed Shared Care Record (NZ National Health IT Board, 2010), to health information reference models that can be used to implement a wide variety of EHRs e.g. openEHR's clinical information model (openEHR Foundation, 2013b).

The structure of the EHR is organised in sections that are containers that hold a specific type of entry. Each section being for a specific type of entry. The EHR can itself be considered a container for the sections. Some example sections taken from currently used EHRs, or from standards for EHRs are:

- Medications (ASTM, 2005)
- Past Medical History (Health Level 7 International, 2004)
- Problem List (openEHR Foundation, 2007)
- Diagnostic Tests Results (ISO, 2008a)
- Alerts (ISO, 2008b)
- Allergies and Adverse Reactions (New Zealand Clinical Documents Templates CDA for Medications and Adverse Reactions; built on HL7 CDA) (HISO, 2013)
- Post Discharge Plan (national hospital Electronic Discharge Summary (EDS) for New Zealand, for transfer of care from secondary to primary healthcare practitioners; being piloted at Counties Manukau District Health Board)
- Social History (Proposed Shared Care Record for New Zealand; content model based on ASTM's CCR) (HISO, 2012b)

A Venn Diagram shows a high level, conceptual representation, of the structures in Figure 4.1. This also shows the discrete, self contained, nature of health information entries. Figure 4.2 shows a model of the same structures, that moves from the conceptual representation to how the structures are implemented via a tree data structure.

Features In Common With Object Oriented Paradigm

Before moving to examples of currently used health information models, a comparison between the design of the health information models, and a popular software development paradigm, Object-Oriented(OO) programming, is given. The comparison is limited to the concept in OO programming of types, or classes of objects and their extensions, and does not speak to the OO concept of describing the behaviour, or functions, of objects.

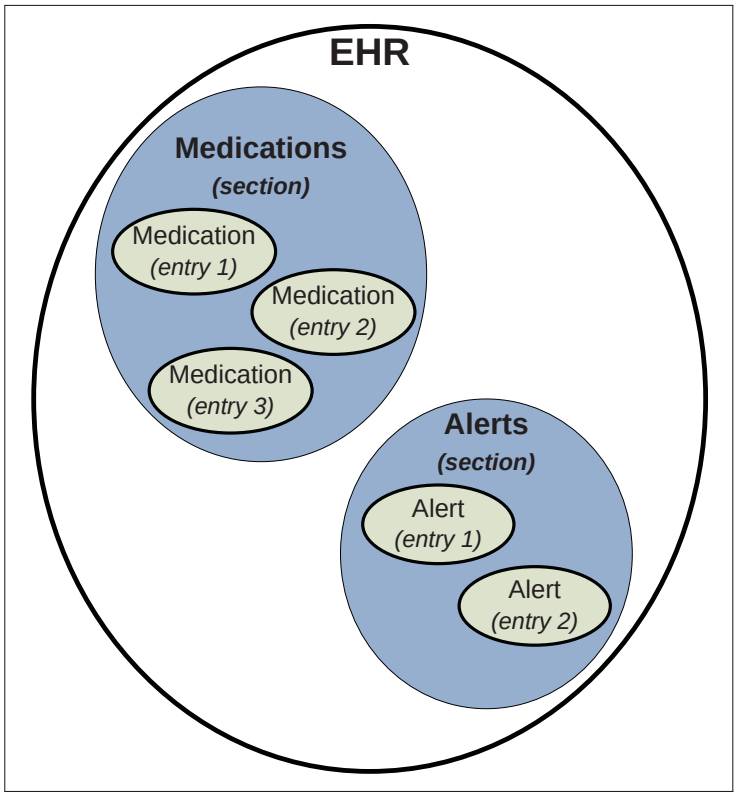


Figure 4.1: Venn Diagram showing conceptual representation of health information model as containers.

A class is a generalised description (template) that describes a collection of certain types. The class contains data fields that describe attributes for the type the class represents. An object is an instance of a specific class (Pressman, 2005; Horstmann, 2008). Classes can inherit from a more general type into more specialised type. Figure 4.3 illustrates the concept where Person is the most generalised type (class), that specialises into Student, Lecturer, and Administrator. Administrator is further specialised into a Financial Administrator, and an Academic Administrator.

Student is a specialised type of Person. Every Person has a name, and date of birth that Student would inherit, but not every Person has a student ID. Student ID is special to the Student class. For example the instance (object) of the student class Student1, given the name Stephen and a date of birth as a data attribute (from Person class), with another data attribute, Student ID, from the Student class.

The OO type structure is similar to the health information models described in Section 4.2. With the health information models, an entry is a general type. A medication, alert, or problem list entry is a specialised type of entry. Every entry has an identifier and a source e.g. an ID number and an actor that made the entry. The identifier and source are

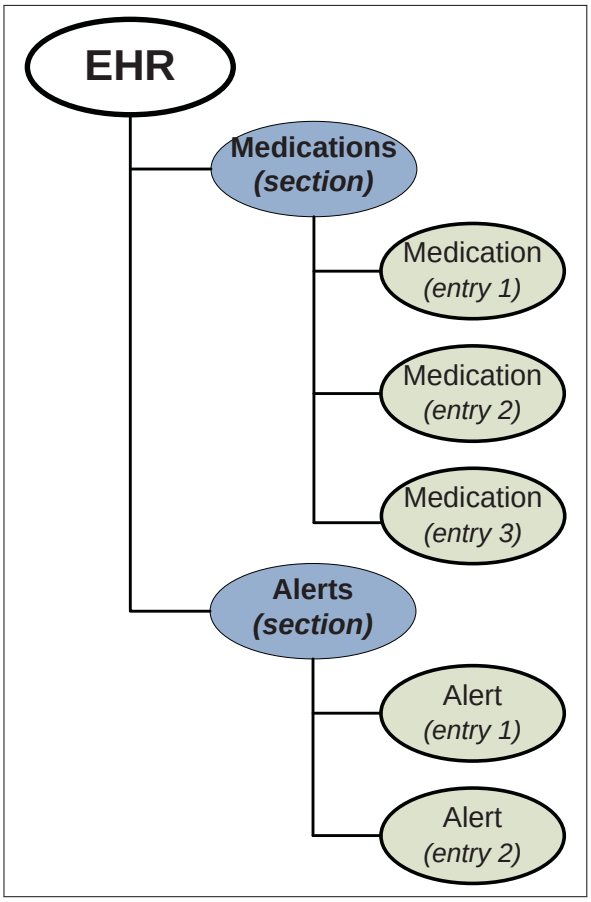


Figure 4.2: Health information model showing implementation as tree data structure.

common attributes of an entry, with a specialised type of entry e.g. medication, having further attributes such as brand name of medication and dose. Figure 4.4 shows this structure. Several of the health information models acknowledge the design similarities. For example the HL7 version 3 Reference Information Model (RIM) is described by a set of classes using an Object-Oriented modelling language (Health Level 7 International, 2013d), and ASTM International's Continuity of Care Record standard defines entries such as medications as data objects (ASTM, 2005).

The use of types and type specialisation is also reflected in the section structure of the health information models. Classes of types are grouped together into containing sections. For example, medication entries in a medications section, and alert entries in an alerts section.

The type structure concepts of Object-Oriented programming are discussed here as they are also reflected later in the development of the Quality Criteria Model (QCM). They are further related as objects are a special type of tree data structure. The QCM (shown in Figure 4.8) shows that Quality Criteria (QC) share some common characteristics.

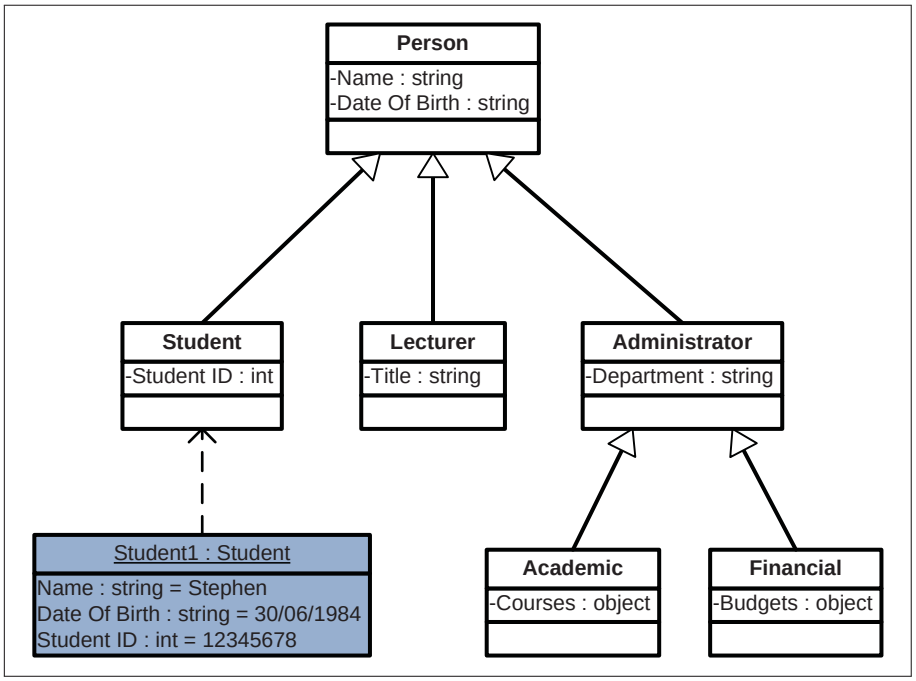


Figure 4.3: Object-Oriented type specialisation.

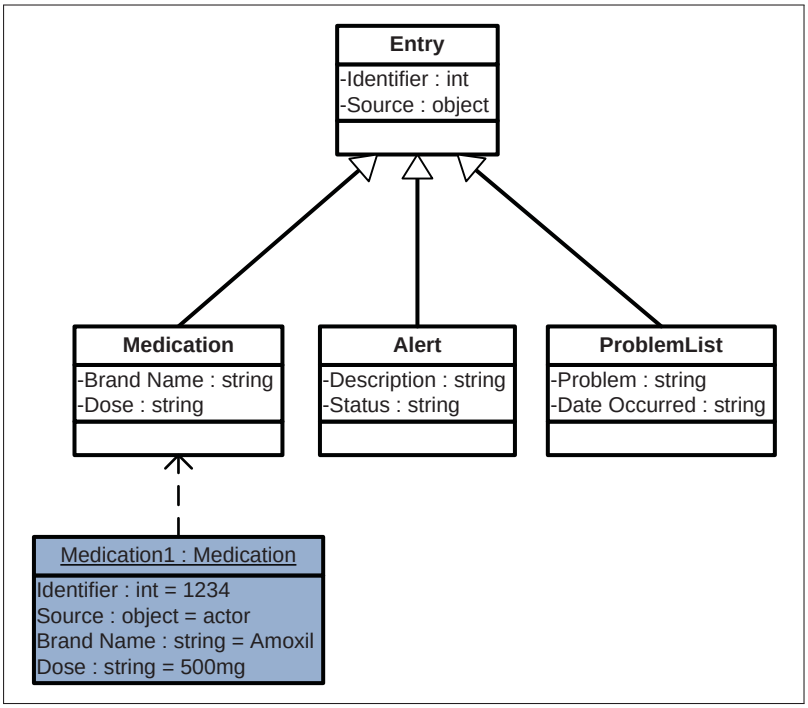


Figure 4.4: Health information entry type specialisation.

However, each type of QC has specialised attributes, as with specialised types. Further, an instance of a QC only takes from the QCM what is needed to represent the specialised type of QC being instantiated, with an instantiated QC becoming a Health Information Quality Indicator (HIQI). The following section, Section 4.2.1, introduces and discusses some specific examples of health information models, standards, and currently used records.

4.2.1 Health Information Model - Examples

The United Kingdom's National Health Service's Summary Care Record (SCR) is a centrally stored health summary. It is intended to provide information when other records are unavailable (Greenhalgh et al., 2008b). Technically it is a tree based data structure where nodes on the tree take the form of attributes (leaves) and containers (intermediate nodes) from the record, with their associated attribute values. For example it has a repeat medication section, or container for repeat medications. Here the nodes are attributes such as the date/time prescribed, medication name and dosage instructions. The particular attributes are the leaves of the tree with nothing below them.

openEHR is a specification that allows for implementation using a variety of software and tools e.g. Extensible Markup Language (XML) or any object oriented programming language. As such it takes the form of a tree data structure. All clinical entries are modelled using archetypes, which can be thought of as objects, or types, that represent re-usable models of domain concepts e.g. a clinical event.

The American Society for Testing and Materials International (ASTM) published ASTM E2369 - 05 Standard Specification for Continuity of Care Record (CCR). The CCR is used to hold the most relevant information for a patient's healthcare e.g. clinical and demographic information. Its specific purpose is to communicate this information as an aggregation of pertinent health information that will follow a patient. It differs from ISO 18308 as it specifies it must be implemented such that it strictly adheres to an XML schema and supplied implementation guide (ASTM, 2005). ASTM have produced an XML schema for implementation. As it is XML-based, it has a tree data structure with sections, or containers holding specific types of entries e.g. a results section for any test results.

Health Level 7 (HL7) is an American National Standards Institute (ANSI) approved standards development organisation. HL7 founded its reputation on messaging standards that set out data format and content to provide for effective communication between different health information systems. HL7 messages are hierarchical structures that can be implemented using tools that implement tree data structures such

as XML and/or Object-Oriented programming languages. The HL7 Clinical Document Architecture (CDA) standard can be used as a medical record and is an XML-based mark-up standard, and therefore tree based. HL7 continues to actively work on standards for clinical content for an EHR. For example, in collaboration with ASTM, they have developed the Continuity of Care Document standard that incorporates aspects of ASTM's CCR content model in HL7 CDA-R2. A new standard under development by HL7 is Fast Health Interoperable Resources (FHIR). Its focus is on speed, and ease of, implementation using web-standards, such as XML, JSON, HTTP, and Atom feeds (Health Level 7 International, 2013b). The FHIR standard incorporates many features for health information models discussed in previous chapters e.g. an entry as a discrete, logical unit of medical data. A normative version of the standard is hoped for by 2015.

The International Organisation for Standardisation (ISO) produced ISO 18308 Health Informatics - Requirements for an Electronic Health Record Architecture. As an international standard it specifies the architectural requirements for an electronic health record (EHR) e.g. "the EHR shall be able to represent actions considered, planned and/or performed" (ISO, 2008b). As such it does not detail a data structure to be used. However, openEHR, discussed previously, uses this standard extensively.

EN 13606: 2007 Health Informatics - Electronic Health Record Communication is a standard from the European Committee for Standardisation (CEN) and also now published by ISO. It draws heavily on aspects of openEHR, especially in its use of archetypes to represent clinical information. As with the CCR it is not a full electronic health record (EHR) standard, but rather a standard for the exchange of EHR extracts. It does not specify any implementation, or data structure, requirements but is hierarchical in nature and the technical approach is generally the same as that of openEHR.

This is only a sample of ways health information may be stored or communicated. It is with these types of models that the stakeholders for this research (GPs) will be presented with health information. All models reviewed here are hierarchical in nature. The models use container sections to hold entries of the same type. Those that are implemented, use tree data structures to implement this container hierarchy. For standards that do not specify an implementation, they can be implemented in this way. It is not surprising that this type of data structure is used in these models as trees are easily extendible and able to be processed by machine. With an understanding of common ways of modelling health information, we move to modelling Quality Criteria (QC) with a view to using these models for processing QC over health information.

4.3 Health Information Model Used For This Research

As stated in Chapter 2 section 2.4 the health information focus in this research is health information used out-of-context. It is information that could be an aggregation from multiple sources, in line with the direction proposed for future access to health information (NZ National Health IT Board, 2010). It is information for use at point of care. Therefore, the information model for this research is conceptually, a view, or fragment, of information made available from multiple sources (potentially), and presented at point of care. For example, in an unscheduled, or emergency care setting where the patient's health information is needed, but is being used outside of the context for which it was originally gathered.

For potential health information that the Quality Criteria Model (QCM) is applied to, the same three areas of health information as used in the interview process (please refer to Chapter 3, Section 3.5.4), were used. To reiterate, these were:

1. Adverse reactions
2. Medical history
3. Medicines list

These areas represent the types of entries contained in this information model. As shown later in Figure 4.5, just as a Quality Criteria (QC) can not exist without an associated record entry, an entry can not exist without an encapsulating record. The record, treated here as an information fragment, provides structure and meaning for the entries it contains.

For this purpose the record, or record extract, can be modelled, and contains the following elements:

- **ID**: A unique identifier for the entire record i.e. a unique identifier for the collection of entries.
- **Version**: The iteration number for the record.
- **Date and Time**: When was the record created.
- **Patient**: Who is this record for.
- **Section/Container for adverse events**: A list of alerts pertaining to any adverse drug reactions this patient may have experienced.

- **Section/Container for medical history**: A list of entries with medical events for this patient.
- **Section/Container for medicines**: A list of medications the patient has been prescribed.
- **Actors**: A list of actors involved with this patient e.g. health practitioners or organisations that have contributed entries to this record.
- **References**: A list of references for this record e.g. any external links, relevant comments for entries or electronic signatures provided with the record.

In summary, the health information model for this project provides a means to an end. The health information model is a vehicle for delivering health information to the GP, that is relevant to the current context of need, at point of care. The model used ensures health information is available to the Quality Criteria Model for processing of Quality Criteria.

4.4 Quality Criteria Model - Background

The identified Quality Criteria (QC) from Chapter 3 are modelled into a formal structure that can be used to represent these identified QC. Once a QC is determined, it becomes a Health Information Quality Indicator (HIQI). The Quality Criteria Model (QCM) thus developed is used to express QC i.e. each QC can be represented using the QCM. From this unified model all QC can be represented by using different parts of the QCM, dependant on what QC is being represented at the time. Put another way, not all parts of the model are needed to represent each specific QC. Before moving to this however, it first must be clear as to how the QC came about, what they conceptually represent, and how they were interpreted. A brief review of this is now given.

The QC were investigated from the standpoint of the "importance in knowing" (please refer to Chapter 3, Section 3.2 for definition) in aiding in the assessment of health information quality. As well, the addition of new criteria was on the basis of "what additional information, that could be provided with health information, would aid in the assessment of that health information". This is demonstrated from examples of questions asked during interviews such as:

> "When thinking about believing recorded medical information, what background information on the entries would persuade you that using this information would be safe?"

> "What is it about recorded information on a patient's drug allergy status that would provide you with confidence that the information can be safely relied upon and used in clinical decisions?"

It is also clear from the rating exercise undertaken by the GPs. The GPs were asked to rate criteria for their "importance in knowing" when wanting to use the health information they are presented with. For example for Criterion 4 (C4: Attestation)[1] this is questioned as follows:

> "How important is it to know that the entry in the record has been attested to?"

and for Criterion 1 (C1: Contradictory Entries):

> "How important is it to know that an entry in the record is contradicted by another entry?"

The approach leads to the need to model QC in terms of their absence or presence, or known/not known status. With this type of indicator the GP has additional information, previously identified as being important to know i.e. Quality Criteria, with which to assess the quality of the health information they are presented with.

This approach is different from approaches presented in literature on data quality where data quality dimensions, or attributes, are evaluated to a score, value, or measure of some type. For example Scannapieco et al 2004 (Scannapieco et al., 2004) created the Data and Data Quality (D^2Q) model. The D^2Q is a semistructured model that extends and evaluates data from a tree structured data model along four data quality dimensions. These dimensions were accuracy, completeness consistency and currency. The data model that the D^2Q operated over was derived from data aggregated from geographically distributed, autonomous organisations that had a need to share data. The D^2Q would associate a measure value (from a specified domain of values) for a quality dimension against nodes in the data model. For example an address value from the data model would be associated to Address_Quality in the D^2Q which would then be extended with nodes for Address_Accuracy, Address_Currency etc. In turn these nodes would be assigned a value as an attribute e.g. Address_Accuracy[high].

[1] The criterion has been abbreviated to Attestation. For the rating exercise this was actually listed as Attestation of entries recorded. Please refer to Table 3.7 for complete listing of non-abbreviated criteria names.

4.4.1 Rationale For "Importance In Knowing" Approach

This approach was taken after analysis and interpretation of the interviews conducted with the General Practitioners (GPs). From the analysis, seven conclusions emerged. First, the GPs wanted to know more about the health information they use e.g. how did the information make it into the health information and/or who did it. It was deduced that if an item of health information were given a quality score, value, or measure they would simply want to know how the score, value or measure made it into the health information, or who did the scoring, valuing or measuring. In other words, what is the quality of the quality measure?

Second, if a quality score were given, would it be of use e.g. if presented with a medicine in the information the patient had been prescribed, and it came with a score, on a scale of 0 to 1, of 0.64 for quality, is this of use? From a medicine listing there would be, at minimum, a medicine name and dosage. Is the missing 0.36 of quality in the prescribed medicine, the prescribed dosage or both? If quality were broken down into component dimensions e.g. accuracy or completeness, and the same information was scored at 0.64 accurate or 0.64 complete, the same issue exists. Which part is inaccurate or, if 0.36 incomplete, what is missing? An approach could be to score every data item down to a fine grain level. However, this would "overload" the GP with information, information they need to be able to assimilate within a short time frame e.g. a fifteen minute consultation.

Third, there is the difficulty of determining completeness, that there is no more relevant information to find. Methods for this are time consuming and usually involve comparisons of records (Parkin and Bray, 2009). Noumeir (2012) goes so far as to state that there is no way of automatically testing for this (Noumeir, 2012).

Fourth, is the nature of the dimensions for quality. The literature reviewed in Chapter 2 identified many quality dimensions used when trying to evaluate data quality. Many of these dimensions moved beyond accuracy and completeness e.g. Objectivity (data being unbiased, objective) (Wang and Strong, 1996), Competence (of counter party) (Mayer et al., 1995), Credibility (of data in data store) (Vassiliadis et al., 2000; Jarke and Vassiliou, 1997), and Believability (data is regarded as true and credible) (Pipino et al., 2002; Wang and Strong, 1996). The theoretical framework developed in Chapter 3, Section 3.6 mapped the Quality Criteria (QC) to quality dimensions from Wang and Strong's (1996) framework. For example, QC C4: Attestation was mapped to dimensions that included Believability, while QC C3: Organisational reputation was mapped to dimensions that included Objectivity. This was done with the intent to then use the mapped dimensions in a computation for quality of the data.

Fifth, along with doubts regarding the usefulness of providing a score for quality for the health information, the dimensions as inputs for a computation, mapped in the theoretical framework, were also seen as problematic. The dimensions are too abstract, or too removed from the information, to be seen as useful for the intended audience. The group of GPs interviewed showed that they wanted a very direct relationship with the information and its source.

Sixth, a further issue was evaluation of dimensions. For example, the dimension of Competence is mapped to QC such as C11: Clinician experience and C10: Clinician qualifications. Using these criteria as inputs for the Competence dimension would mean a decision would be taken as to how much experience, and what level of qualification would influence the Competence dimension. The level of influence would be a subjective evaluation, different for different GPs.

Last, is the issue of loss of information should a computation be used. If each dimension were not attached to each data item then compromise would be needed to determine which dimensions would out-weigh others, with consequent loss of information. This issue is illustrated in the work of Jang et al (1995). They were addressing a data quality judgement problem and developed a data quality reasoner. This used a dominance hierarchy of quality dimensions and a reduction algorithm where, based on the dominance relationship, quality dimensions were subsumed as the reasoner progressed over the available quality dimensions (Jang et al., 1995). For example, for the three quality dimensions Credibility, Temporal-effect, and Semantic consistency, when Temporal-effect had a value of tolerable, it would be subsumed when Credibility had a value of high. The Temporal-effect dimension is effectively lost in the computation. As Jang et al (1995) point out this sometimes leads to counter intuitive results.

From these conclusions the decision was taken to leave the evaluation of whether to use the health information i.e. is the information fit for purpose, to the GP, but provide additional information to the user via the identified QC. The GPs have a very specific focus and that is determining whether any information that is available to them is fit to use, safe to use, and will not compromise patient safety. Hence a quantitative score was seen as of little value.

This is not to say that the Quality Criteria Model developed here can not be extended to evaluate the QC into a domain of applicable values for each evaluated criterion. Rather, that the goal was to give an indicator to the GP that these QC are absent or present, known or unknown, and allow the GP to use these results to aid in the assessment of the information quality, and therefore its fitness for purpose.

The following section details a conceptual Quality Criteria Model (QCM) that can be used to describe QC. The QCM is not only used to describe QC but also provides a

structure to represent QC such that the QC can be used computationally.

4.4.2 Quality Criteria Model - Development

Creation of the theoretical framework in Chapter 3, Section 3.6 provided a deep exploration of the Quality Criteria (QC), their relationships with information and actors, and their potential evaluation. As part of the Quality Criteria Model (QCM) development, interaction and relationships between QC are explored, along with components that form QC.

The overarching relationships between QC, data entry, and record are as follows. Quality Criteria (QC) are applied to entries in a medical record. This is the relationship between QC and entries. Entries must be contained within a medical record, or record fragment. The relationship between QC, entry, and record is conceptualised in Figure 4.5. This can be expressed as: a QC can not exist without an associated record entry, so to a record entry can not exist without a containing record. Whether that record is complete or only a fragment, or view, of relevant information from a larger record, is irrelevant for this research. The important point is there is a record that contains entries that QC are applied to.

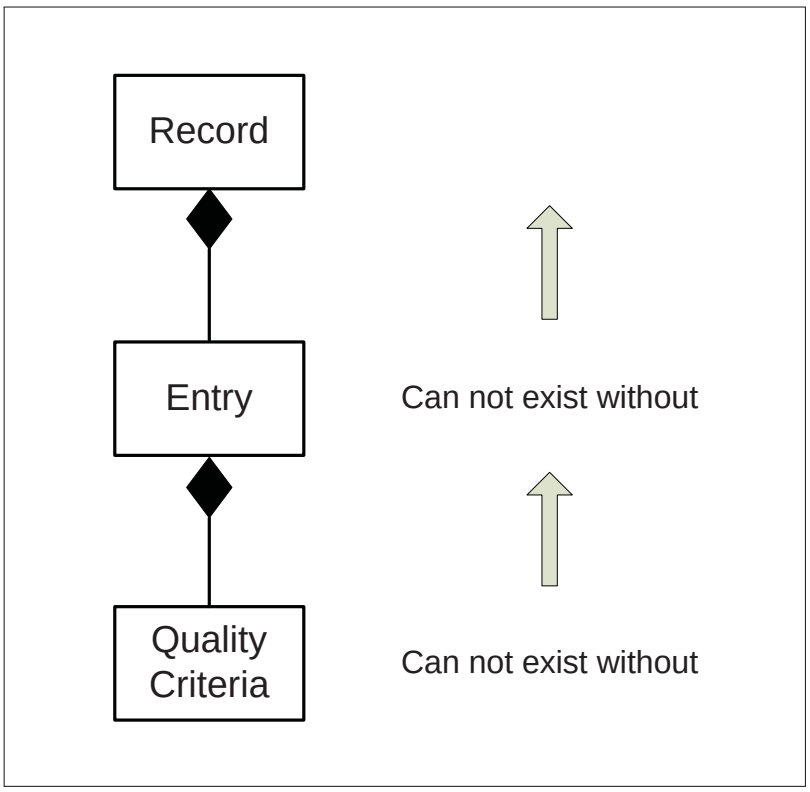


Figure 4.5: Quality Criteria and the EHR

QC are categorised further, building on and extending categories described in Chapter 3, Sections 3.6.1 and 3.6.2. As to QC composition, QC must identify the entry they belong to and a source for that entry, as reflected by analysis and interpretation in Chapter 3, Sections 3.5.6 and 3.5.7. The source is a required component of any entry. The source is defined as an Actor. An Actor in the context of an entry can be a person, an organisation, or an information system. The definition for Actor is taken from ASTM's Standard Specification for Continuity of Care Record (CCR) (ASTM, 2005). The additional categories, QC for the entry itself and QC for information on entry source, define increased granularity in describing QC and QC groupings.

QC grouped by what they say about an entry itself are shown in Table 4.1a. QC grouped by what they say about an entry's source are shown in Table 4.1b. Both groupings use QC short names.

Table 4.1: Quality Criteria - groupings.

(a) Quality Criteria - information about record entry.

QC Grouping On Providing Information (metadata) On Record Entry
• C4: Attestation
• C8: Contextual narrative
• C9: Further information
• C12: Data age
• C13: Change narrative
• C14: Information captured

(b) Quality Criteria - information about record entry source.

QC Grouping On Providing Information (metadata) On Record Entry Source
• C2: Organisational expertise
• C3: Organisational reputation
• C6: Personal experience
• C7: Attestor role
• C10: Clinician qualification
• C11: Clinician experience
• C15: Clinician role

One other grouping of QC was produced. This is on the Electronic Health Record (EHR), or extract from an EHR, as a whole. The EHR can be considered a collection of entries. Two of the QC require more than one entry to be realised. These two are:

- C1: Contradictory entries
- C5: Consistency/cohesion

C1: Contradictory entries requires, at minimum, two entries to contradict each other. C5: Internal consistency or cohesion of record (long name for this QC) directly references the record, or record extract, as a whole. Consistency/cohesion is across multiple entries. The consequence is that these two QC are realised against the EHR document, or EHR extract as a whole. The groupings described help inform the development of the Quality Criteria Model as they place the QC in context of the type of information

they are related to i.e. an entry, a source for the entry, and in some cases the EHR itself.

The three groupings are dimensions from the health information through which QC emerge i.e. three dimensions, entry, source, and EHR document. Each QC only belongs to one grouping, with QC being mutually exclusive along the three dimensions. Figure 4.6 is a graphical illustration of the three dimensions for QC from health information.

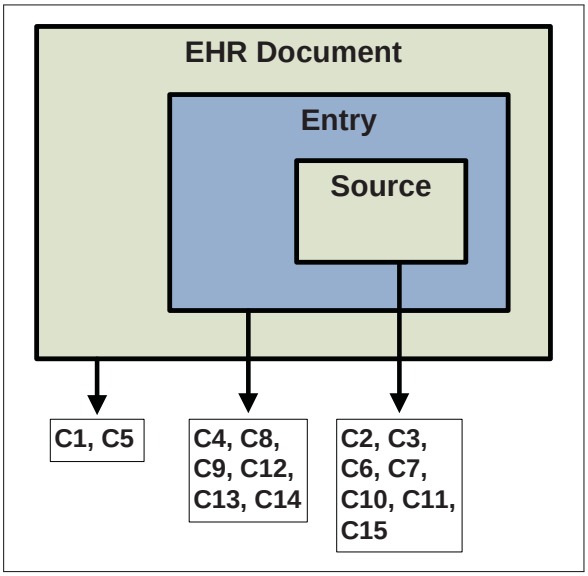


Figure 4.6: Three dimensions for health information for QC.

The Quality Criteria Model (QCM) must take the features/relationships, described into account. The QCM model is required, as it will be necessary to construct Quality Criteria (QC) in a form that is capable of being applied to, and processed with, health information. In essence the QCM allows for the QC to be used as a basis for Health Information Quality Indicators (HIQI). HIQI being realised, or instanced, QC.

4.4.3 Quality Criteria - Theoretical Formalisation

The first step is to formalise how QC are composed. Expression (4.1) shows a QC (criteria C) model composed of an entry (E), and and actor (A). The entry represents those QC associated with the entry (information), and the actor with those QC associated with the actor. Using previous analysis, groupings and relationships for QC, QC require an entry within health information to operate over. With an entry, is a source. The source is called an actor. This is clearer when a QC is instanced upon being realised, or detected as being present. Expression (4.2) shows this, where the entry is now a set of the QC that are associated with the entry, and the actor is now a set of QC

associated with the actor.

$$C\{E, A\} \tag{4.1}$$

$$C\{E(c_1, c_2 \ldots c_n), A(c_1, c_2, \ldots c_m)\} \tag{4.2}$$

When multiple QC are yielded for an entry, or actor, the QC is a composite, or nested i.e. the defining criterion, that initiated the search for others, is composed of others inside the sets for entry and actor. If an entry, or actor, only yields a single realised criterion, the criterion is said to be atomic. An example using the criterion C4: Attestation illustrates this. Expression (4.3) shows a composite criteria, where C4: Attestation for an entry has been found, along with C7: Attestor's role. It may seem odd that C4 is not in the set for the entry. This is because the two sets, for entry and actor, require an identifier. The identifier of C4 is enough to link it to an entry, as there are no criterion that belong in both the entry grouping and the actor grouping. Expression (4.4) models an atomic version for C4: Attestation, where the only criterion realised is that the entry is attested to e.g. the attestor's role, and anything else regarding the entry itself, or the actor, is unknown.

$$C4\{e(\varnothing), a(C7)\} \tag{4.3}$$

$$C4\{e(\varnothing), a(\varnothing)\} \tag{4.4}$$

Each entry, and entry source, is searched in turn. Should this search realise a QC, the QC is instanced and the entry and source are searched for further QC. Using the previous example the process would be as follows. C4: Attestation is detected, instanced and added, as shown in Expression (4.5). Next, the entry and source are searched for further QC instances, and in this case, the role of the actor attesting to the entry is found (C7: Attestor role). This QC is instantiated and added, as shown in Expression (4.6). Further searching yields that there is a contextual narrative for the event for the entry (C8: Contextual narrative). This is instanced and added, as in Expression (4.7). It is also found that there is a second source (actor) for the entry, as it is possible to have more than one source for information in an entry. It is found that qualification/s are listed for this second actor (C10: Clinician qualifications), along with their years of experience. The C10 and C11 criteria have been found, are instanced, and added to the new actor as in Expression (4.8). This is the processing model for QC, and continues

until no more QC are found for the entry, or its associated actors. Processing then moves to the next entry. The expressions used represent the sets of realised QC for the entry and the entry's source/s (actors). Figure 4.7 represents a conceptual view of the process.

$$C4\{e(\varnothing), a(\varnothing)\} \tag{4.5}$$

$$C4\{e(\varnothing), a(C7)\} \tag{4.6}$$

$$C4\{e(C8), a(C7)\} \tag{4.7}$$

$$C4\{e(C8), \{(a_1(C7)), (a_2(C10), (C11))\}\} \tag{4.8}$$

Figure 4.7 depicts an example for processing of an entry. It shows only those QC that are entry related, and objectively assessed, being initially searched for. However, it is possible that actor related QC can also be found without an entry related QC being found e.g. C10: Clinician qualification could be found for an entry's source, but nothing further is known of the entry itself. This is a potential example and not all possible outcomes are represented. For example, C9: Further information could also have C10: Clinician qualification and/or C11: Clinician experience detected for its source. Also, without additional information, the subjectively evaluated QC, C2: Organisational Expertise, C3: Organisational reputation, and C6: Past personal experience, can not be evaluated. The additional information would need to be a way of determining reputation or expertise, such as pre-defining these ratings for a system to use.

4.4.4 Quality Criteria - Applying the Formalisation

Quality Criteria (QC) are now modelled in detail, where required features needed to realise, or instantiate, QC are defined. The features of the QCM used for an individual QC are dependent on the QC being modelled. For example, a QC instanced for a source of an entry e.g. C10: Clinician qualifications, will not need the ability to record if there was a contextual narrative for an entry i.e. C8: Contextual narrative. Figures 4.8a and 4.8b show a graphical representation of the Quality Criteria Model (QCM) in two parts.

Elements of the model are defined as follows:

- **Label**: An identifier for the QC being instantiated e.g. C4: Attestation, or C4.

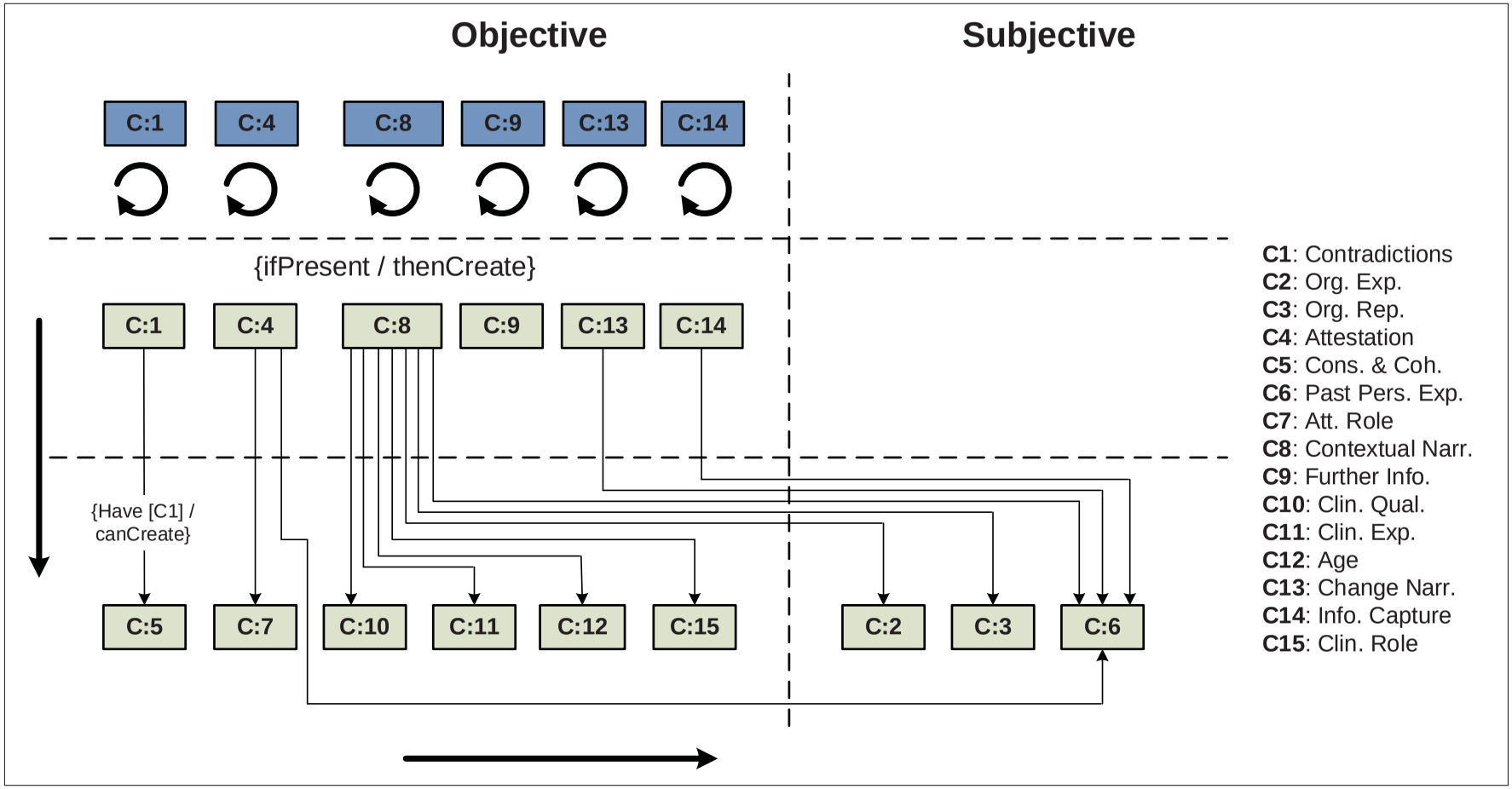
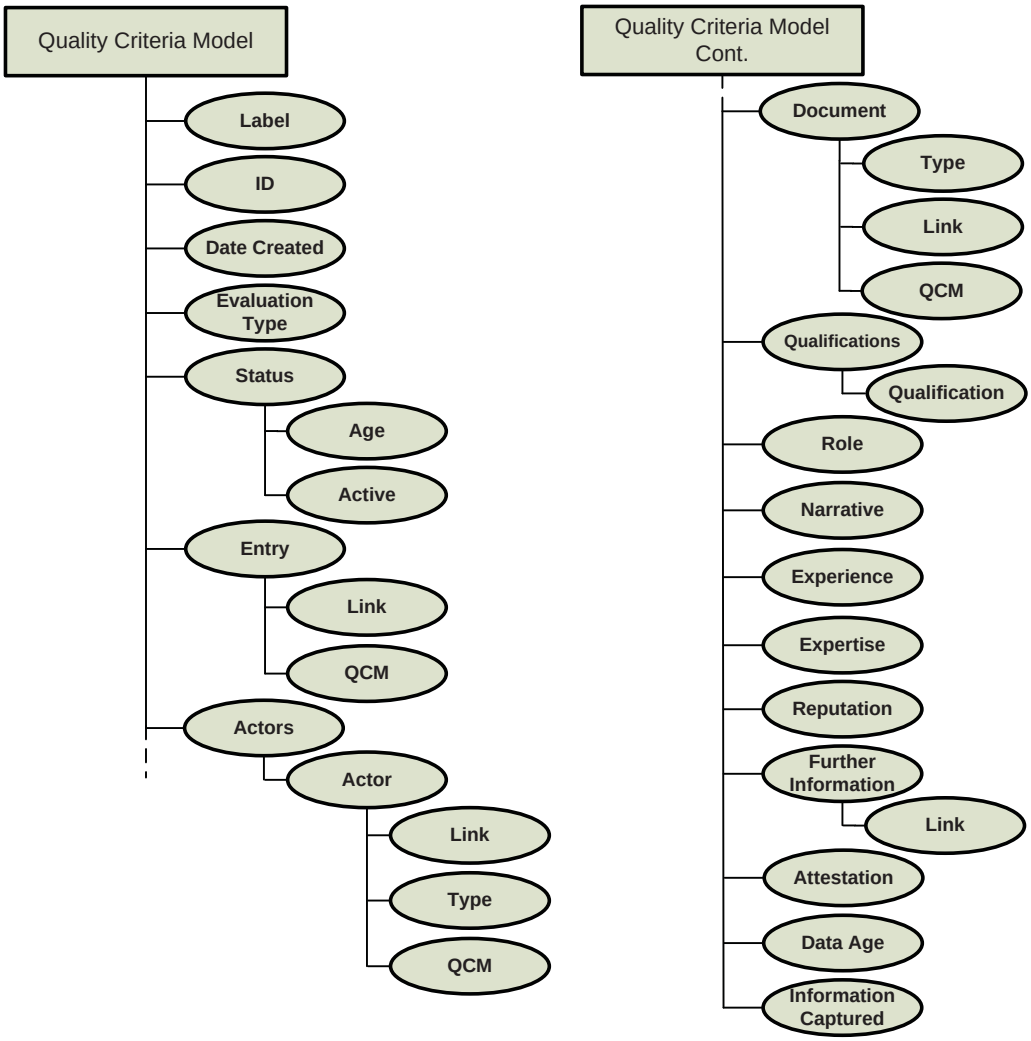


Figure 4.7: QC - Conceptual processing model.



(a) Quality Criteria Model - Part 1. **(b)** Quality Criteria Model - Part 2.

Figure 4.8: Quality Criteria Model

- **ID**: A unique identifier for the QC.
- **Date Created**: The date the QC was created.
- **Evaluation Type**: Is the QC objectively or subjectively evaluated.
- **Status**: Contains two sub-elements; Age and Active. Age represents the amount of time since the QC was last evaluated. For active this represents whether the QC is active or not i.e. is relevant in the current context.
- **Entry**: Contains two sub-elements; Link and QCM. Link contains a pointer to the unique identifier for the entry. Each QC may only be associated with one entry.

- **Actors**: A container for an actor, or actors, as it is possible to have multiple actors for a single entry. An Actor element contains three sub-elements; Link, Type, and QCM. Link is a pointer to the actor. Type refers to the type of actor it is i.e. a person, an organisation, or an information system. QCM is a container for all instanced QC for this actor.

- **Document**: Contains three sub-elements; Type, Link, and QCM. Type is the type of record e.g. discharge summary, HL7 message, or Shared Care Record. Link is a pointer to the unique record ID. QCM is the container for instanced QC for the record document as a whole e.g. C1: Contradictory entries or C5: Constenct/cohesion.

- **Qualifications**: Has one sub-element; Qualification. Container for a list of qualifications and actor (person for this type of data) may hold.

- **Role**: This will be either the role of a clinician i.e. C15: Clinician role or C7: Attestor role. Role is dependent on the identifier for the QC being instanced. For example if the QC were C7: Attestor role (what the QC is labelled as), then this element would contain the role for the attestor to the entry, or a pointer to where the role can be found.

- **Narrative**: There are two types of narrative. A contextual narrative (C8) and a change narrative (C13). The identifying label for the instanced QC will determine what narrative appears here. Alternatively, this could contain a pointer to a narrative of the correct type, or could be an indicator that a narrative exists for the entry.

- **Experience**: There are two types of experience. C10: Clinician experience or C6: Personal experience. As with Narrative, what appears in the Experience element is dependent on the instance of QC i.e. if the Label for the QC were C6: (personal experience), the experience element would contain a value for the paste personal experience with the actor.

- **Expertise**: Should the label for the instantiated QC be C2: Organisational expertise, the expertise element will contain the organisation's level of expertise.

- **Reputation**: As for Expertise, but the element would have the organisation's reputation rating.

- **Further Information**: Container for pointers to further information for the entry. Alternatively, could contain the further information.

- **Attestation**: Binary value for entry attestation. Indicates an entry attested to, or not.
- **Data Age**: Element contains a value for the age of the data entry.
- **Information Captured**: Element contains data on where the entry was captured. Alternatively, could contain a pointer to data on where the entry was captured.

As with which QC are realised, depending on whether the QC are associated with an entry, the source for an entry, or the record document as a whole, the Quality Criteria Model (QCM) highlights dependency relationships between elements of the model. Dependencies in the QCM are based on the type of QC being instantiated. The type determines what elements are used from the QCM. For example, where a QC is realised, and instantiation begins, it is given a label, or type. If the QC were labelled C6 (C6: Personal experience), then the experience element from the QCM would be used and given a value for past personal experience (not clinician experience) with the source of the entry. In this respect the Experience element from the model is polymorphic. Designing the QCM in this way reduces the number of elements required to represent all QC. These polymorphic elements are shown in Figure 4.9.

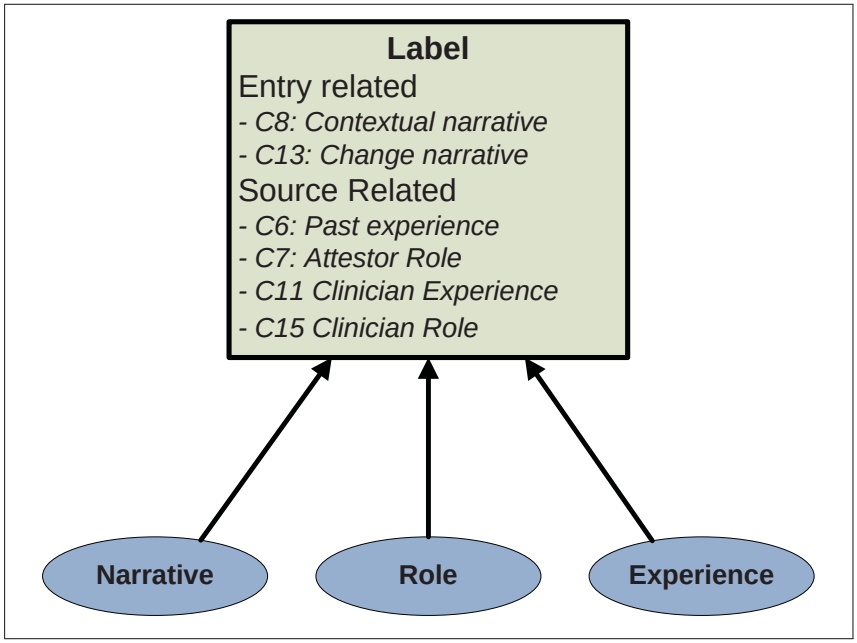


Figure 4.9: QCM - Polymorphic elements dependent on QC type.

Beyond the polymorphic elements, other elements used from the QCM are dependent on the identifier label when a QC is instanced. This is seen with the groupings of QC

based on association with the entry, entry source, or record document. For example if the QC is related to the source, then elements such as Qualification may be used and not elements such as Narrative. One other type of dependency is found in the QCM. This is for C7: Attestor role. The C7:Attestor role QC can not occur on its own. It being dependent on C4: Attestation being realised first.

An illustrated example for detecting the presence of Quality Criteria (QC), realising a QC on detection, and population of the Quality Criteria Model (QCM) to produce a Health Information Quality Indicator (HIQI), helps in understanding the process. The example will be the same as used for the description of the processing model and how realised QC are built up in sets associated with an entry and/or an entry's source.

To begin, there is some health information organised in sections with discrete entries within the section. For example, a medications section contains individual medication entries. The process begins with each entry examined in turn. In the example shown in Figure 4.10 each medication entry is examined to determine if Quality Criteria are present.

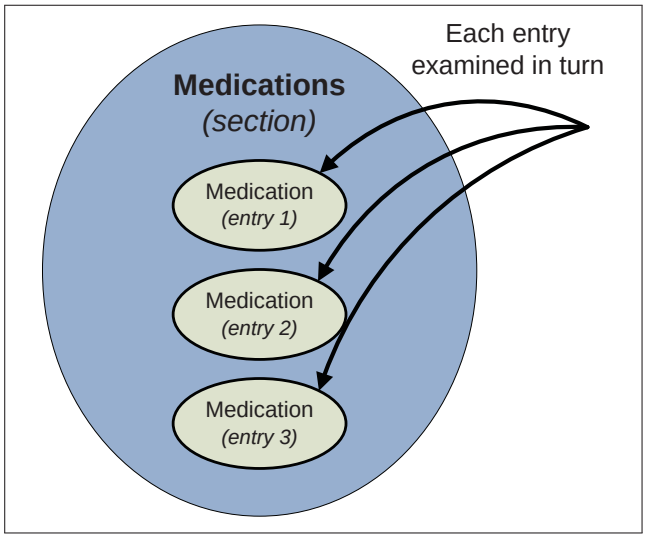


Figure 4.10: Quality Criteria processing - entries examined.

In Figure 4.11 the example finds that a medication entry has been attested to. Detection of attestation could happen in several ways. In Figure 4.11 the source could have a role of "Attestor" (Continuity of Care Record from ASTM International), or the entry itself could have been attested to, as with HL7's Clinical Document Architecture (CDA). How it is represented in the health information is not important, the fact it has been attested to, is. The QC C4: Attestation has been realised. QC C4 is instanced, using the Quality Criteria Model (QCM). The realised QC is now an indicator that attestation is present for the entry. The QCM representing the indicator is shown in Figure 4.12.

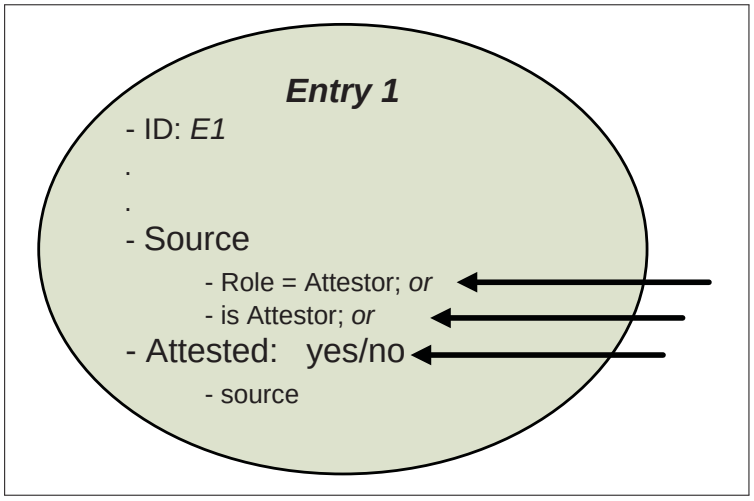


Figure 4.11: Quality Criteria processing - attestation detected.

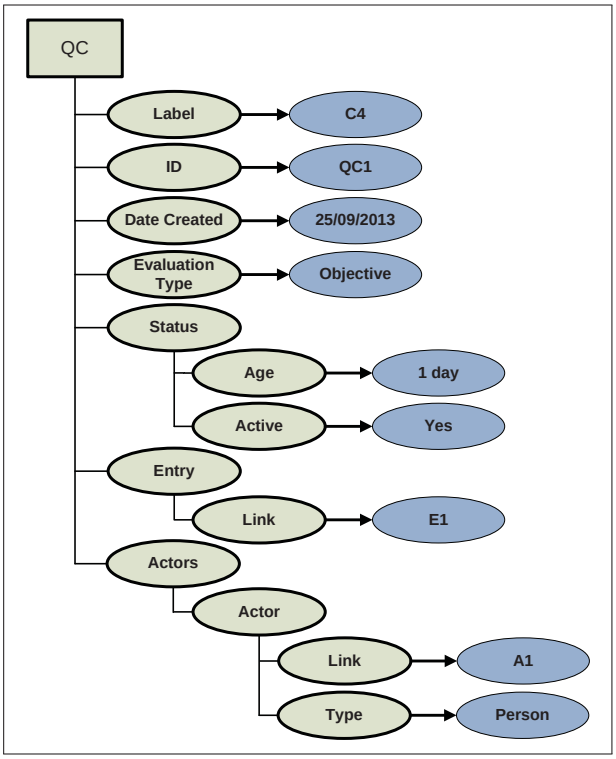


Figure 4.12: Quality Criteria processing - C4: Attestation instanced.

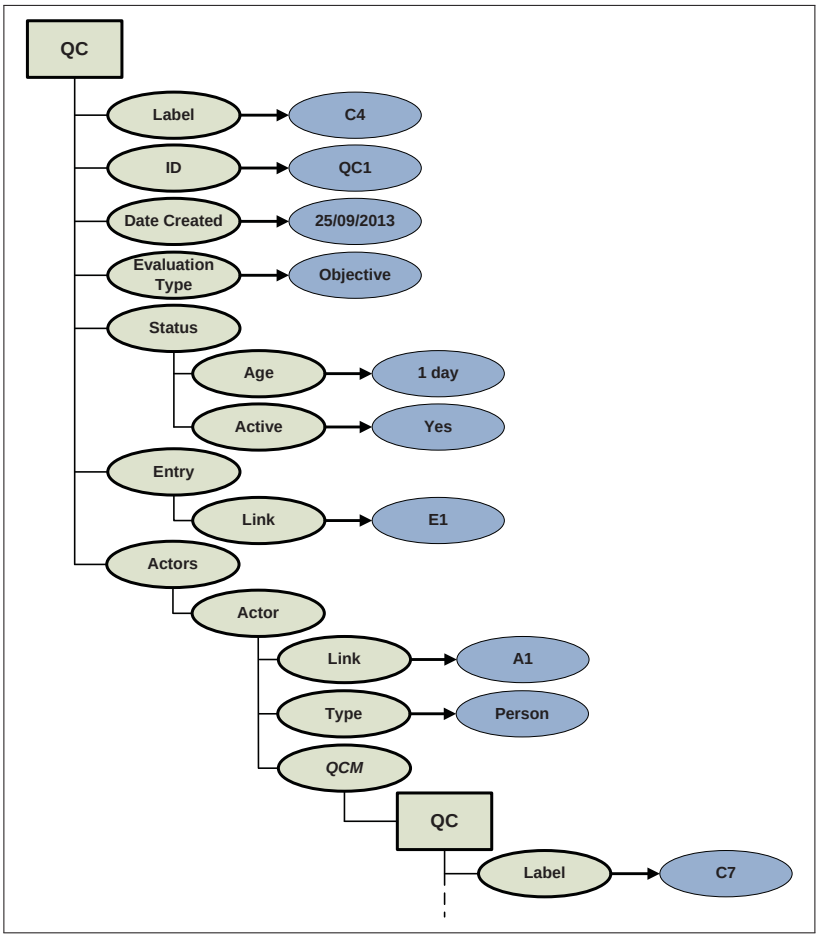


Figure 4.13: Quality Criteria processing - C7: Attestor role instanced and added.

With the detection of a QC for the entry, a search of the entry is triggered, where all QC a looked for within the entry. QC C7: Attestor role is detected. C7 is instantiated using the QCM and associated with the source (actor) for the entry, as shown in Figure 4.13. The C7 instance of QC is extended from the point in the QCM used to hold instances of realised QC for the actor. The content for the new QC (C7) is shown in Figure 4.14 where the value for the Role element is included. The instance of C7 includes the link to the record entry as well as the link to the actor. This may seem unnecessary as the instance is contained within another instance that already contains the links. However, it is done in this way because if the nested structure were to be flattened, or if all QC instances were stripped off the record for secondary analysis purposes, then this way it is possible to identify the entry, and actor, the QC is for. The Document element, and sub-elements (Link), also allows for identification of the originating record, or record extract, as well as QC instances that are record specific i.e. C1: Contradictory entries and C5: Consistency/cohesion. The Document element is not shown in Figure 4.14 as

it is not relevant to the detection example.

$$C4\{e(\varnothing), a(\varnothing)\} \tag{4.9}$$

$$C4\{e(\varnothing), a(C7)\} \tag{4.10}$$

$$C4\{e(C8), \{(a_1(C7)), (a_2(C10), (C11))\}\} \tag{4.11}$$

Figure 4.12 shows the process at the same stage as Expression (4.9), taken from the previously used QC composition example. Figure 4.13 shows the equivalent stage to Expression (4.10). The illustrated example for the Quality Criteria Model (QCM) proceeds until the final stage is reached, represented by Expression (4.11) and Figure 4.15 Date Created, Evaluation Type, and Status elements have been removed from this Figure. At this point no further Quality Criteria (QC) are realised for the entry being examined.

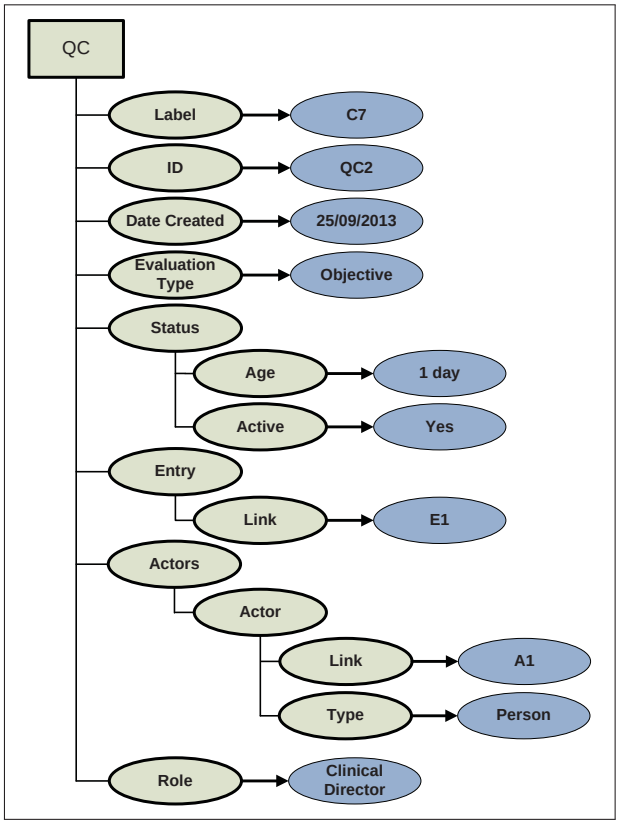


Figure 4.14: Quality Criteria processing - C7: Attestor role content.

A final figure, Figure 4.16, for this example shows the content for the instanced Quality Criteria (QC), C8: Contextual narrative. This illustrates an entry group QC, and also how the type of QC instanced influences the content. The C8 QC extends the QCM container for the entry, with the Narrative element taking on the specialised type of contextual narrative. The Narrative element could contain the text for the contextual narrative, a pointer to the contextual narrative or a boolean value that represents that a contextual narrative exists for this entry. Figure 4.16 shows the instanced C8 QC.

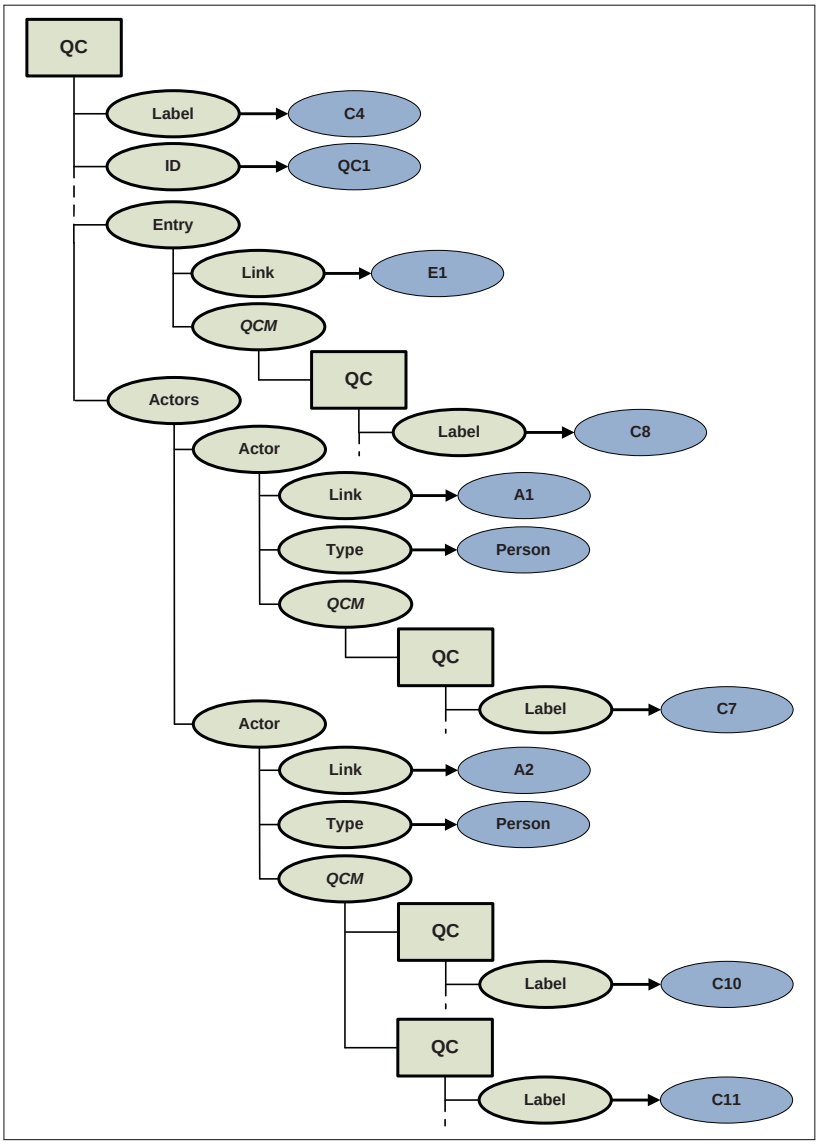


Figure 4.15: Quality Criteria processing - result of full search of entry.

The C8 instance has an actor link, to a second actor (id A2 in the example). The example also shows that QC are realised for this actor; C10: Clinician qualification and C11:Clinician experience. The actor link for C8 provides a reference to the actor, that

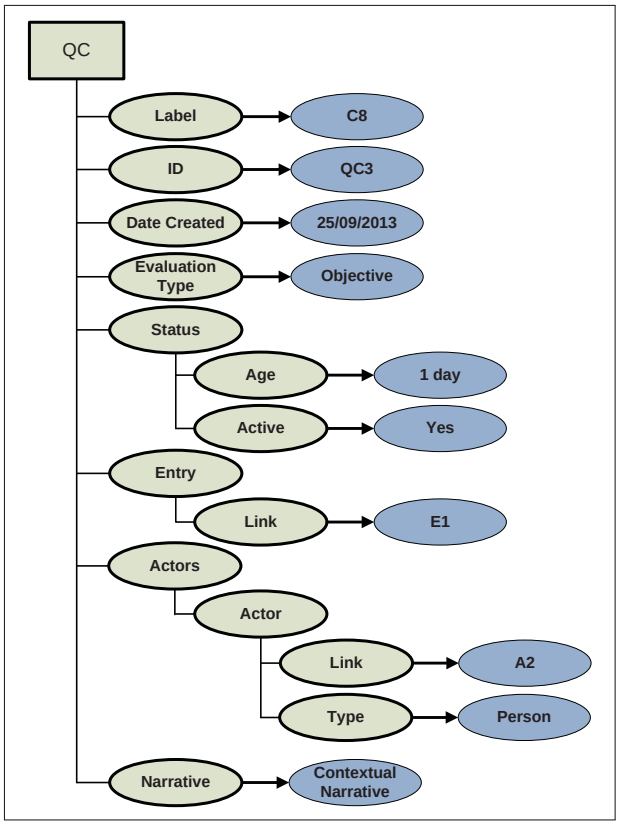


Figure 4.16: Quality Criteria processing - C8: Contextual narrative content.

in turn allows for an identification that the actor's qualifications and experience are known, for the source that provided C8: Contextual narrative.

4.4.5 The Quality Criteria Model - Summary

The Quality Criteria Model (QCM) provides the following features:

- A recursive structure that allows nesting of Quality Criteria (QC) instances within a containing QC instance.
- A single QC instance of the QCM contains all realised QC for an entry i.e. at the top level of the nested structure, one QC for each entry.
- The recursive structure allows for capturing information such as source of source, should this be required. This could be necessary as the actor recording the source for an entry may not be the actual source e.g. clinical notes transcribed by a third party.

- Unique identification for every QC instance, its originating entry, and/or document, allows for flattening of recursive structure should separation of QC instances be required e.g. for secondary analysis of QC, or for application of the QCM within a system that does not require a nested format.
- Based on features from Object-Oriented (OO) methodology, such as type specialisation and polymorphic elements, the number of model elements required to represent all QC is reduced.
- Use of pointers to data elements reduces data redundancy e.g. pointers to entries and actors within a record, rather than needing to record them again within the QC (data normalisation).

The features of the QCM, and its method of development, allow for implementation using a variety of methods such as XML-based methods or Object-Oriented methods. The variety of implementation methods is in line with other models, from standards and currently used Electronic Health Records, used with health information; as was highlighted when health information models were reviewed.

A second model was developed; a processing model for use when applying the QCM to health information entries/documents. Examples were given to illustrate the processing model's operation. The model implements a breadth, then depth search algorithm. The breadth search sequentially looks for QC within entries in turn i.e. for argument's sake it will look for C1 across all entries, then C2, and C3 etc. The depth search is triggered on an entry on detection of the first QC being realised for the entry being searched. The depth search looks for all instances of QC within an entry. On completion of the depth search of an entry, the algorithm returns to the breadth search for the next available entry.

Processing efficiency is enhanced when the depth search of an entry is triggered. For example, should the first QC realised for an entry by the breadth search be C4, the depth search needn't look for C1, C2, or C3 as the breadth search has already looked for these in prior sequences. The approach requires that the breadth search be targeted at elements known as the location for the QC being searched for in the current sequence, so as to negate the need to search the entire entry. For example, when the breadth search reaches C4: Attestation in the sequence of QC, it should know how attestation is provided for in the type of record, and entry, being searched. It would be looking for an instance of the Attester class with HL7, or a role of "attestor" for an actor in the CCR.

The QCM and the processing model are implementable and able to be described using methods used with other models for health information. The two models are also

implementable and able to be described using accepted techniques from the software development field.

4.5 Chapter Summary

Two models have been developed and described. These were the Quality Criteria Model (QCM) and the model for processing results of applying the QCM to health information. The second model provides enhanced information annotated to the health information elements. The enhanced information gives form to identified desired features of health information (Quality Criteria) discovered via the triangulation of sources. Processing health information to highlight these features, captured by the QCM, and deriving Health Information Quality Indicators (HIQI), is to assist the GP in their judgement of the fitness for use of the health information.

Chapter 5

Prototype Development

5.1 Chapter Overview

At the end of the previous chapter, Chapter 4, two models had been developed i.e. the Quality Criteria Model (QCM) and the the process model that applies the QCM over health information to realise Health Information Quality Indicators (HIQI). In this chapter, these two models are implemented in a prototype to demonstrate that the models are able to be used to detect the identified criteria in health information, instantiate them using the QCM, and add them to the health information as HIQI. HIQI are not user interface components, though could be represented this way. Rather, they are the product of a positive result of a search for QC. A prototype system is developed to demonstrate data structures and their use.

5.2 Rationale For Prototype

Criteria used by a group of GPs, to determine if health information they wish to use is fit for purpose, have been discovered. These criteria were analysed and a rating for each determined. The rating was based on the importance in knowing the criteria are present, to the GP (refer Chapter 3). There may also be value in indicating those not present. The criteria were formally defined using the Quality Criteria Model (QCM), to provide a format for recording, or instantiating, a QC. A second model was produced. This was a processing model that showed how health information could be examined to detect Quality Criteria and produce instances of the QCM to represent what was found in the health information (refer Chapter 4) and realise (HIQI).

As this research is a Proof of Concept (POC) study, the evolution described in the

previous paragraph is a natural progression from discovery to description to representation, via the described models. As the final part of the POC, it is now required that it be demonstrated that it is possible to implement both the Quality Criteria Model (QCM) and the processing model in a computer system. The system will demonstrate that it is feasible for health information to be processed to detect Quality Criteria (QC), record what is known about the criteria using the QCM, and lastly, to add this to the health information as HIQI, to be later made available to the GP. A prototype was developed to demonstrate the required functionality.

5.2.1 Why A Prototype?

There are many reasons for prototyping. The following are some of these:

- can demonstrate that a model, or concept, is able to be implemented within a system.
- to test a new design to simulate a full system or a material part of it.
- provide specifications for an actual working system, rather than a theoretical one.
- to test the validity, or usefulness, of a system or part thereof.
- are able to better technically understand the problem at hand and to quickly exclude approaches that don't work to focus on approaches that do work (National Instruments, 2009).
- to create discussion on the proposed solution the prototype is being developed for.
- to aid in patent applications or to help determine pricing for a new invention (National Instruments, 2009).

Of these, several make a prototype a logical choice for this research. First is the need to demonstrate that the previously developed models, the Quality Criteria Model (QCM) and processing model, are able to be implemented within a system. Second, is to better technically understand the problem and to allow quick development of solutions. Last, is to create discussion around the difficulties and opportunities that development of the prototype has raised, as part of the overall research questions.

5.3 Domain Environment

The domain for the prototype's intended functionality is briefly reviewed. The domain environment for the prototype is the use of health information at point of care. Specifically, General Practitioner's (GPs) use of information in a consultation with the patient they are seeing. This is out-of-context use of information (refer Section 2.4 for a definition for out-of-context use).

This could be information from different sources and in different formats such as information that has been aggregated, for a particular patient, at the point of care. It could be health information from the GP's Practice Management System or information the patient has brought with them e.g. the patient's medicines list. However, these processes, by which the GP gains access to the health information, is outside of the scope of this research. How the information is obtained is not important. What is of interest is that the GP has obtained electronic health information, and now must assess its quality, in order to determine its fitness for purpose.

5.4 Health Information Model To Be Used - The CCR

For the prototype, it is necessary to have electronic health information of some sort to operate over. For this purpose the American Society for Testing and Materials' (ASTM) Standard Specification for Continuity of Care Record (E2369-05) was chosen. This will be referred to as the CCR.

The ASTM International standard for CCR is defined as:

> "a core data set of the most relevant administrative, demographic, and clinical information facts about a patient's healthcare, covering one or more healthcare encounters" (ASTM, 2005)

It is further described as a means for aggregating all pertinent data for a patient to allow this to be carried forward to another healthcare setting to support the continuity of care for the patient (ASTM, 2005). The definitions fit well with the domain for health information use, described for this research.

The standard specifies these requirements using the eXtensible Markup Language (XML). XML defines the coding required when a CCR is created and describes the XML elements available to record health information. To facilitate this, a CCR XML schema file is used. This validates that any instance of a CCR created, conforms to

the CCR standard. The ability to validate a CCR document is important for prototype development. This is because using the CCR schema allows for validation that the health information input for the prototype is a conforming CCR document. It also allows for the processed output from the prototype to be validated. This confirms that what is output by the prototype still conforms to the CCR standard. The validation of input and output of health information for the prototype is important as it means the prototype system is certain what it is dealing with on input, and also allows the output to be in a form (CCR standard) that can further be used by systems expecting health information in a CCR format.

The CCR is divided into three sections; the Header, the Body, and the Footer. The Header defines document parameters and contains items such as version number, unique identifier, the patient the CCR is for and who has generated the CCR. The CCR Body contains patient clinical data. The Body section is the only one of the three that is explicitly named within a CCR document, as a single CCR Body element. The approach taken by ASTM is to treat each entry as a discrete data object with sections (containers) defined in the CCR to contain data objects of the same type. For example, in the Body section there is a single element called Alerts. Alerts are used to list and describe things such as a patient's drug allergy (ASTM, 2005). This in turn contains Alert entries as discrete data objects representing each individual alert for the patient. Another example is the Medications element, of the Body section, that is a container used to hold individual Medication entries for the patient.

The Footer of the CCR contains data objects that represent all of the actors involved in the CCR. It also lists information on external references, any text comments regarding entries and a list of signatures for the CCR document. The Footer section follows the same approach as the Body section with container elements holding the instances of the data. For example the Footer section has a single element Actors which contains individual Actor entries. The same is true for the other three containers in this section References, Comments and Signatures. This section is key to how the CCR's structure operates.

The structure allows for a normalisation of data within the CCR. Normalisation means a data object is stored only once within a CCR document and thereafter referenced via a link, reference or pointer. For example you can store data about an actor, who has contributed to the CCR in some way, as an Actor entry within Actors. This actor may be involved in multiple entries within the CCR e.g. may have entered an alert for this patient as well as being the actor who prescribed medications to the same patient. This actor is the source for both of these entries. The Alert entry contains a pointer to the Actor entry and the Medication entry contains a pointer to the same Actor entry. The

reference pointer approach is used in several health information models e.g. HL7's new FHIR standard, as well as in the development of the Quality Criteria Model (QCM) to be implemented in the prototype. The reference pointer structure is illustrated in Figure 5.1.

This brief description of the CCR is to illustrate the reason the CCR was chosen for use with the prototype. Other models and implementations were investigated for use, from HL7's Clinical Document Architecture (CDA) to openEHR. To summarise, the reasons for choosing the CCR were:

- the rationale for the CCRs creation closely follows this research's specified operating domain i.e. data (potentially aggregated from multiple sources) used at point of care, to follow a patient, to facilitate the patient's continuity of care.
- the treatment of entries as discrete data objects allowing for interrogation of entries.
- the ability to use pointers to extend data objects e.g. an Alert entry with a pointer to the Alert's source (actor), or a pointer to a generated Health Information Quality Indicator (HIQI), once a Quality Criteria (QC) is realised.
- the CCR's content model is to be used as the basis for New Zealand's (NZ) Health Information Exchange (HIE) Content Model, for interoperability of systems within the NZHS (HISO, 2012b).
- richness of CCR's clinical data representation.
- provision of XML-schema and implementation guide with the CCR standard.

Due to the way the QCM is structured, as well as the structure of many health information models, the QCM could be implemented in a variety of ways. For example, instead of the nested recursive structure that conveniently collects all Quality Criteria (QC) instances for any one entry, the recursive structure could be flattened, with each individual QC referenced by its unique ID attribute. The QCM always retains which entry, and document, any QC instance is related to. This could also be extended beyond document and entry, with additions to the QCM, having references to systems and repositories where QCM instances exist. Therefore it is not necessary to physically co-locate instances with their associated entries, and/or documents. CCR provides for internal/external references, as do the HL7 standards (both normative and in development) HL7 CDA version 2, CCD, and FHIR, and also the openEHR Archtype Object Model (openEHR Foundation, 2013a).

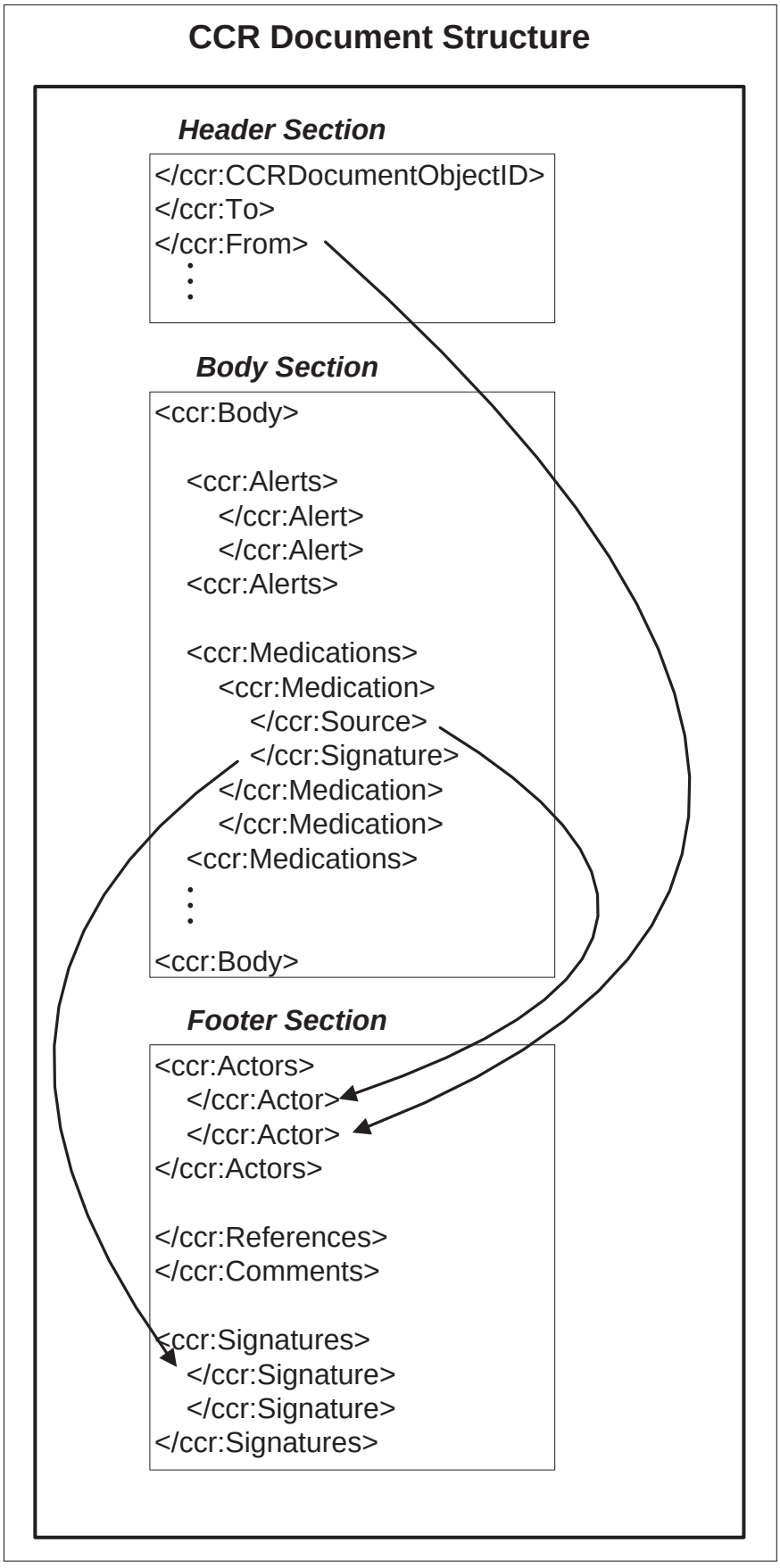


Figure 5.1: CCR document structure.

5.5 Prototype - Development Environment

The development environment was selected based on the functionality the prototype is required to deliver. Notice was taken of tools commonly used when developing software in the health information domain e.g. XML technologies such as XML-schemas, Object-Oriented (OO) programming languages such as Java and C#, and querying technologies such as XQuery/XPath, Structured Query Language (SQL), and Microsoft's .NET Framework's Language Integrated Query (LINQ) libraries.

The CCR standard for health information states that when the CCR is implemented in an electronic format, adherance to an XML schema is required (ASTM, 2005). XML was also chosen for implementation of the the Quality Criteria Model (QCM), developed in Chapter 4. XML was used for the following reasons:

- XML is a well established technology, and frequently used for implementation of structured, and semi-structured data, that includes implementations for Electronic Health Records e.g. CCR, and HL7 standards.
- XML-schema allow for validation of XML data structures that ensures correctness of data, and data structures being used. This aids system that use XML to be certain of the structures the system is to deal with. The CCR standard includes an XML-schema for this purpose.
- The ability to create user defined data types, and extend existing data types. Also, many in-built data types are available in XML.
- Multitude of tools available for creating, and manipulating, XML data and data structures.
- XML is both human readable and machine processable.

An OO programming language was chosen for development, specifically C# from .Net Framework version 4. The choice being due to familiarity with the OO programming paradigm, and its frequent use in development of health information systems. This frequent use can be seen where reference implementations of standards for health information are developed. For example, openEHR offer development tools written in Java, while HL7 offer reference implementations for standards (e.g. FHIR) written in Java, C#, and Delphi, all third generation OO programming languages. Further, open source projects, developed by third parties, provide implementations e.g. open source CCR implementation in C# (Ranjan, 2013).

C# was also chosen as it allowed access to the LINQ Application Programming Interface (API), specifically LINQ to XML functionality. Although familiar with XQuery/XPath languages that allow for querying, and accessing path structures in XML documents, LINQ libraries from the .Net Framework were used. This was for the following reasons:

- Is part of the core libraries of the .Net Framework, requiring no additional libraries added to the project, as is necessary to enable XQuery within C#.
- With LINQ, XML is instantiated directly as a first class object, and does not require the use of the Document Object Model (DOM), or the Simple API for XML (SAX) that require parsers, readers, and writers to manipulate XML.
- An XML Element object, or XML Document object, instantiated using LINQ, is directly accessible using the LINQ API, as the object is the data.
- Retains the ability to work with the DOM and SAX models, for interoperability.
- Can instantiate XML fragments and manipulate them, unlike the DOM model where to instantiate an XML-fragment, you must create a containing document first.
- XML creation, using functional construction, looks very similar in code to the XML it is creating, unlike the DOM that builds an XML tree from the bottom up.
- Provides querying of XML structures, based on XPath, with two choices of syntax; one being very similar in structure, and therefore familiar to, Structured Query Language (SQL), long used to query relational databases.

With choices of XML for data modelling, and C# for programming logic and querying of data, the choice of development tool was straight forward. The choice being Microsoft Visual Studio 2010. This provides an Integrated Development Environment (IDE) with debugging, modelling, and code support tools.

5.6 Prototype - Data Structures

The following section is divided into two subsections. The first subsection identifies key components and strategies used to define the Quality Criteria Model (QCM) in XML. For a full specification of the QCM, as an XML-schema file (`QCSchema.xsd`), please refer to Appendix C.1. The second subsection identifies key portions of the CCR

health information model used. The emphasis is on CCR components used for detection, realisation, and processing of Quality Criteria (QC). Examples are given of how QC could be detected within Electronic Health Record content that has been based on the CCR model.

5.6.1 Prototype - QCM Definition

The QCM requires formal definition of data types and cardinality constraints before it can be incorporated into the prototype. An XML-schema was developed for this purpose, the QCSchema. With XML-schema it is possible, and sometimes desirable, to incorporate, or re-use XML types and elements from previously existing schema. For QCSchema development, the base schema from the XML standard (XMLSchema) and the CCR schema were imported to allow use of their types.

The imports are declared at the beginning of the QCSchema as attributes of the `xs:schema` element. The `xs:schema` element defines this as a schema definition and is the opening and closing element of any XML schema. To aid in using imported types, namespaces are declared, where the imported types are given a prefix to distinguish them, via the `xmlns` keyword, followed by the desired prefix e.g. `xmlns:ccr` for CCR types. The `import` and `namespace` declarations can be seen in Listing 5.1. The `targetNamespace` attribute for the schema element declares the namespace that this XML schema is intended to target, or validate i.e. what namespace this schema file describes. The `elementFormDefault="qualified"` states that elements and attributes declared in this schema are in the `targetNamespace`.

The three explicit namespace declarations follow, using the `xmlns` keyword. The first is `xmlns:"urn:qc-models:QC"` for the `targetNamespace`, and has no prefix. Any elements or types declared in the QCSchema will not have a prefix. The following two `xmlns` statements declare prefixes (`ccr` and `xs`) for elements, or types, used in the QCSchema that are from the CCR and XMLSchema namespaces respectively. These imported types or elements will have prefixes. The last line in Listing 5.1 is an import statement and tells QCSchema where to find the elements and types from the CCR schema. No import statement is required for the XMLSchema namespace as it is the XML schema standard, and is implicit. This is an introduction to XML schema functionality. It is given here to highlight the ability of XML schema to be built from re-usable types from other schema, and to be extended by the addition of specialised types, for specific use cases. For further detail, please refer to the XML standard developers website at www.w3.org/standards/xml.

Listing 5.1: Type imports and namespace declarations.

```
<xs:schema  targetNamespace="urn:qc-models:QC"
   elementFormDefault="qualified"
   xmlns="urn:qc-models:QC"
   xmlns:ccr="urn:astm-org:CCR"
   xmlns:xs="http://www.w3.org/2001/XMLSchema">
   <xs:import namespace="urn:astm-org:CCR" schemaLocation="CCR.xsd"/>
   .
</xs:schema>
```

An example of the functionality clarifies use. Figure 5.2 shows the first part of the QCM. The first four elements: Label, ID, Date Created, and Evaluation Type, are all created in the QCSchema by using three different types.

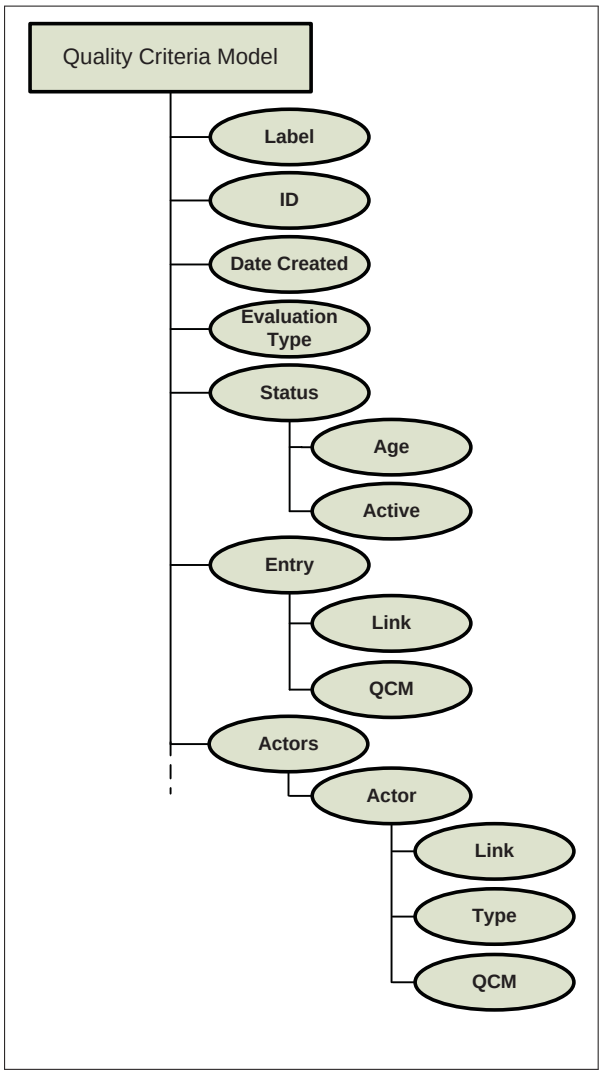


Figure 5.2: Quality Criteria Model - Part 1.

The different types are seen in Listing 5.2, with the declaration of the first four elements of the QCM, after the schema declarations. In the example, definition of the QCM begins after the comment line `<!-- Quality Criteria Model. -->`. The element `quality-criteria` is declared. It is declared as a complex type (`complexType`) with a sequence of elements to follow. The `sequence` element explicitly defines that the following elements must be in the specific order from the schema for any XML instance validated against the QCSchema. The declaration as a `xs:complexType` declares the entire QCM definition as a data type, and able to be referenced from other elements, and schemas. The sequence of elements begins with `label`, declared as an XMLSchema type string (`xs:string`) from the XMLSchema namespace. The `qc-id` element is declared the same. The third element, `date`, is declared as a CCR data type (`ccr:DateTimeType`) from the CCR namespace. The last element in the QCSchema fragment is `evaluation-type`, and is declared as `type="eval-type"`. The eval-type type has no prefix, as it is a type declared within the QCSchema, as a global type. Cardinality constraints for `evaluation-type` are truncated to ... for space.

Listing 5.2: Three types from different namespaces.

```
<xs:schema  targetNamespace="urn:qc-models:QC"
   elementFormDefault="qualified"
   xmlns="urn:qc-models:QC"
   xmlns:ccr="urn:astm-org:CCR"
   xmlns:xs="http://www.w3.org/2001/XMLSchema">
   <xs:import namespace="urn:astm-org:CCR" schemaLocation="CCR.xsd"/>

<!-- Quality Criteria Model. -->

 <xs:element name="quality-criteria">
  <xs:complexType>
   <xs:sequence>
    <xs:element name="label" type="xs:string" minOccurs="1" maxOccurs="1"/>
    <xs:element name="qc-id" type="xs:string" minOccurs="1" maxOccurs="1"/>
    <xs:element name="date" type="ccr:DateTimeType" minOccurs="1" maxOccurs="1"/>
    <xs:element name="evaluation-type" type="eval-type" ... />
    .
   </xs:sequence>
  </xs:complexType>
 </xs:element>
</xs:schema>
```

Global types for the QCSchema are not declared inside the `quality-criteria` element, but as type declarations at the same level as the quality-criteria element. Declaration at this level allows for use of global types in other structures. This is seen in Listing 5.3 where `eval-type` is declared as a simple global type, restricted to the `xs:string` type, and also enumerated i.e. is constrained, and can only be one of two declared (enumerated) values, objective or subjective.

Listing 5.3: Simple global type declaration for `eval-type`.

```
<xs:simpleType name="eval-type">
  <xs:restriction base="xs:string">
    <xs:enumeration value="objective"/>
    <xs:enumeration value="subjective"/>
  </xs:restriction>
</xs:simpleType>
```

The recursive nature of the QCM is made possible by declaring the model's definition as a global element, in the form of the element declared as a complex type named `quality-criteria`. The following example illustrates this. The QCSchema fragment in Listing 5.4 shows the structure of the element `actors`, that contains a listing of all the `actor` elements, that contain realised Quality Criteria (QC) for actors of the entry. The `actor` element definition has had the cardinality constraints truncated for space reasons. The last element declaration in the inner sequence contains the following: `ref="quality-criteria"`, followed by cardinality constraints. The `ref` keyword instructs the schema to refer to the element that is its value. The `ref` keyword is used to reference a globally declared element, in this case the global element `quality-criteria`, just as the `type` keyword is used to reference a globally declared type. Referencing the quality-criteria element here, means the entire QCM model is available at this point, to record a QC instance for the `actor`. Figure 5.3 shows a graphical representation of the recursion.

Listing 5.4: QCM global type `actor-type`.

```
<xs:complexType name="actor-type">
  <xs:sequence>
    <xs:element name="link" type="ccr:ActorReferenceType"/>
    <xs:element name="type" type="xs:string" minOccurs="1" maxOccurs="1"/>
    <xs:element ref="quality-criteria" ... />
  </xs:sequence>
</xs:complexType>
```

The QCM `actor` element is of type `actor-type` (Listing 5.4, declared as global type

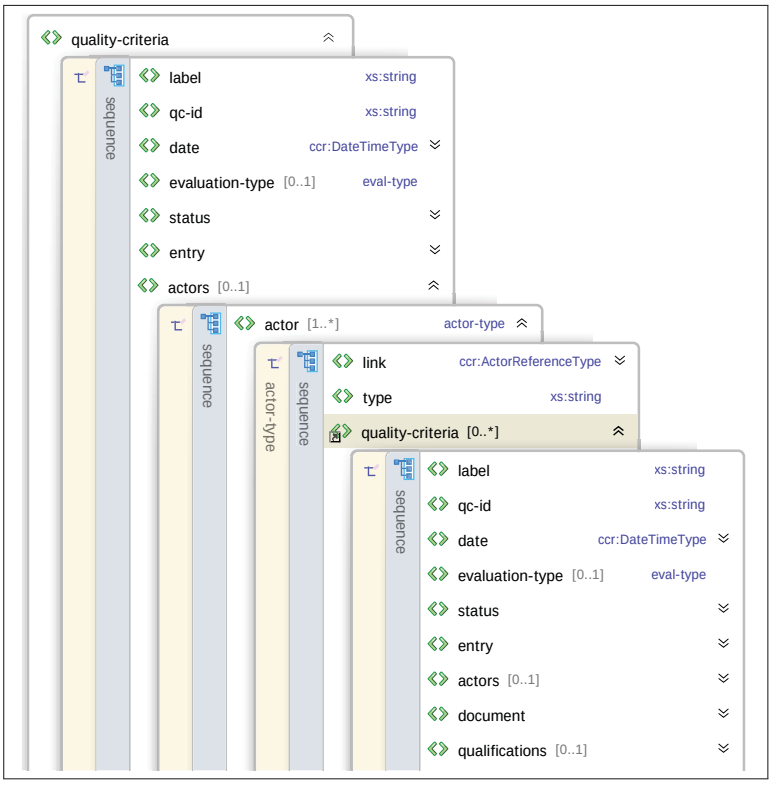


Figure 5.3: Recursion of QCM.

within the QCM. `actor-type` is extended using the CCR type `ActorReferenceType` that contains two elements; an `ActorID` and an `ActorRole`, for referencing and describing the actor involved. The `actor` element is extended with a `type` element (`xs:string` type), and a new instance of the QCM that would begin with the element `quality-criteria`, where all realised QC for this actor would be recorded, using the QCM model described by this globally declared element. The `actor` element would then contain the elements: `ActorID`, `ActorRole`, `type`, and a recursion of the QCM beginning with the `quality-criteria` element. Whether all elements are present, or how many of each, is dependent on the cardinality constraints declared for the elements.

Elements that record data values for QC e.g. elements `role`, `narrative`, `qualification`, or `attestation`, do so via a globally declared complex type, `qc-data`. The `qc-data` type is shown in Listing 5.5. It is a complex type as it contains other types, not only strings. The `xs:choice` element defines that only one of the defined elements in the sequence is to be used, per instance of the type.

Listing 5.5: Global complex type `"qc-data"`.

```
<xs:complexType name="qc-data">
 <xs:choice>
```

```
  <xs:sequence>
   <xs:element ref="ccr:InternalCCRLink"/>
   <xs:element name="description" type="xs:string"/>
   <xs:element name="true" type="xs:boolean"/>
  </xs:sequence>
 </xs:choice>
</xs:complexType>
```

The first choice element is a reference that points to where the data is found e.g. for QC C8: Contextual Narrative, the reference would point to where the narrative could be found. The second choice (description) would contain the contextual narrative itself, as a string. The third choice (true) is a boolean type and would indicate that yes, a contextual narrative for this entry exists. Conversely, it could indicate, no, a contextual narrative does not exist. Three choices are available, as it is unknown at the time of processing, in what way the data will have been stored. It also may be desirable not to use the `xs:choice` constraint so that combinations of values could be used e.g. a true value plus a reference to where the narrative exists.

5.6.2 Prototype - CCR Components Used For Detection Of QC

As with the QCM, the Continuity of Care Record's (CCR) elements, data types, structure, and cardinalities are defined in an XML schema. The schema (`CCR.xsd`) uses a combination of globally defined types and elements, as well as types defined in the XML standard schema. The globally defined `ContinuityOfCareRecord` element forms the basis for all CCR content. The top level elements of the CCR are seen in Figure 5.4.

Within the `ContinuityOfCareRecord` element, the `Body` element contains all clinical content related to the patient the record is for. The clinical content sections are seen in Figure 5.5. As described in Chapter 3, Section 3.5.4, the three areas focused on for the interview process were: adverse events, medical history, and medicines list. The three areas map to CCR content sections as follows:

1. **Adverse Events** to **Alerts**.
2. **Medical History** to **Encounters**.
3. **Medicines List** to **Medications**

The medical history domain could be mapped to more than one section e.g. could include Immunizations section. However, for simplicities sake in development, a one to

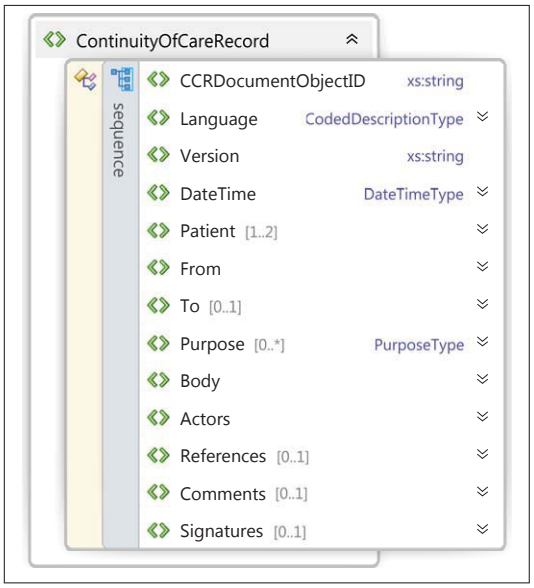


Figure 5.4: CCR top level elements.

one mapping was decided on. Each CCR section contains individual Alert, Encounter, and Medication entries, as shown in Figure 5.6. The figure shows individual entries are defined by types. An `AlertType`, `EncounterType`, and `StructuredProductType` respectively. The sections have cardinality of 0..1, meaning that a section does not have to be present but should it be, the section element only occurs once. The individual entries' have cardinality 1..*, meaning that should a section be present, it must contain at least one entry, but may contain many. Of interest, unlike when CCR types are used in the QCM schema, when used within the CCR schema it is unecessary to prefix the types with `ccr`, as the types are within the target namespace.

The global complex data type `AlertType` illustrates how CCR builds types from others, and also extends types to provide specialisation and specific meaning for instances. An `Alert` entry is of `AlertType`. `AlertType` is an extension of the `CCRCodedDataObjectType` complex data type. The `AlertType` is constructed from the `CCRCodedDataObjectType` plus two referenced global elements: the `Agent` and `Reaction` elements. The extension elements, `Agent` and `Reaction`, record the agent that caused the adverse reaction, and the reaction that resulted (ASTM, 2005), thus specialising the `CCRCodedDataObjectType`. The `CCRCodedDataObjectType` is used in many entries within CCR, and illustrates a core re-usable type that forms the basis for specialisation for specific requirement. The same paradigm is used with Object-Oriented types, described in Chapter 4, Section 4.2. The `AlertType` structure is illustrated in Figure 5.7, along with detailing how the `CCRCodedDataObjectType` is itself built from other global types e.g. `DateTimeType`.

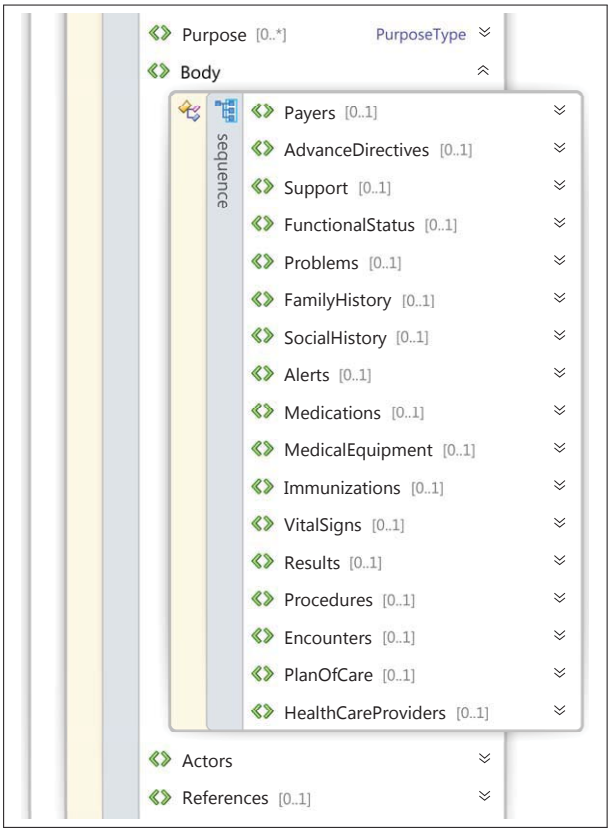


Figure 5.5: CCR clinical content sections.

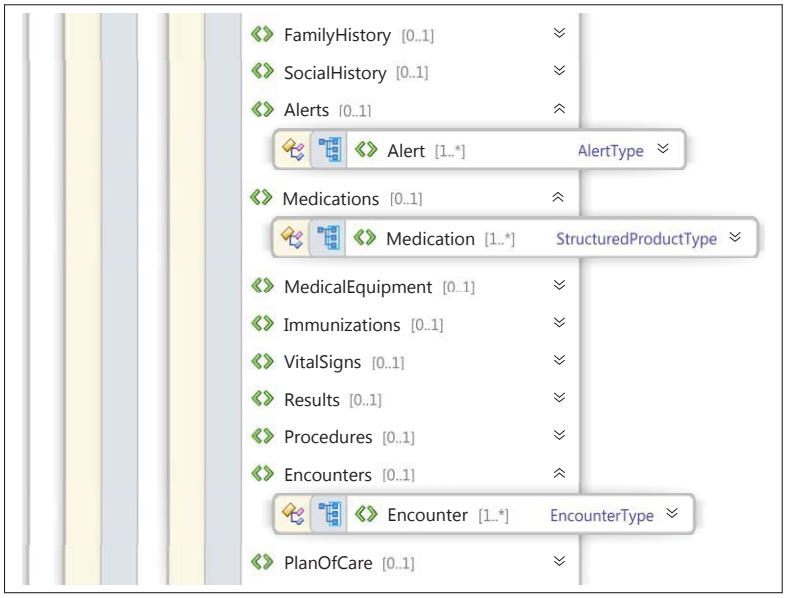


Figure 5.6: CCR individual clinical content entries for sections.

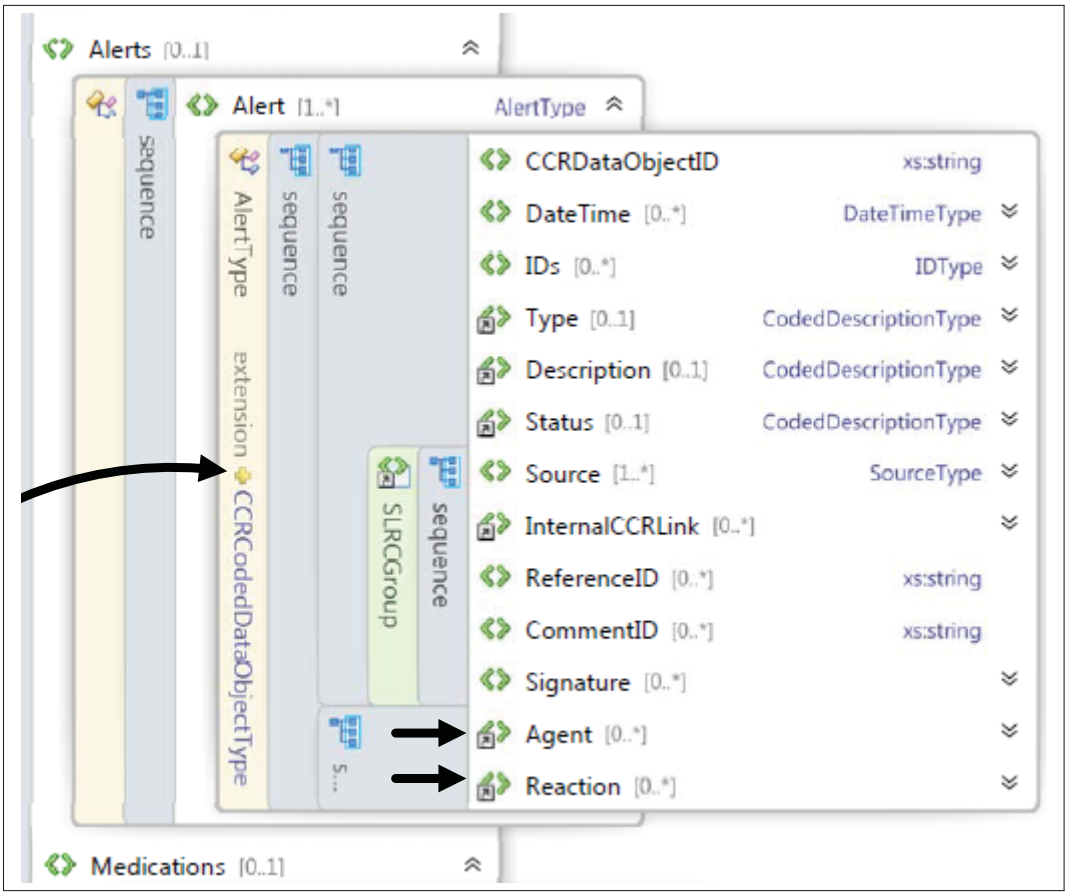


Figure 5.7: CCR `AlertType` showing core type plus extensions.

The type re-usability, plus specialisation, provides several benefits, as when used with OO programming, such as:

- Able to define core data elements/types to be used multiple times as the basis for multiple elements.
- Core data elements/types can be specialised by additional elements/types to represent specialised data, thus easing the creation of complex data items.
- Makes XML schema easier to use and understand, as based on well understood, and validated, principles from other domains e.g. OO programming.
- Core (generalised) types are re-used as the core for multiple specialised types, developing systems for processing data, constructed using such techniques, is easier, as systems ultimately need to deal with a less elements, as they are based on re-usable, generalised types.

Table 5.1: Quality Criteria potential detection parameters.

QC - Short Name	Detection Detail From Entry/Source
C4 : Attestation	• is entry attested to? • is entry source attestor?
C7 : Attestor role	• if entry is attested to, is role defined for source?
C8 : Contextual narrative	• is a reason for an entry given? • is Context provided for an entry?
C9 : Further information	• are references/pointers to more information provided?
C10 : Clinician qualifications	• if details for source are provided, are qualifications given? • can qualifications for the data source be derived?
C11 : Clinician experience	• are details for years of service of source provided? • can years of service be derived from source details?
C12 : Data age	• if date for an entry is given; calculate the age of the data?
C13 : Change narrative	• is a reason for a change to an entry given?
C14 : Information captured	• do details for an entry state where the entry was recorded? • can where an entry was recorded be derived from entry/source details?
C15 : Clinician role	• are details of the role for the source provided? • can the role of the source be derived?

Before introducing specific CCR components, and how the specialisation of types can assist with Quality Criteria (QC) detection, a listing of QC used in development is given. The focus is why they were chosen for development, and their potential means of detection.

Not all QC were chosen for implementation. Those QC classified as being of a subjective evaluation type (see Chapter 3, Section 3.6.2) were excluded in the first instance. This was because, to evaluate subjective criteria e.g. C3: Organisational reputation, requires data for evaluation that is not currently available with health information. Nothing available in the CCR specification would support this type of evaluation. This is not to say that future additions to health information communications could introduce the required data. Alternatively, the system surrounding the health information could factor in this data sourced from elsewhere e.g. the proposal in New Zealand to produce league tables for organisational performance. Two other QC were also excluded at this stage. This was C1: Contradictory entries and C5: Consistency/cohesion. These were excluded as their focus is on multiple entries, and/or the Electronic Health Record (EHR) document as a whole (refer Chapter 4, Section 4.4.2 for detail). Therefore, the QC looked at were objectively evaluated QC that related to single entries, or single entry sources. These are listed in Table 5.1, along with potential means for detecting the QC.

Two CCR data structures are used to implement detection of QC in the following

example. The two structures are the globally declared, re-suable types: `CCRCodedDataObjectType` and `CodedDescriptionType`. The `CCRCodedDataObjectType` has been discussed previously and forms the basis of all of the CCR's clinical data entries e.g. Alert, Medication, and Result. The `CodedDescriptionType` is a smaller data structure and is a component re-used in many elements, and types, within the CCR, including the `CCRCodedDataObjectType`. The `CCRCodedDataObjectType` is shown in Figure 5.8 and shows the use of `CodedDescriptionType` in its construction.

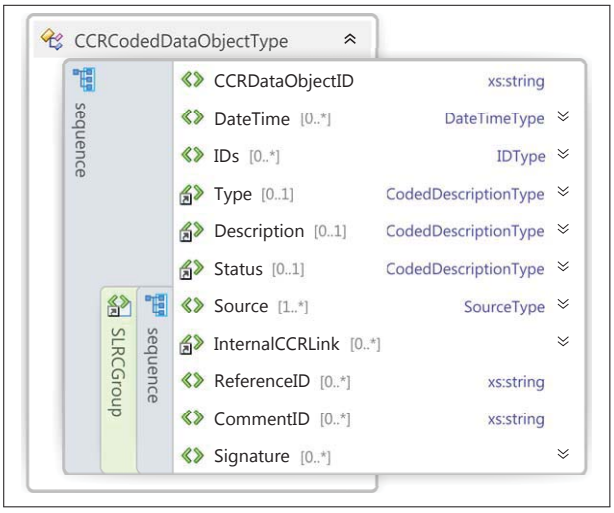


Figure 5.8: `CCRCodedDataObjectType` type contents.

Detection of the first QC listed in Table 5.1, C4: Attestation, is used to illustrate the use of the two CCR components. C4: Attestation can be detected in several ways, depending on individual usage of the CCR. In this example, two ways are illustrated. The example uses an Alert entry, that is an extension of the core type, `CCRCodedDataObjectType`.

The first method utilises the Source element defined in `CCRCodedDataObjectType`. For all clinical entries present in a CCR document, the `Source` element is required i.e. if an entry is present, it must contain a source. As shown in Figure 5.9 the source element can point to an actor via the `Actor` element, within source, of type `ActorReferenceType`. The `ActorReferenceType` is composed of two elements; `ActorID` that points to the actor entry in the CCR Footer section, and `ActorRole`, that is composed of the `CodedDescriptionType` global type.

The CCR specification states that a `CodedDescriptionType` supports the use of either simple text strings (unstructured data), through to complete, structured, discrete data. The intention is to support a variety of systems, to support use of CCR e.g. a system that can only generate a text string (ASTM, 2005). The desired flexibility is seen with the three elements that make up the type: `Text`, `ObjectAttribute`, and

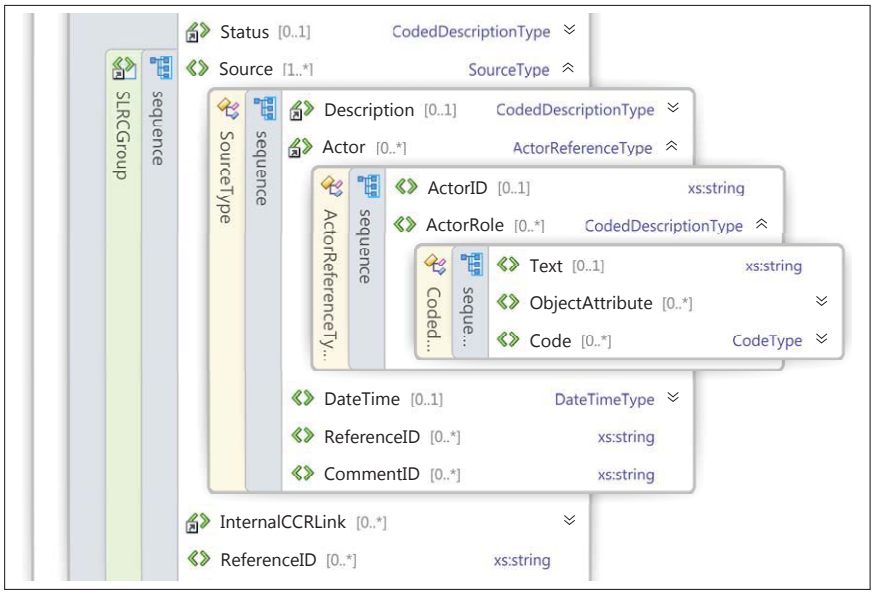


Figure 5.9: Source element within `CCRCodedDataObjectType`.

`Code` seen in Figure 5.9. `Text` is for an unstructured string (limited to one instance), `ObjectAttribute`, and/or `Code` for structured data. `ActoreRole` can therefore be defined in many ways, depending on individual implementations for the CCR. This ranges from a free text string description, to a coded entry from a structured clinical vocabulary e.g. SNOMED CT, or ICD-9. For the example being given, the free text string is used, where the `ActorRole` text element content is "Attestor". Listing 5.6 illustrates this. The `ActorRole` element, used in this way, allows for detection of attestation. A system e.g. the prototype implemented, would look for a source, where the `ActorRole` value was "Attestor". As multiple sources for an entry is permitted, the same actor could be referenced, but with different roles e.g. attending physician.

Listing 5.6: `ActorRole` with text value = Attestor.

```
<ccr:Source>
 <ccr:Actor>
  <ccr:ActorID>A001</ccr:ActorID>
   <ccr:ActorRole>
   <ccr:Text>Attestor</ccr:Text>
   .
   </ccr:ActorRole>
</ccr:source>
```

To demonstrate flexibility in CCR implementation a second method, for the example of detecting attestation, uses the `CommentID` element of `CCRCodedDataObjectType`.

`CommentID` references (links to) a comment instance (of complex data type `CommentType`), within the `Comments` section of the CCR Footer. An instance of `Comment` provides a comment to a data object e.g. entry, but should not contain core relevant clinical or administrative data (ASTM, 2005). The `CommentType` that defines a comment is seen in Figure 5.10.

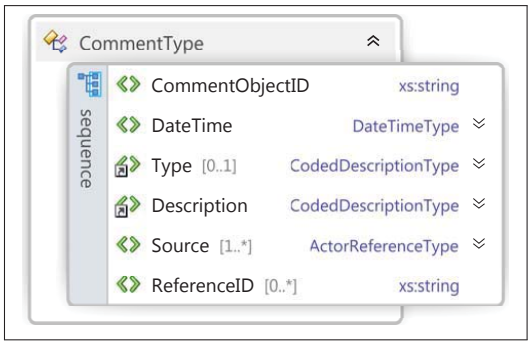


Figure 5.10: CCR `CommentType` describes a `Comment` instance.

Detection of attestation of an entry would occur using the `Type` element within an instance of a comment, referenced from the entry. The `Type` element, of `Comment`, is an instance of `CodedDescriptionType`. The `Type` element would use the `CodedDescriptionType`'s `ObjectAttribute`, element to indicate attestation. `ObjectAttribute` can contain `Attribute`, `AttributeValue`, and `Code` elements. Figure 5.11 illustrates this structure within a `Comment`. To indicate attestation, `Attribute` and `AttributeValue` elements are used. `Attribute` is a string, with value "Attestation". `AttributeValue` would have its `Value` element equal to "Yes", or "True". It could also have a value of "No", or "False", if a negative indication of attestation were required. An instance of this is seen in Listing 5.7 Detection would occur by following `CommentID` references from an entry, and checking for comments whose `Type` element has an `ObjectAttribute` element whose `Attribute` element has a value of Attestation.

As `Type` and `Description` elements within `Comment` are of `CodedDescriptionType`, either could be used to capture attestation. The example outlines two CCR data structures that would allow for capture and detection of attestation for an entry. Both structures, `ActorRole` and `Comment` provide for capturing the source of the attestation, which is required for attestation of clinical data.

Listing 5.7: Comment used to capture attestation.

```
<ccr:Comment>
 <ccr:CommentObjectID>C001</ccr:CommentObjectID>
 <ccr:DateTime></ccr:DateTime>
 <ccr:Type>
```

```
  <ccr:ObjectAttribute>
   <ccr:Attribute>Attestation</ccr:Attribute>
   <ccr:AttributeValue>
    <ccr:Value>True</ccr:Value>
   </ccr:AttributeValue>
  </ccr:ObjectAttribute>
 </ccr:Type>
 .
</ccr:Comment>
```

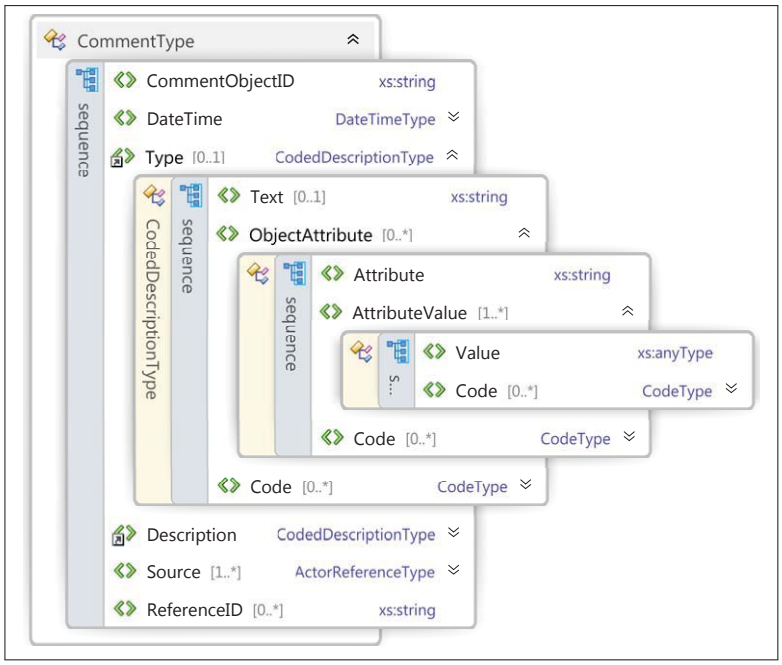


Figure 5.11: CCR `CommentType` structure.

For data structures that support attestation, several health information models, besides the CCR, provide for capturing attestation of clinical data. This was why attestation was included as a candidate quality criteria from the literature in Chapter 2. Examples of this include HL7's new Fast Healthcare Interoperable Resources (FHIR). The FHIR standard provides document attestation codes for attestation of FHIR documents (Health Level 7 International, 2013a). The attestation is at the document level, rather than at an entry instance level. openEHR also supports attestation via its conformance to the ISO 18308 standard (Beale, 2006). Having explicit data structures would further facilitate detection of the QC, C4: Attestation.

The example given for the QC, C4: Attestation, was for a QC related to an entry. An example related to an entry's source, highlights the CCR data structures that allow

detection of source related QC. A source is an actor within the CCR. An actor is either a person, organisation, or information system. An actor is defined in the CCR by the complex global type `ActorType`. Elements for `ActorType` are seen in Figure 5.12. The `ActorObjectID` is the ID referenced from elsewhere in the CCR document by the `ActorReferenceType`, described in the previous example. Within `ActorType` is a `Source` element to reference the source of an actor instance.

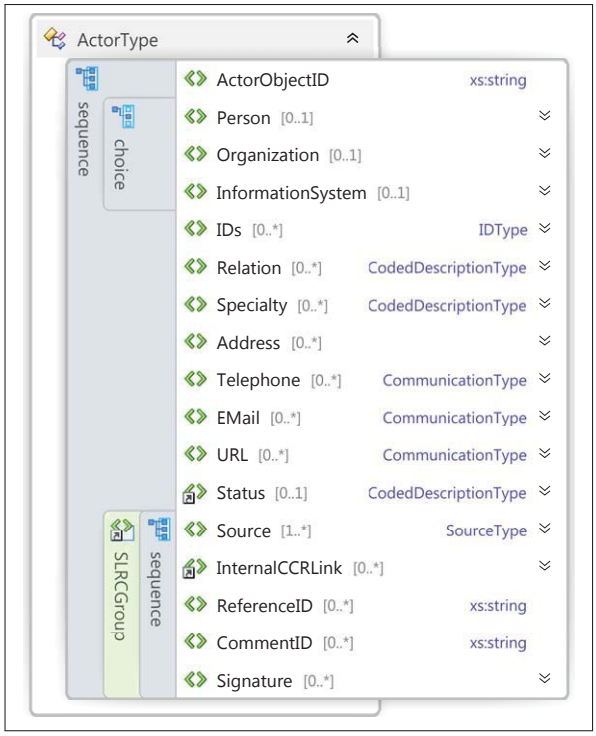


Figure 5.12: CCR `ActorType` structure.

The example illustrates the data structures used for detecting C10: Clinician qualifications. For the instance of `Actor` in the example, the choice from the three options immediately after `ActorObjectID` would be Person. Within `ActorType` the element `Specialty` can contain multiple instances (cardinality 0..*) of `CodedDescriptionType`, as seen in Figure 5.13. As there can be multiple instances of `CodedDescriptionType` within the `Specialty` element, each instance would contain a qualification/specialty. This could be an unstructured text string, structured `ObjectAtrribute`, or a `Code` representing the qualification/specialty. A specialty being a field of focus for the health practitioner. A combination of these is also permitted, however each instance of the `Specialty` element would only record a single specialty, or qualification. Codes for specialties would come from a clinical terminology vocabulary such as SNOMED CT. Listing 5.8 shows a specialty element populated using all options i.e. unstructured text string, `ObjectAttribute`, and `Code`.

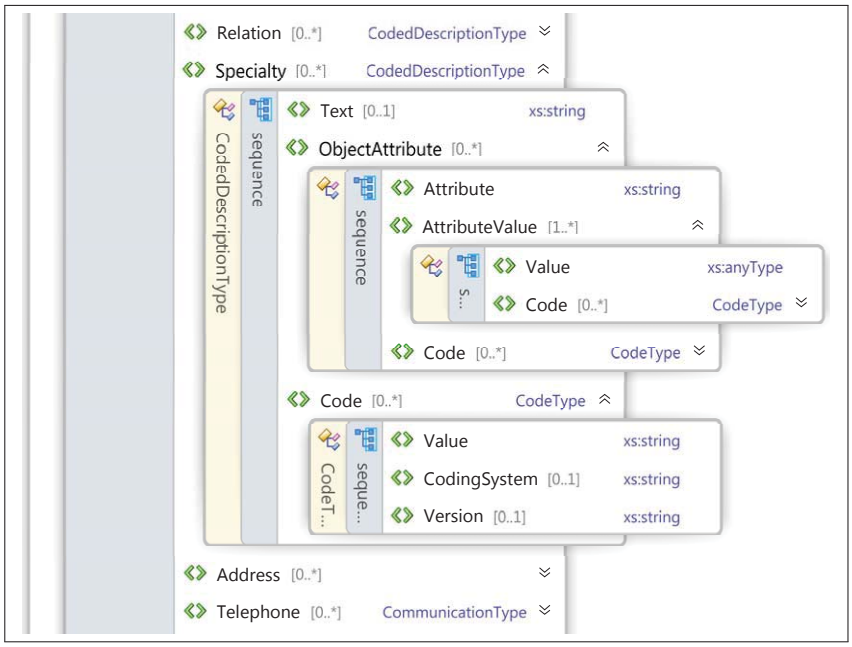


Figure 5.13: CCR `ActorType`: `Specialty` element structure.

Multiple `ObjectAttribute` elements are permitted for a `Specialty` element allowing for recording of different attributes for a specialty. For example, the specialty example shown in Listing 5.8 could contain a second attribute Experience, with a value indicating the number of years experience with the specialty. Thus capturing data for C11: Clinician experience. The re-usable type structure used with the CCR specification is highlighted with the examples given here. This is further demonstrated with the `ActorRole` element of `ActorReferenceType`, used previously to demonstrate that an actor referenced, was in the role of attesting to an entry. The `ActorRole` element of `ActorReferenceType` can have multiple instances (cardinality 0..*). Therefore, for each actor referenced, multiple roles could be specified. Returning to C4: Attestation example, attestation was capture by having a role as Attestor. The QC, C7: Attestor role (the role the person was in when attesting to the information), could be captured by a second instance of `ActorRole`.

The last data structures emphasised, are the `ReferenceType`, and `CommentType`, previously used in the Attestation example. The CCR standard defines a clear cut difference between References (`ReferenceType`), and Comments (`CommentType`). References are sources/locations outside the CCR document. These can be URL's, articles, patient documents (paper or electronic), images, or other materials of assistance to the CCR user. Comments are any data associated with data within a CCR document. Both are discrete instances contained within their respective sections in the CCR Footer. Both

have `Source` elements, with a `Reference` element referencing external data or location, and a `Comment` element being able to refer to a `Reference`, via `ReferenceID`. These structures would capture the following QC: C8: Contextual narrative, C9: Further information, C13 Change narrative, and C14: Information captured. References and Comments highlight the pointer/reference nature of the structure , and implementation, of a CCR document.

Listing 5.8: Populated `Specialty` element for `ActorType`.

```
<ccr:Specialty>
 <ccr:Text>Clinical oncology</ccr:Text>
 <ccr:ObjectAttribute>
  <ccr:Attribute>Clinical speciality</ccr:Attribute>
  <ccr:AttributeValue>
   <ccr:Value>Clinical oncology</ccr:Value>
   <ccr:Code>
    <ccr:Value>394592004</ccr:Value>
    <ccr:CodingSystem>SNOMED CT</ccr:CodingSystem>
    <ccr:Version>1201 Intl</ccr:Version>
   </ccr:Code>
  </ccr:AttributeValue>
  <ccr:Code>
   <ccr:Value>394658006</ccr:Value>
   <ccr:CodingSystem>SNOMED CT</ccr:CodingSystem>
   <ccr:Version>1201 Intl</ccr:Version>
  </ccr:Code>
 </ccr:ObjectAttribute>
 <ccr:Code>
  <ccr:Value>394592004</ccr:Value>
  <ccr:CodingSystem>SNOMED CT</ccr:CodingSystem>
  <ccr:Version>1201 Intl</ccr:Version>
 </ccr:Code>
</ccr:Specialty>
```

With the ubiquitous use of `CodedDescriptionType` throughout CCR data structures, it would be possible to capture the QC listed for References and Comments, within entries themselves e.g. as an attribute of the `Description` element within an Alert entry. However, this approach does not support the CCR standard's emphasis on data normalisation, where data is recorded once, and referenced, should it be needed again. The same approach is used to reference instances of QC, represented by the Quality

Criteria Model (QCM), from entries, and sources, that realise QC.

5.7 Prototype - Function and Operation

The purpose of the prototype was to address Research Questions 4 and 5 (see Chapter 1, Section 1.2.1). The prototype provides three areas of functionality to achieve this. These are:

1. Discovery of Quality Criteria (QC) present within a CCR specified document, or document fragment.
2. Capture of QC data, via creation of instances of the Quality Criteria Model (QCM).
3. Tracking of references between QC instances and the data sources for those instances via indexing.

The prototype function is conceptualised as a black-box system in Figure 5.14. When viewed as a black-box the prototype functions by taking, as input, a CCR document, and outputting the CCR document with an index of references for any QC realised, along with the captured QC information. The functionality is fulfilled via a search algorithm, a formally specified model (QCM), and an index data structure. The search algorithm is a collection of individual search functions tailored to discover individual QC. For example, one search function to discover C4, and another to discover C7, and so forth, for the list of QC being searched for. Importantly, the prototype does not alter the clinical content of the original document.

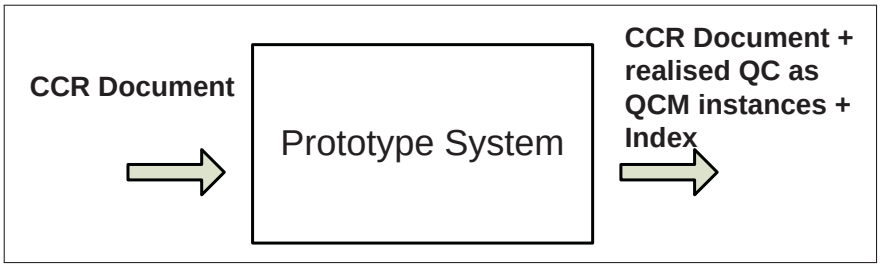


Figure 5.14: Prototype functionality as black-box system.

5.7.1 Prototype - Example Operation

Prototype operation begins with the reading in of the health information document to be examined. The document is validated by the system, against the CCR schema,

to ensure the health information is a valid instance of the CCR. The search operation begins by creating an index of the search functions to be used. The index is initialised as a one dimensional array, with each position in the array corresponding to a search function. For the example operation the first position is the search function looking for QC C4:Attestation. The search function for this criterion uses the mechanism described in Listing 5.6 where the `ActorRole` element of the `ActorReferenceType` for the entry's source has a value "Attestor". Figure 5.15 represents the search function index and contains the objectively evaluated QC from Table 5.1.

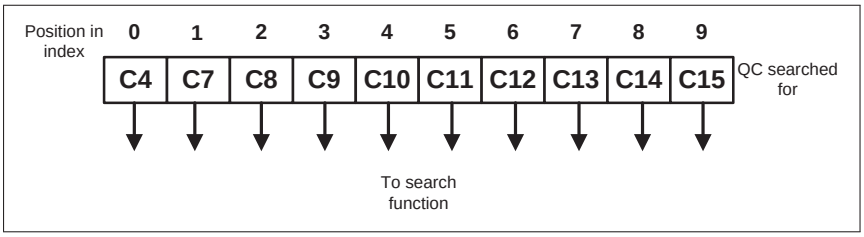


Figure 5.15: Index for search functions.

The entries from the CCR specified health information are now searched. The search begins with the first search function indicated by the search index. In the example, the search for C4 is in first position of the index and therefore, the function for discovering the presence of C4 in entries is launched. Each entry is visited in turn by the C4 search function. Figure 5.16 conceptually shows the process. The figure shows four entries, with QC associated with those entries that can potentially be discovered, for illustrating the example.

Entry 1 from Figure 5.16 is searched, but does not have C4 Attestation to be found. The search moves to look for C4 at Entry 2. At Entry 2, C4 is found. At this point the search algorithm changes to a different mode. The search now runs all search functions on Entry 2, from those listed in the search index (see Figure 5.15) i.e. the algorithm now runs the search for C7, C8, C9, through to the last search in the index, C15. Also, at the point C4 is discovered for Entry 2, an instance of the QCM is generated, and C4 is recorded by the instance for Entry 2. As the examination of Entry 2 proceeds, further discovered QC are added to the instance of the QCM. For Entry 2 this would mean that C4, C8, C11, and C12 are recorded by the QCM. The QCM is validated against the QCSchema to ensure it is a valid instance of the model. An index is generated that links the QCM to its source entry (Entry 2) via the unique IDs available for both entry and QCM. The index is created as a two dimensional array. As the QCM recursively contains all QC for Entry 2, only one ID for the containing QCM is required to reference Entry 2. Once all searches from the search index have been run for Entry 2, the search returns to looking for C4 in the next available entry, in this case Entry 3.

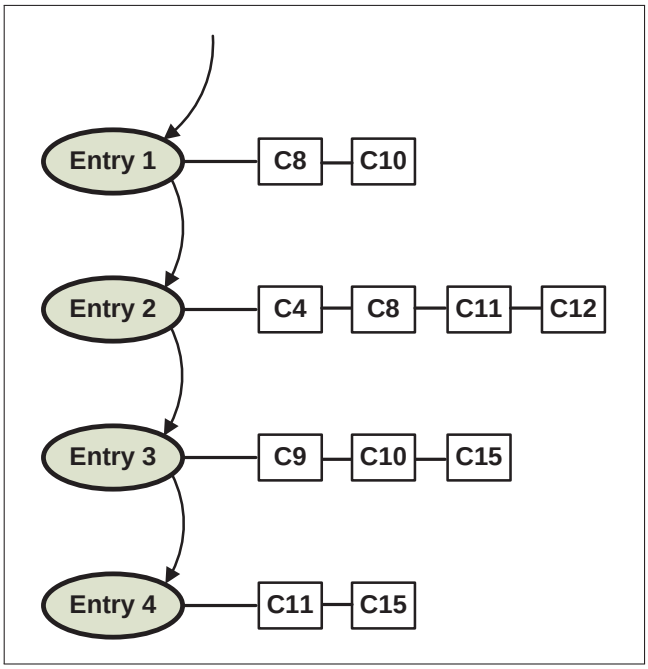


Figure 5.16: Search operation.

Once all entries have been examined for the presence of C4, the search moves to the next search function from the search index and begins searching entries again. The events from the previous search pass modify subsequent search passes. First, with the search for C4 exhausted, the C4 search function is removed from the search index. Also, the QC C7 Attestor role is dependent on C4 Attestation i.e. without an entry first being attested to (C4) there can be no finding for C7 Attestor role, the role the person was in when they attested to the information. Therefore, the search function for C7 is also removed from the search index. Second, Entry 2 triggered a search by all search functions on discovery of C4. Entry 2 does not need to be searched again and is excluded from further passes by search functions remaining in the search index. Figure 5.17 shows the state of the entries and the search index after the first pass of the search algorithm, ready for the second pass, beginning the search for C8.

On the second pass, C8 is discovered for Entry 1. All remaining search functions in the search index are executed on Entry 1. A QCM instance is created where all discovered QC for Entry 1 are recorded. Last, the created QCM is validated, and the index linking QCM instances to entries is updated. Search operations continue until the search index is empty, or until there are no more entries to examine.

At the completion of the search the prototype holds the original CCR document with entries, discovered QC for the entries captured in the form of instances of the QCM, and an index that links entries to their QCM. Figure 5.18 shows the state of the index

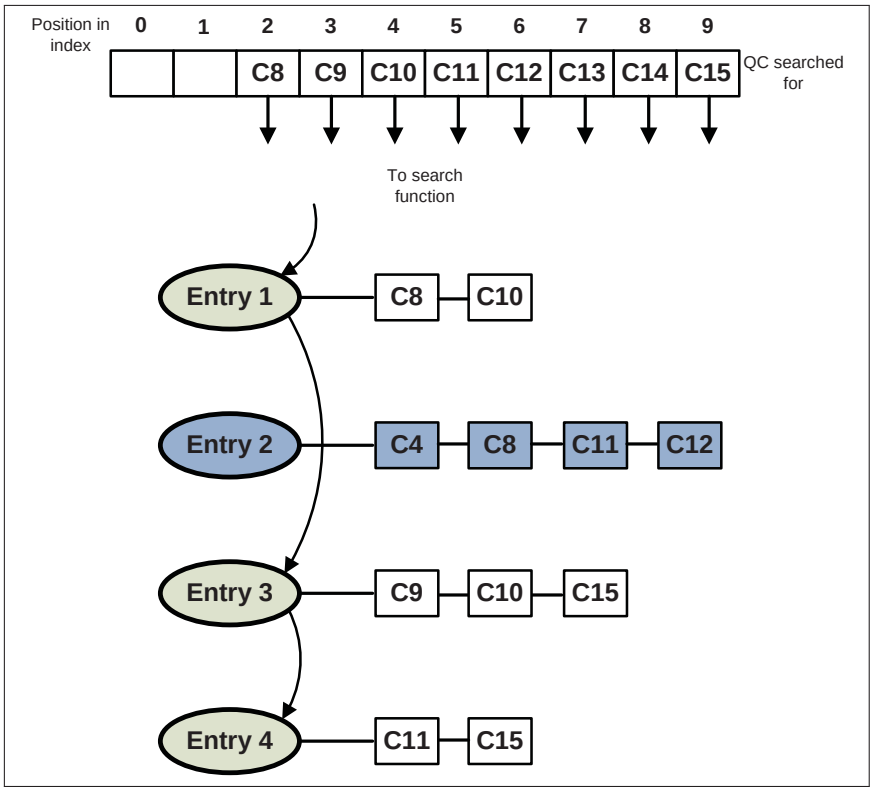


Figure 5.17: State of entries and search index after first pass of search algorithm.

referencing entries to QCM instances. As an example of a QCM instance, Figure 5.19 shows the QCM for Entry 2. The figure does not include all elements that are recorded e.g. figure does not include the `document` element that references the entry's containing document. Using the reference index, and the in-memory copies of the QCMs, the QCM instances are able to be written to the CCR document, as Comment entries within the Comments section of the CCR document. Once written to the CCR document the document is once again validated against the CCR schema to ensure it is still a valid instance of a CCR document.

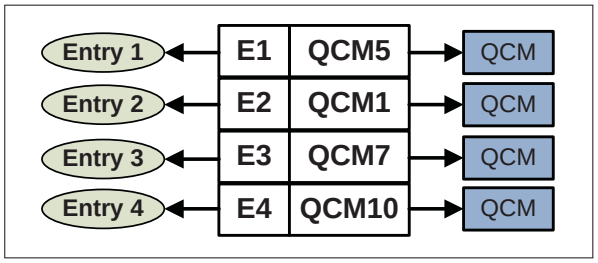


Figure 5.18: State of index linking entries to QCM instances.

The functionality of the index provides for storage of QCM instances outside the original document. When QC are recorded via the QCM, not only is a reference recorded

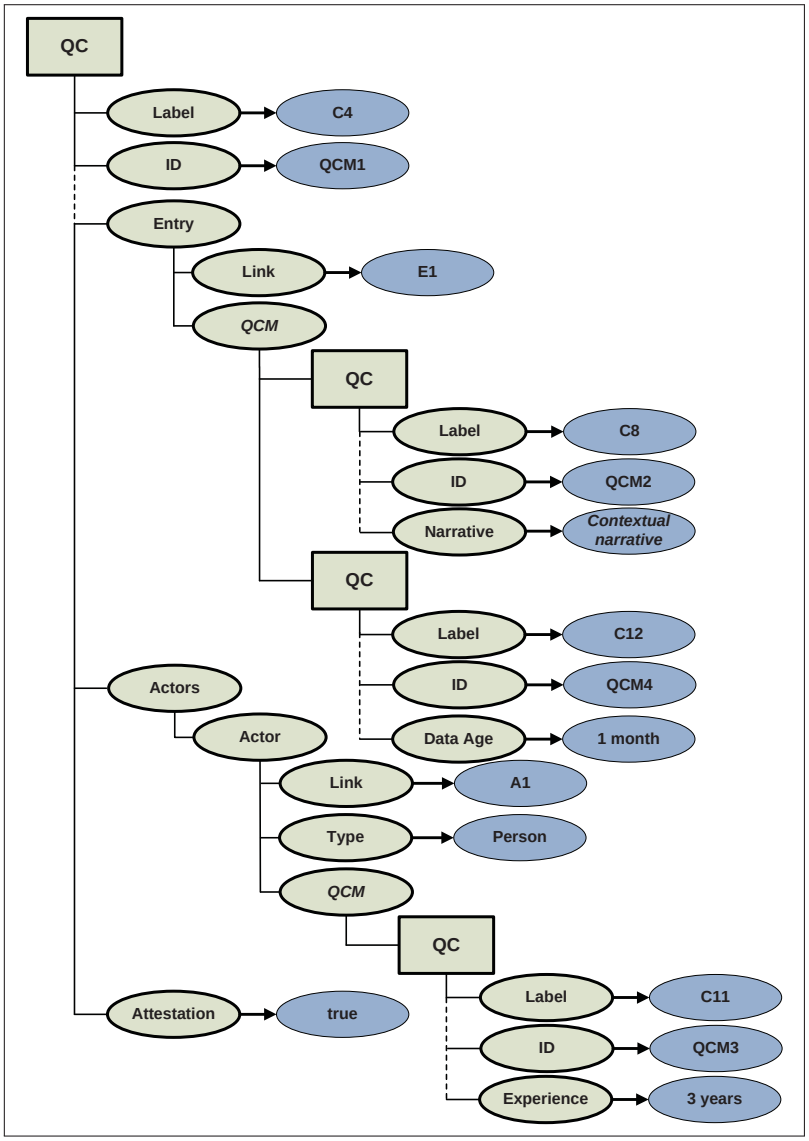


Figure 5.19: QCM instance after examination of Entry 2 from example.

to the originating entry, but also to the entry's containing document. Therefore QCM instances could be stored separately from the entries and linked by the referencing index. This is in line with several health information strategies, where clinical information for patients will be stored in repositories and aggregated at time of use e.g. the New Zealand Health Sector's proposed Health Information Exchange (HIE) based on Regional Clinical Data Repositories (R-CDR) for Cross Enterprise Document Sharing (XDS) services (NZ National Health IT Board, 2010; HISO, 2012a; NZ National Health IT Board, 2013).

The given example shows prototype operation. The pseudocode shown in Algorithm 1, conceptualises prototype execution operation and shows the steps for the search algorithm.

```
Data: Collection of health information entries
Result: HIQI; a collection of instantiated QC
initialization;
while entryIdx.count != 0 & functionIdx.count != 0 do
    run functIdx[current](entryIdx[current]);
    if functIdx[current](entryIdx[current]).return = true then
        generate QCM;
        update QCMidx;
        update entryIdx;
        udate functionIdx;
    else
        run functIdx[current](entryIdx[current]).next;
    end
end
```

Algorithm 1: Prototype program flow.

5.8 QCM Application With Other Health Information Models

Many health information models provide mechanisms that can be used to to provide enhancements to the core model. One such model is the new standard under development from HL7. The new standard is the Fast Healthcare Interoperability Resources (FHIR) specification. The FHIR specification is used here to illustrate how QC, represented via the QCM, could be used with FHIR. This is to show that the QCM is not limited to one type of health information model.

As stated by HL7, the philosophy for FHIR is a base set of resources to satisfy most

Table 5.2: FHIR extension definition fields (partial).

Field	Required?	Path in Profile (from Profile.extensionDefn)	Description
Code	Required	.code	The name that is used as a code in a resource to identify this extension - unique in the context of the defining profile
Context	Required	.contextType and .context	The context of this extension. See above. The context has two parts: a type, and a path which supplies the details
Short Definition	Required	.definition.short	A brief description of the extension used in the XML descriptions when the extension is referenced in a profile
Definition	Required	.definition.formal	A formal statement of the meaning of the content of the field
Requirements	Optional	.definition.requirements	Identifies the reason the extension is needed
Comments	Optional	.definition.comments	Additional other information about the extension, including information such as use notes
Type	Required	.definition.type	The type(s) of the extension. This SHALL be a valid FHIR data type as described above, or empty, if the extension will contain other extensions

common use cases for health information. However, it does provide an explicit, built-in extension mechanism, to capture remaining content as needed (Health Level 7 International, 2014). It is this extension mechanism that could be used to incorporate instances of the QCM into health information created using FHIR.

Extensions in FHIR are defined using a profile that describes the extension element. The fields used to define an extension are shown in Table 5.2. Not all fields available are shown in the table; only those relevant to this example. The fields are used to identify, describe, give reason for, and define a type for the extension.

The last field from Table 5.2, Type, specifies that the extension being defined must be of a valid FHIR data type. The data types are given for the FHIR specification and include a type labelled Attachment. The Attachment type can be used to contain, or reference, data content defined in other formats (Health Level 7 International, 2014). The Attachment data type could contain, or reference, instances of the QCM created for QC detected in FHIR resources used to represent health information.

5.9 Chapter Summary

Examples for prototype operation demonstrate that the meta-data in health information may be used to detect QC, with instances captured, using the QCM model. A search algorithm is used for this purpose. At the conclusion of the chapter, the possibility for use of the QCM with an alternative health information model was explored.

Chapter 6

Discussion

6.1 Reflections On Research Questions

The stated goal of this thesis was to provide a proof of concept for the overarching research question:

> What can a model based approach provide the health practitioner, when using electronic health information, within a clinical decision making context, such that the health practitioner is assisted in the judgement of the quality of the health information?

The question was broken down into five further questions. The five questions were tangible and able to be answered directly. This section reflects on these specific questions, and how they were addressed, and answered, through the thesis.

Research Question 1:

> What criteria do GPs use when evaluating quality of electronic health information?

The primary motivation for this research was to provide assistance in the judgement of health information quality by a GP at point of care. The rationale being that for safe use of health information, information must be of high quality i.e. fit for purpose. A GP is required to judge information fitness for purpose, in a clinical context. They seek to eliminate adverse impacts to patients that may result from basing decisions on the use, or not, of health information.

Users of information decide whether, or not, the information is "fit for purpose", with studies such as from Wang and Strong (1996), demonstrating the ability to conceptualise underlying aspects of data quality (Wang and Strong, 1996). The seminal work by Wang and Strong (1996) provided a starting point that showed that criteria for quality judgement of information could be defined and investigated. It was also made clear that user's decision to use, or not to use, information is affected by criteria for information quality they deemed important to know, when making such judgements.

Criteria for judgement of quality of health information needed identifying. Also involved with the criteria were the processes by which the general practitioner (GP) used them in the judgement of health information. The criteria, and the process used by the GP when they are present or absent, to determine usage of health information, are inextricably linked i.e. the criteria's presence, or absence, affects the process of judging health information quality.

The link highlights the special circumstances within the profession of GP, encapsulated by codes of ethics such as the New Zealand Medical Association's Principles of Ethical Behaviour that all medical practitioners must acknowledge and accept, where the first principle is to always consider the health and well being of the patient as the first priority (New Zealand Medical Association, 2013). The code of ethics is informed by previous, and current, codes from other sources such as the Hippocratic Oath and the World Medical Association's Declaration of Helsinki. It highlights the application of practical wisdom, or *phronesis*, touched on at the start of Chapter 1, discussed in detail in Section 6.3.1, and part of the reflective nature of the health profession. These codes of ethics constrain, and inform, a GP's assessment of health information.

With the importance of the relationship between quality criteria for data, and their use in judging data quality established, candidate criteria for assisting in judging data quality, within the specific domain of health information, were identified. Identification was achieved via a triangulation of sources. Sources were:

1. Academic literature e.g. a study investigating how users of medical information websites determine good quality ones from others (Wilson, 2002).

2. Standards produced for collecting and communicating health information e.g. ASTM International's ASTM E2369 - 05 Standard Specification for Continuity of Care Record (CCR), and the International Organisation for Standardisation's ISO 13606 Health informatics Electronic Health Record Communication.

3. Currently used, and proposed, electronic health records (de-facto standards) e.g. New Zealand's proposed Shared Care Record, and MedTech Global's MedTech 32 Practice Management System.

The sources for criteria for quality evaluation, specific to health information and especially specific to GP's judgement of the quality of health information, provided an incomplete picture. The gap was filled by an interview process with GPs. The interviews not only filled the identified gap, but also evaluated the candidate criteria, already identified, with the GPs.

Through analysis and interpretation of interview transcript data, combined with a criteria rating exercise, candidate criteria were developed into Quality Criteria (QC) for judgement of health information. QC were well supported by themes identified throughout all interview transcripts, as well as the rating exercise scores. The thematic analysis and rating exercise were used to reduce the number of candidate criteria, developed into QC, by retaining only those as QC that were most important to the GP when evaluating quality of health information.

QC are features of data most valued by the GP when evaluating health information "fitness for purpose" i.e. its quality, and whether the GP has confidence they can safely rely on, and use the information. A GP, from the group interviewed, evaluates information they are presented with. None accepted, or rejected, information at face value. The evaluation process is against an internal framework, used to assess the information. The internal framework being the process followed with quality assessment.

The QC, successfully identified in this research, are quality criteria applied, or looked for, during the assessment process and fulfil underlying questions from the GP. Some QC relate to judging whether there is sufficient information, others whether the information triggers a need to look further e.g. does not "make sense" to the GP for some reason. Still other QC cause judgement of the information, based on tacit knowledge e.g. prior experience, or community knowledge. A large portion of the evaluation strategy can be stated as: "Who did this?", "Why is it in the record?", and "How did it get there?".

Research Question 2:

> What information (meta-data), that currently accompanies electronic health information, can determine these criteria?

With Quality Criteria (QC) identified, there was a need to identify if current mechanisms for health information recording and communication, provided for capturing/determining QC. Sources investigated for candidate quality criteria in Chapter 3, were also investigated for potential to capture the criteria. In turn, this meant the ability to capture QC.

Sources investigated were not restricted to the domain of health. Domains indicated in Chapter 3, Section 3.2 were investigated for meta-data for assessment of data quality,

and data quality in decision making. Studies ranged from data tagging (with meta-data) (Shanks, 2001; Price and Shanks, 2010), with scores for quality dimensions such as accuracy (Fisher et al., 2003), to process meta-data that describes the provenance of the data at hand (Shankaranarayanan et al., 2006) i.e. how it was created, processed, and stored, this included the Massachusetts Institute of Technology's Total Data Quality Management Program (TDQM) (Wang, 1998), and the Canadian Institute for Health Information Data Quality Framework (CIHI, 2009). Much of the work was database related, across all domains, with the application of data quality dimensions not only to data, but functions performed on data e.g. data aggregation (Parssian, 2006).

If using data, meta-data is also being used. The studies, frameworks, and standards, investigated how meta-data could be used to aid the data user, through adding meta-data e.g. tagging, that assisted the data user in their use of the data, or using meta-data to aid in the production of quality data. Using meta-data to value abstract concepts found in data quality dimensions such as accuracy, can have counter-intuitive results, where tasks using data for decisions are involved. For example, the conclusions of some studies into data tagging was that applying meta-data this way, did not enhance decision making for experts in the decision making context (Fisher et al., 2003; Price and Shanks, 2011). Including a score for a quality dimension e.g. accuracy, in some cases, increased decision time, even when the same decision was reached (Price and Shanks, 2010). The conclusions added weight to the decision in this research, to not apply meta-data to health information as scored quality dimensions (in the abstract), rather as indicators for information the GP had declared as important to know, when assessing health information quality.

The search revealed the importance of meta-data, not only for data quality work, but for data generally. Meta-data is used to describe data, to add information about data. This is true of health meta-data as well, and highlighted by research involving standardised meta-data repositories/registries for describing health data (Kerr, 2006; Australian Institute of Health and Welfare, 2013), and meta-data utilised for assessing quality of health practices, health outcomes for patients, generation of quality health data (Brouwer et al., 2006), and retrieval of health information from clinical databases (Brown and Sönksen, 2000). With the development of candidate quality criteria for health information, meta-data was encountered extensively. This was from academic studies into reputation systems for on-line Personal Health Records (van Deursen et al., 2008), to provision in health information standards and currently used medical records, for capturing (via meta-data) candidate criteria, as well as finalised QC, such as attestation of information (Beale, 2006).

Table 6.1: QC that can be captured by current meta-data use.

QC Determined Using Currently Available Meta-data
• C4: Attestation • C7: Attestor Role • C8: Contextual narrative • C9: Further information • C10: Clinician Qualifications • C12: Data age • C13: Change narrative • C14: Information captured • C15: Clinician Role

Most QC were able to be determined from, or captured by, currently used meta-data. Table 6.1 shows the QC that are able to be determined from implementation of health information standards, or from currently used records used for health information.

Mechanisms for capturing QC C4: Attestation, C7: Attestor Role, C10: Clinician Qualifications, and C15: Clinician Role are described in Chapter 5 as part of describing data structures used in prototype development that used an implementation of the CCR standard. C9: Further Information, and C13: Change Narrative were also discussed, where the CCR Comments and References data types are used to refer to additional internal and external data, pertinent to a record entry. C12: Data Age is a derived attribute, reliant on a date and time stamp for a record entry, an extremely common meta-data attribute. The QC C14: Information Captured is easily represented in the CCR as an actor reference. All QC in Table 6.1 are able to be represented using implementations of the CCR standard, with many also being able to be captured by other standards' implementations, and currently used medical records e.g. openEHR, HL7 CDA, or New Zealand's proposed Shared Care Record.

Research Question 3:

> What additional information (meta-data), that is not currently available with electronic health information, could determine these criteria?

Several Quality Criteria (QC) would not be able to be implemented as meta-data using current implementations of standards, or currently used medical records, without additional data, meta-data, or mechanisms within the data structures of health information. While possible to force current mechanisms to capture these QC, the added functionality likely breaches the current understanding for the prescribed use of the

standards, or currently used records. For example, using the CCR data type InternalCCRLink (used for referencing one CCR data object to another), it may be possible to show that two entries contradict each other i.e. C1: Contradictory entries. However, this is outside the scope the CCR standard defines for the InternalCCRLink; where the semantics for the LinkRelationship element would be an enumerated value i.e. from a pre-defined list (ASTM, 2005). It is not to say that a value such as "conflicts with" could not be added to an enumerated list. Or that, as the LinkRelationship element is of type string, that a system could not use a value such as "conflicts with".

More difficult would be using current meta-data elements to determine QC such as C2: Organisational Expertise, C3: Organisational Reputation, C6: Personal Experience, or C11: Clinician Experience. The difficulty in these QC lies in the nature of the QC, that being they are subjectively evaluated. By definition, the subjective nature of the QC means that even if they were determined/captured with health information, any determination of these QC would not apply across all GPs. For example, one GP may consider an organisation that was the source of an entry to have an excellent reputation, while another GP may disagree. Regardless, the subjective QC cannot be represented without additional meta-data to indicate their presence or absence.

The lack of meta-data for these QC is for several reasons. Health information standards generally do not define how they are implemented, they only define the composition of an Electronic Health Record (EHR), or function that EHRs should fulfil. For example, ISO 18308 Health informatics Requirements for an electronic health record architecture states that "The EHR shall be able to represent intentions, goals and care plans" ISO (2008b). It does not state the data, or meta-data, required to fulfil this requirement.

One other reason is that the QC not provided for with current meta-data, are one level further removed from the data content. This is seen with C2, C3, C6, and C11. These QC provide information on the source of an entry (the entry being data content). The source itself is meta-data, with information on the source being meta meta-data. Meta meta-data is already able to be represented in some cases e.g. C15: Clinician Role, the role describes something of the source. However, role is easily defined, and therefore easily able to be included, while information based on subjectively defined concepts is not. This is one of the reasons why these subjectively defined QC were not implemented using the prototype.

Research Question 4:

> If present, are the criteria able to be detected by a system?

Those Quality Criteria (QC) able to be captured using current meta-data used for implementation of Electronic Health Record (EHR) standards, and currently used medical

records, can be detected by a system. Prototype development in Chapter 5 provided evidence, and examples, of this. Demonstration of functionality was the reason a prototype was chosen for implementation, being a common purpose for prototypes.

Not all QC, that could be determined by current meta-data (refer Table 6.1), were implemented in the prototype. The reasons for this were threefold. First, resources and time restricted the researcher's ability to implement detection of all detectable QC. Second, it was felt that demonstrating the data structures, and mechanisms, within a record based on an implementation of the CCR standard was able to demonstrate how it could be achieved. Last, the logic of how this could be achieved was the important part, as the research was not to demonstrate a single implementation, but rather the mechanism that could be developed, and applied, across different health information models.

Research Question 5:

> If detected, are the quality criteria able to be modelled with the electronic health information and indicated to the clinician?

The Quality Criteria Model (QCM) successfully captures instances of realised Quality Criteria (QC). The example health information model used with the prototype (CCR) allows inclusion of additional information via the `CodedDescriptionType` complex data type. Within the `CodedDescriptionType` is the ability to add an `ObjectAttribute` element, with a `Value` element contained within the `AttributeValue` element. The `Value` element is declared in the CCR schema as being of type `xs:anyType`. The `xs:anyType` declaration supports the `Value` element being of any simple, or complex, data type. The XML-Schema type `xs:anyType`, is the root of the XML type definition hierarchy tree, where every type definition is an extension, or restriction, of some other type definition (World Wide Web Consortium (W3C), 2013). The declaration as `xs:anyType`, provides for the `Value` element, of the `AttributeValue` element, for a new `ObjectAttribute`, to take the form of any XML type.

When QC are detected, or realised, from health information, they are captured by instantiating the QC via the QCM. The instance of the QCM is an instance of a complex data type beginning with the XML element `quality-criteria`, and is defined by the QCSchema model. QCSchema conforms to the XML-Schema standard. Therefore, it can be added to CCR based health information, by taking the value of `Value` element, of `xs:anyType` data type, within the `CodedDescriptionType` CCR complex data type. This mechanism allows for insertion of the QCM into a CCR based EHR, carrying the results of successful searches for QC (HIQI) with the health information.

With the `CodedDescriptionType` data structure (complex data type) used throughout the CCR health information model, the QCM could be applied to a CCR based document at multiple points. For example, both `Reference`, and `Comment`, elements' type definitions have `Description` elements of type `CodeDescriptionType`. The QCM could be applied as a `Reference`, or `Comment`, should the QCM instances be required physically with health information, and the health information would still conform to the CCR standard. The prototype examples used a `Comment` element.

As discussed in Chapter 4, Section 4.4.2, QC are realised through the dimensions of an entry, a source for an entry, or a document as a whole. The recursive structure of the QCM provides for capturing multiple QC within one root instance of the QCM. A single Comment, referenced from each entry, source, or document, via the Comment ID, would capture all instances of QC for each entry, source, or document.

After the detection of QC, a map is generated, containing both the IDs for each entry, source, or document, that has realised QC, and the IDs for the associated Comments. A Comment is effectively a set of QC, as the QCM captures multiple QC, within a single Comment. With this map, indexing instances of the QCM (realised QC) to the sources (entry, source for entry, and document) for those instances, Health Information Quality Indicators (HIQI) can be generated. The indexing process provides for referencing each set of QCM instances, within a single Comment element, to the source for the QCM instances i.e. the information entry, a source for an entry, or document as a whole. HIQI are then derived for each instance of QCM, and are able to be applied to the source for those instances.

With QC detected, and captured, and originating data for the QC able to be referenced, producing HIQI for the health information can occur. This satisfies the proof of concept requirement and answers Research Question 5. The form of these indicators is beyond the scope of this research. Realised QC can be modelled with health information, and made available to the system, that is then able to present them to the GP as HIQI. However, an example could be a simple list of entries, with their associated QC. The entry/QC pair becoming a HIQI. Alternatively, the QC could be annotated to an entry, with the annotation becoming a HIQI. This study was a computational approach to HIQI i.e. using a system to detect, and model QC with health information, with Human Computer Interaction (HCI) interface aspects, not included.

6.2 Thesis Contributions

The thesis set out to identify criteria used by a group of general practitioners when assessing health information, to show these criteria are able to be detected, formally

modelled, and incorporated with the health information. The successes in these areas is where the thesis makes its major contributions. However, as the project incorporated research from multiple disciplines, it required application of thought from these disciplines, resulting in several minor contributions. Contributions are now discussed.

- *Identification of criteria used by a group of practising GPs for assessing quality of health information, that assists them in the determination of the health information fitness for purpose.* The criteria were identified by the GPs, as important to know when making judgements of data's fitness for purpose in clinical situations. This contribution set the groundwork and required analyses into the GPs approach to evaluation of information quality, for clinical use. The fact that these GPs consistently evaluated information, as to whether it "ticked" boxes that satisfied their expectation for quality. Alternatively, that evaluating information, or finding more information, was necessary when flags were raised by the evaluation process e.g. knowing something of the source of the information.

- *Development of thematic maps that capture the combination of criteria, and the process used when applying the criteria.* The thematic maps model the internal framework applied by the GPs when assessing health information, as they do not just accept information, but constantly evaluate it for the purpose of accepting, or rejecting, its content. There is a process that all GPs in this study followed, largely summarised by questions regarding the information: Who did it?, Why is it there?, and How did it get there? The thematic maps capture this with their emphasis on tacit knowledge, that can be sub-divided into experience, and community knowledge.

- *Identified that Quality Criteria (QC) are able to be formally defined, and categorised.* The QC are able to be defined through three dimensions of the health information: entry, source for entry, or whole document. The QC are evaluated for what information they add to these dimensions. The QC can be categorised by what type of evaluation is required to realise the QC i.e. a subjective, or objective evaluation. The QC can also be categorised by what they relate to. For example, do they relate to an entry, a source for an entry, or to the document as a whole. This last categorisation is mutually exclusive e.g. QC for an entry are not able to be applied to the source for an entry.

- *Development of a model for capturing the semantics of Quality Criteria, the Quality Criteria Model (QCM).* The QCM provides a formal mechanism for capturing QC, and storing them within a machine processable data structure. The QCM is defined explicitly through an XML-schema, the QCSchema, that applies tried

and tested technology for capturing, and processing, semi-structured data. The QCSchema provides functionality for system validation of any instance of the QCM, thus ensuring that QC instanced, using the QCM, conform to the model.

- *Demonstrated that QC are able to be detected using meta-data, within a current health information model, the CCR, without the need to alter, or extend, the CCR model.* The application, and knowledge of, strategies such as data as objects, and data type specialisation, greatly aided in the development and success of this approach. The flexibility of the CCR, through use of these types of strategies, provided this mechanism. These techniques are also used within other health information standards e.g. the developing HL7's FHIR standard, as demonstrated with the example that used the FHIR Extension mechanism (see Chapter 5, Section 5.8). This shows the potential use of the QCM with other health information models.

- *Demonstrated that QC, once detected/realised, can be instantiated and incorporated into health information, via the QCM, and their presence indicated to the GP as Health Information Quality Indicators (HIQI).* The QCM is able to be applied to a health information model, the CCR, without requiring alteration, or extension of, the CCR specification. This is of advantage and shows QC, via the QCM, are able to be modelled with the health information they enhance. The prototype demonstrated the process; showing QC detected are transformed into instances of the QCM. The instances are applied to health information, using mechanisms available from the CCR. The prototype maintained link references between QCM instances, and the sources for these references. The links are required to process QC, modelled by the QCM, into system generated HIQIs. In other words, the captured QC, and what the QC were detected from, is the information required for a system to generate HIQI. The prototype demonstrated the required functionality.

- *Found that additional meta-data is required to detect QC, categorised as subjectively evaluated.* Subjective evaluation proved one of the most challenging aspects of QC i.e. how to detect, and incorporate into health information, subjective QC? Subjective QC could not be incorporated into the prototype, as no health information models, for clinical information in the research context, provide for evaluation/detection/incorporation of subjective QC e.g. past experience, reputation, or expertise based QC. However, there is no doubt that these criteria are valued by the GPs, and speak clearly to reliance on experience, and community knowledge, in health information evaluation. There is also the issue that while some QC have been categorised as Objective, a case could be made that some are

Subjective e.g. C5: consistency/cohesion, as one person may evaluate a record as consistent, but another may not.

- *Theory that providing scores for abstract quality dimensions, and applying them to health information has limited utility for GPs as professionals bound to limit adverse harm to patients in clinical decision making.* Supported by the concept of *phronesis*, or practical wisdom, as applied to the profession of general practitioner. Having the information is not enough. There are things that need to be known, or the GP feels they can't safely rely on the information. The GP needs to satisfy their internal framework, such that they are confident in the safe use of the information, providing an ethical decision, and protecting themselves (medico legal requirement) i.e. took all care. That the profession of general practice brings serious challenges when it comes to information use, in clinical situations.

Many areas of research, and domains they involve, were brought to bear in addressing the research questions, and thus making the above contributions. Though this multi-disciplinary approach brought with it its own considerable challenges, it provided the researcher with many philosophies, and techniques, to address the topic of investigation. This thesis represents a distillation of the investigation, containing the core elements needed to give the reader the purpose, outcome, and contribution of the research. It can not hope to cover all avenues explored, literature, thoughts, and outcomes from this project.

6.3 Implications For Theory And Practice

The research conducted for this thesis has raised several issues. The issues are centred on information use, and evaluation for information quality, in clinical context. The issues were revealed not only through the original research conducted for the thesis, but also due to extensive investigation of the many, and varied, research domains investigated for this project. The interviews with the general practitioners (GPs) revealed these issues were common across those GPs interviewed

6.3.1 Access Alone Is Insufficient

Being professionals, the general practitioners (GPs) must first and foremost exercise a duty of care toward their patients. This is encapsulated in the New Zealand Medical Associations Code of Ethics. Just having the information is not sufficient. With the

duty of care, the GPs do not trust, or distrust, information at face vale, but constantly assess the information for its fitness for purpose.

From the group of GPs interviewed it is clear that they approach information, for clinical decision making, in a specific, practical way. It is an approach reflected by other professionals, charged with a duty of care e.g. engineers, and reflects the privileged status of the professions as a group in society (Pitman, 2012). The application of *phronesis*, or practical wisdom, has been studied since Aristotle distinguished it as an intellectual virtue in Book VI of the Nicomachean Ethics. It has been applied to the study of the professions, including healthcare. *Phronesis* is about not ignoring practical wisdom, and being caught in routine, that does not take account of the current situation, but is followed anyway. As an example, at times in healthcare, protocols may be followed regardless of context, or what is best for the patient, as routines such as protocols provide a defence, should treatment go badly. Practical wisdom is the opposite; it is about basing decisions on expertise, using reflection to break routine, thus increasing personal responsibility (Frank, 2012).

The concept of practical wisdom is related here, as it applies to how the GPs interviewed judge information. They do not just accept information in a routine manner, but apply expertise, based on the current context, with the well-being of the patient uppermost. There is routine in the assessment of information, but it is in the process, the ticking, or not, of boxes, but not routine on how similar information would be applied to different patients, in different contexts.

When providing information to the GPs, this philosophy should be noted. It is unreasonable to expect that the information should just be taken on faith, that simply having the information is of utility. This is especially true as the amount of information likely increases, and as the GP becomes more remote from the source of the information i.e. as information is aggregated from multiple sources. It would be well to take this in to account for systems intended to deliver information. Increased information volume, without the desired accompanying information used for assessment, may just cause less information use, as users feel unable to rely on it, and it would take too long to find what they want.

6.3.2 Form Is Important

Using health information is beyond a semantic interoperability, or format, issue. These issues are systems issues, with the focus appearing to be on how to easily get information into a system, storage and retrieval, security of access, and transmission. These are not end user issues, as identified in this research. The group of GPs in this research (the

end users), look for criteria within what they are presented with to assess information fitness for purpose. This is a form issue. Focus, for the end user, should be on supplying the health information in a form that allows for ease of assessment. It should provide the criteria used in assessment "up front" so the GP does not need to take extra steps to find these criteria, to make the assessment. To match the information form, to the GPs internal framework, used for assessment. The Quality Criteria Model (QCM) may be a step in this direction, by modelling criteria for assessment.

It may sometimes appear that assessment is not taking place, as the particular information strikes the GP as being of quality. However something has always keyed them in to this conclusion. Through tacit knowledge, experience and community knowledge, it has happened very quickly, but the assessment has still taken place. This is the application of the GPs practical wisdom. Instead of the focus being on systems just getting the information to the GP, though this is important, it should also take into account the way the information is assessed. The GPs are looking for information on the content, should something within the information trigger a need to look further. From the GPs interviewed this occurs frequently.

The Electronic Health Record (EHR) entry should be the logical, self-contained, discrete unit, for communicating health information. Communicate, rather than transmit, is the correct term, as communication requires a mutually understood channel. A logical unit is required, as proposed strategies for communicating health information are focusing on aggregation of data from disparate sources, where documents containing information will be fragmented (based on required context), and assembled "on the fly" e.g. New Zealand's National Health IT Plan (2010). The entry should carry with it all data (meta-data) that is required for the end user to be able to use the information. One aspect of this use, is to ensure it can be used safely, with a judgement of the information fitness for purpose.

6.3.3 Subjective Criteria Are Problematic

Those Quality Criteria (QC) that are unequivocally subjective i.e. reputation, past experience, or expertise based, will be non-trivial to implement in a system. Not just how to evaluate them, but also their representation. Also, their very nature, being subjectively determined, makes a "one size fits all" approach, not only difficult, but probably undesirable.

Proposed strategies for the NZHS may provide some help. For example, the proposed introduction of league tables for New Zealand health organisations could be leveraged

where organisational expertise (C2), or organisational reputation (C3), criteria are involved. It is likely that the best approach for subjective criteria is some form of peer, or community, rating from GPs. This approach is likely to encounter heavy resistance, even if confidentiality, or anonymity could be preserved. It may sound counter-intuitive to talk of anonymity for something like reputation of sources. However, a potential solution could be a confidential rating for sources of information, where peers confidentially rate one another, and the rating is anonymously applied to information. This would rely heavily on an ethical use of ratings by peers.

6.3.4 Difficulties with Multiple Entry Criteria

Entries will not appear in health information in isolation. They occur as collections i.e. in a document structure of some type. This is true for current health information. Two Quality Criteria (QC) relate specifically to a collection of entries (C1, and C5), not single entries, or the source for an entry. These are the three dimensions entry, source and document, described in Chapter 4, Section 4.4.2 . If information sources are to become more connected, as is proposed, potentially aggregated from disparate sources, for delivery of information to the end user, the issues with these two criteria become even more important.

Even if the aggregation of data is limited to that required for the current context of need, there is the potential for a lot of information. Having the ability to have quickly highlighted, where contradictions occur (C1: Contradictory Entries), was rated highly by the GPs interviewed. Leaving it to the GP to decipher this criterion may result in the criterion being missed, or too much time taken, especially for information they are not familiar with, or when information volume is high.

C5: Consistency/cohesion is a quality criterion that evaluates the collection of entries as a whole i.e. Does the collection/record “hang together” as a whole, or “make sense” as a whole? C5 is categorised in Chapter 3, Figure 3.9, as being an objective criterion. This was based on finding contradictory entries, implying inconsistency, as per processing example in Chapter 4, Figure 4.7. However it could be argued that the nature of the question, just posed, means it could be interpreted as a subjective criterion. Either way, assessing if a collection of entries is consistent/cohesive, especially for information from disparate sources, and not seen before, would be challenging, for both human or system. Having said that, it is deemed by the GPs interviewed to be a highly rated criterion, and therefore important to know when judging information.

6.4 Thesis Limitations

There are limitations to this thesis. These are identified, and discussed in the following points.

- The prototype was not formally evaluated. This research was to provide a proof of concept, to demonstrate the the approach taken is feasible. The prototype was a model, computational, and feasibility demonstrator. A formal evaluation would have required the implementation of a user interface for GPs to use the system. It was felt that anything short of what the GPs were use to seeing, would have been too abstract for them to evaluate the purpose of the test e.g. evaluating the interface, not the content. Development of an interface would be a project in and of itself. It would need to include representation of HIQI. The representation used would also influence the evaluation of HIQI, with the need to find an appropriate representation, or implementation of multiple representations. This, again, would lead to an evaluation of the representation, not the content (HIQI).
- The subjective categorisation of some QC could be disputed. For example, C5: Consistency/cohesion could be interpreted as objective as well as subjective. C11: Clinician experience, could also be interpreted as a subjective criterion, rather than objective i.e. some may accept that a certain number of years in a role qualifies as experienced, while others may judge additional factors are required for experience.
- The small number of GPs interviewed is a limitation. This was outside the control of the researcher due to the low number of positive responses to requests. However, as this was a proof of concept study, and results are not being claimed as generalisable across all GPs, it is not seen as an issue, but an opportunity for future research. Also, those interviewed showed consistent themes, and agreement, across criteria and findings.

6.5 Recommendations

It is accepted that information quality issues within health information should be addressed in a holistic, design oriented approach Orfanidis et al. (2004); Kerr (2006); Mettler et al. (2008). This is not only true for health information but also for information quality across other domains. For instance the Total Data Quality Management (TDQM) methodology, where information is treated as a product that moves through a manufacturing system (Wang, 1998).

However, the focus from these studies has been on the production, storage, and management of quality information, along with inter/intra organisational communication. With the move toward aggregation of health information from disparate sources at point of care and at time of need, in essence "on the fly", assessment of information quality, at that specific point in time, has not been addressed. Certainly, it may be possible to declare that information aggregation comes from sources that have high quality information, but as shown in this thesis, this is not sufficient for the group of GPs interviewed, as they themselves need to justify acceptance, or rejection, of the information and can't rely on evaluations made by others. This will be further exacerbated as more health information becomes available for use and the GP is further removed from any knowledge of the information source and production methods.

It is recommended that health information quality issues surrounding the assessment of information quality at time of use by GPs, as shown in this thesis, be treated with the same importance as the holistic, design oriented approach accepted for production and storage of information. Otherwise efforts in making more health information available, for use in clinical decision making at point of care by GPs, may not realise its full potential.

Another recommendation would be that the "entry" be the unit of information, pulled together on request. Several models use entries in this way. Usually with a surrounding document. Entries can already contain all information necessary to reference a containing record/document, the subject of the entry, and authentication, allowing for combination of entries into a document when needed.

Chapter 7

Conclusions And Future Work

Rationale for this research was the knowledge that within the New Zealand Health and Disability Sector (NZHS) major initiatives are moving toward linking health information systems with the goal of making health information much more freely available for use by health practitioners, at point of care. Some of the key drivers for this, spelled out in the *National Health IT Plan 2010*, are to reduce costs, increase efficiencies, and improve health outcomes, to be achieved by leveraging the use of health information. This raised questions as to the use of health information by health practitioners; for this thesis a group of practicing GPs. Specifically, what is it that the GPs want to know, regarding health information, when assessing the quality of health information?

An extensive review of relevant literature, currently used electronic health information, and standards for the production and communication of health information yielded little for quality criteria used by GPs to help them assess health information quality. With literature, the review moved beyond the domain of health into many other domains that also attempt to address information quality issues. The lack of answers for the research questions necessitated interviewing the GPs to determine the quality criteria important to them to know, and how those quality criteria contribute in the assessment process. The interviews provided much valuable information and were a very informative component of this thesis.

Interview results, analysis, and interpretation showed the GPs will evaluate health information they are presented with. The assessment is based on an internal framework that utilises tacit knowledge, community knowledge, and past experience. The GP assesses whether the information "smells" right, that it is safe to rely on, and is fit for purpose. This may not be what is consciously thought during the process, but it is the result. Information may fail this test, triggering another process. This may be due to not enough information to decide, or, that the information does not hold together, does

not "make sense" when evaluated against their internal framework or in the current consultation.

The process the GPs would enter, upon questioning any information, is they need to investigate further. The GPs can not simply ignore, or accept, the information, without first justifying why they should ignore/accept it. This is in line with the reflective nature of the profession they are in, and the requirement to act in the best interest of the patient. The Quality Criteria (QC), formalised in this thesis, have been identified as key when there is a need to assess health information.

The identified QC reflect three areas of detail for health information that the GPs investigate when assessing health information quality. The three areas are details for the entry itself, the source for the entry, and how multiple entries hang together as a whole i.e. the consistency and cohesion for a collection of relevant entries. This can be summarised as questions by the GPs that featured strongly from the interviews, namely, Why is it (entry) in the record?, Who did it?, and How did it get there?

With the identification of this set of QC, it remained to formally define the QC and demonstrate that they could be detected, if present in health information, and the semantics of the QC captured to be made use of as Health Information Quality Indicators (HIQI). The definition of HIQI being realised QC, or with the prototype, the positive result from search functions, rather than an actual user interface component that the GPs would see. HIQI could be used later as the basis for any user interface component, used to communicate to a GP that instances of QC exist for the associated health information.

The Quality Criteria Model (QCM) was developed, and successfully demonstrated that it was capable of representing all instances of QC, from the set of identified QC. The QCM providing a formalised representation for describing QC. The QCM was formally implemented in an XML-schema (QCSchema), that described the model in a way that would allow it to be implemented, and used, within a computerised system. It was shown that the QCM is flexible enough to be used in more than one way. It was shown that it could be incorporated directly into health information, using a health information model (CCR) that can form the basis for real world EHRs. The QCM could also use unique identifiers, integral to the design of QCM, to store QC instances separate from health information, and be referenced from the health information, linking QCM instances to the information they describe.

The recursive structure of the QCM also allows the QCM to act as a container, storing all realised instances of QC for an entry in a single data structure. This allows for a one to one correlation between an entry, a discrete entry in an EHR, to the known, realised QC for the entry. The recursive structure does not limit the ability of the QCM

to be flattened, with single instances of QC being recorded per instance of the QCM, should this be required e.g. for grouping together of QC of the same type for secondary analysis.

The last two research questions (see Chapter 1, Section 1.2.1) were addressed with the development of a prototype system. The prototype successfully showed that QC, if present within health information, could be detected using the prototype. Detection of the presence, or absence, of QC was not the original premise for the use of QC. That being to use them to produce a score for the health information quality, the reasoning for this change is sound and has been discussed, and justified, in Chapter 4, Section 4.4.1. The prototype further showed that QCM instances could be created for detected QC, and validated as legitimate instances of the QCM, using the developed QCSchema. The prototype used a current health information model, the CCR standard from ASTM International, without alteration to the CCR standard. This adds further weight to demonstrating successful use of the prototype, and the QCM.

However, not all QC could be used with the prototype in its current form. These QC were those requiring subjective evaluation such as C3: Organisation reputation. The reason being their subjective nature that is entirely dependant on the individual GP. Additional information would be needed for the prototype to use these subjective QC. This is not to say that they are of less value, indeed some of the subjective QC rated very highly with the GPs, reinforcing their value, and the fact they can not be ignored if such a system is to be further developed.

With the prototype the last of the research questions was successfully addressed. This allowed a move to discussion, and the successful completion of this thesis. The discussion described how the research question had been addressed and the contributions of this thesis made clear.

7.1 Future Work

The thesis presented here dealt with questions of determining health information quality, particularly criteria used by a group of GPs during that process. Issues discovered during the course of the research project highlight several questions for further research. If addressed, these questions would further extend the knowledge gained in the course of this study. It is considered that each of the topics in the following sections could be a research project in and of themselves.

7.1.1 Generalisability of Quality Criteria

The set of QC successfully identified from this research need further investigation. This is because, while found to be important to know for the group of GPs interviewed, and part of these GPs assessment of health information, it remains unknown whether the set of QC are important to all GPs. A study to determine the generalisability of these QC would be of value in this respect. The question could be addressed in several ways such as surveys/questionnaires, to further interviewing of a larger pool of GPs.

The question could be further extended to health practitioners in other domains of health. For example, secondary healthcare such as hospitals, could be investigated with a similar approach to that used for this thesis. This could include not only hospital doctors, but nursing staff as well. Another domain of health that could be of value would be ambulance staff. This is in line with making health information available to these first responders. This includes not only emergency care situations, but also in their duties for patient transfers.

7.1.2 User Interface for Representation of QCM Instances to Users

The prototype developed during this research project was used to demonstrate two things: that QC, if present, could be detected in health information, and that the prototype could use the QCM to capture and store realised QC. As a prototype, it is not a system that could be used at point of care by GPs. Part of developing a production system from the prototype would be in developing, and testing, a user interface for communicating the results of searching health information for occurrences of QC.

The question could not simply involve a user interface, but a search for an optimal interface for the system. This would be in the domain of Human Computer Interaction (HCI) and would include questions as to the type of display used for the interface e.g. could be annotation (could be many forms for this) of QCM instances for realised QC directly to the relevant entries, to simply listing the QCM instances and their generating entries.

7.1.3 Application Of Subjective QC

One of the most interesting discoveries during this research project was subjectively evaluated QC, and their importance. Future questions to be answered would be how to incorporate these type of QC within the prototype i.e. how to perform evaluation for such QC as C6: Past experience (with information from a particular source).

Several approaches could be used to attempt to resolve this question. For example, a community rating system such as those used for rating sellers on internet sites such as TradeMe or eBay. However, this is likely to meet strong resistance from a professional community of health practitioners. An alternative could be to allow the system to be customisable for individual GPs where the GP could flag sources for reputation based on their own past, subjective experiences.

7.1.4 Evaluation of Search Algorithms for Detecting Presence of QC

The prototype developed for this thesis demonstrated a search algorithm for detecting the presence of QC within an example health information model. Alternative search algorithms would be possible and could be investigated. Alternative algorithms could be focused on increased efficiency of searches, alternative methods and ordering, or implementation using different health information models. For example recursive algorithms for recursive data structures.

Search algorithms could also be employed on collections of health information and not just at time of use. Health information entries being the discrete, logical unit, that searches for this thesis involved could be pre-searched in readiness for aggregation at point of use. This could be applied for use with Regional Clinical Data Repositories (R-CDR), proposed as part of the *New Zealand National Health IT Plan*.

7.1.5 Development of Prototype Toward Production System

The prototype was used to demonstrate functionality that directly addressed research questions for this thesis. It was not designed for use as a production system. However, its function could be extended and developed for use as a production system. Work would be focused on usability aspects such as user interface, integration with systems from the health domain, and thorough testing of these new developments.

Alternative design, tools, and development environments could be explored. This may be required for integration into current environments as the prototype is unlikely to be a standalone application when used in a production environment. This is standard practice with software development. The development toward a production system, though considered a highly desirable goal, would not be a trivial exercise. Issues highlighted here, such as user interface design and systems integration, are complex tasks, requiring in-depth requirements analysis, good design and implementation principles, and extensive testing.

7.2 General Conclusions

If more health information is to be made available for GPs to use in clinical treatment situations at point of care, then consideration needs to be given to how GPs will approach the use of this information. This is especially true as the GPs are likely to become further removed from any personal experience of the source for the information and how it has been made available to them.

There is no doubt that increased availability of a patient's health information at point of care for clinical decision making could yield efficiency gains and cost reductions. However, if GPs feel unable to rely on the information, or feel they will need to spend too much time and effort to justify its use, they are unlikely to exploit the benefits to their fullest potential.

This thesis supports these claims with the following general conclusions:

1. Quality Criteria (QC) used by GPs when assessing health information are able to be identified. Further, the process by which the GPs apply those criteria in assessment of health information quality can be understood. The successful discovery in this thesis of these two points can be used to further leverage the use of health information.

2. The QC can be formally defined and modelled, thus providing the ability to capture QC semantics. If QC can be captured, they are able to be used in support of assessment of health information quality.

3. QC can be detected using computer systems and meta-data from health information models. Additional meta-data may be required to allow for detection and capture of some QC. For example, those QC that are based on subjective assessment by the GP. As subjective QC are rated highly by the group of GPs interviewed for this thesis, the additional meta-data should be explored with a view to incorporation into health information.

As much time, effort, and resources are being invested in developing strategies and systems to make health information more widely available, important aspects as how users of such information determine fitness for purpose should not be ignored. Indeed, this thesis argues that this factor be integral to any proposed strategies and systems, designed to facilitate increased access for GPs to health information in clinical decision making contexts, at point of care.

Bibliography

ASTM (**2005**). *ASTM E2369 - 05 Standard Specification for Continuity of Care Record (CCR)*. American Society for Testing and Materials International.

Atalag, K., Kingsford, D., Paton, C., and Warren, J. (**2009**). Putting health record interoperability standards to work. *electronic Journal of Health Informatics*, 5(1): e1.

Atkinson, R. and Flint, J. (**2001**). Accessing hidden and hard-to-reach populations: Snowball research strategies. *Social research update*, 33(1): 1–4.

Australian Institute of Health and Welfare (2013, METeOR Home). Internet. Retrieved 2013.11.27, from `http://meteor.aihw.gov.au/content/index.phtml/itemId/181162`.

Ballou, D. P. and Pazer, H. L. (**1995**). Designing information systems to optimize the accuracy-timeliness tradeoff. *Information Systems Research*, 6(1): 51–72. . Retrieved from www.scopus.com.

Batini, C., Cabitza, F., Cappiello, C., and Francalanci, C. (**2008**). A comprehensive data quality methodology for web and structured data. *Int. J. Innov. Comput. Appl.*, 1(3): 205–218. doi:10.1504/IJICA.2008.019688.

Batini, C., Cappiello, C., Francalanci, C., and Maurino, A. (**2009**). Methodologies for Data Quality Assessment and Improvement. *ACM Computing Surveys*, 41(3). doi:10.1145/1541880.1541883.

Beale, T. (2006, September). openEHR Release 1 ISO 18308 Conformance Statement. Internet. Retrieved 2011.12.09, from `http://www.openehr.org/programs/specification/releases/1.0.2`.

Blumenthal, D. and Tavenner, M. (**2010**). The Meaningful Use Regulation for Electronic Health Records. *New England Journal of Medicine*, 363(6): 501–504. doi:10.1056/NEJMp1006114. PMID: 20647183.

Boehm, B. and In, H. (**1996**). Identifying Quality-Requirement Conflicts. *IEEE Softw.*, 13: 25–35. doi:http://dx.doi.org/10.1109/52.506460.

Boxwala, A. A., Dierks, M., Keenan, M., Jackson, S., Hanscom, R., Bates, D. W., and Sato, L. (**2004**). Organization and Representation of Patient Safety Data: Current Status and Issues around Generalizability and Scalability. *Journal of American Medical Informatics Association*, 11: 468–478. doi:10.1197/jamia.M1317.

Boyatzis, R. E. (**1998**). *Transforming qualitative information: Thematic analysis and code development*. Sage Publications, Inc.

Braun, V. and Clarke, V. (**2006**). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2): 77–101. doi:10.1191/1478088706qp063oa.

Brouwer, H., Bindels, P., and van Weert, H. (**2006**). Data quality improvement in general practice. *Family Practice*, 23: 529–536.

Brown, P. J. and Sönksen, P. (**2000**). Evaluation of the quality of information retrieval of clinical findings from a computerized patient database using a semantic terminological model. *Journal of the American Medical Informatics Association*, 7(4): 392–403. doi:10.1136/jamia.2000.0070392.

Buccella, A., Cechich, A., and Domingo, G. (**2008**). Analyzing and improving data quality. *Journal of Computer Science & Technology*, 8.

Burgess, M., Fiddian, N. J., and Gray, W. A. (**2004**). Quality Measures and The Information Consumer. In *Proceedings of the Ninth International Conference on Information Quality (ICIQ-04),* pp. 373–388. doi:10.1.1.106.83.

Burgess, M., Gray, W. A., and Fiddian, N. J. (**2002**). Establishing a Taxonomy of Quality for Use in Information Filtering. In *Proceedings of the 19th British National Conference on Databases: Advances in Databases,* BNCOD 19, pp. 103–113. Springer-Verlag, London, UK, UK. Retrieved 2011.08.30, from `http://dl.acm.org/citation.cfm?id=646104.678790`.

Burgess, M. S. E., Gray, W. A., and Fiddian, N. J. (**2007**). Using quality criteria to assist in information searching. *International Journal of Information Quality*, 1: 83–99. doi:10.1504/IJIQ.2007.013377.

Chan, K. S., Fowles, J. B., and Weiner, J. P. (**2010**). Review: electronic health records and the reliability and validity of quality measures: a review of the literature. *Medical Care Research and Review*, 67(5): 503–527. doi:10.1177/1077558709359007.

Chengalur-Smith, I. N., Ballou, D. P., and Pazer, H. L. (**1999**). The Impact of Data Quality Information on Decision Making: An Exploratory Analysis. *IEEE Trans. on Knowl. and Data Eng.*, 11(6): 853–864. doi:http://dx.doi.org/10.1109/69.824597.

CIHI (**2009**). *The CIHI Data Quality Framework*. Canadian Institute for Health Information, Ottawa, Ont. Retrieved 2013.06.19, from `http://www.cihi.ca/CIHI-ext-portal/pdf/internet/data_quality_framework_2009_en`.

Classen, D. C., Pestotnik, S. L., Evans, R., Lloyd, J. F., and Burke, J. P. (**1997**). Adverse drug events in hospitalized patients: Excess length of stay, extra costs, and attributable mortality. *JAMA*, 277(4): 301–306. doi:10.1001/jama.1997.03540280039031.

DiCicco-Bloom, B. and Crabtree, B. F. (**2006**). The qualitative research interview. *Medical education*, 40(4): 314–321.

Dick, R. S., Steen, E. B., and Don E. Detmer, I. o. M., Editors; Committee on Improving the Patient Record (**1997**). *The Computer-Based Patient Record:An Essential Technology for Health Care, Revised Edition*. The National Academies Press. Retrieved 2012.06.28, from `http://www.nap.edu/openbook.php?record_id=5306`.

Dolin, R. H., Alschuler, L., Beebe, C., Biron, P. V., Boyer, S. L., Essin, D., Kimber, E., Lincoln, T., and Mattison, J. E. (**2001**). The HL7 clinical document architecture. *Journal of the American Medical Informatics Association*, 8(6): 552–569. doi:10.1136/jamia.2001.0080552.

Dolin, R. H., Alschuler, L., Boyer, S., Beebe, C., Behlen, F. M., Biron, P. V., and Shvo, A. S. (**2006**). HL7 clinical document architecture, release 2. *Journal of the American Medical Informatics Association*, 13(1): 30–39. doi:10.1197/jamia.M1888.

Dondio, P. and Barrett, S. (**2007**). Computational Trust in Web Content Quality: A Comparative Evalutation on the Wikipedia Project. *Informatica*, 31: 151–160.

Eichelberg, M., Aden, T., Riesmeier, J., Dogac, A., and Laleci, G. B. (**2005**). A survey and analysis of electronic healthcare record standards. *ACM Computing Surveys*, 37. doi:10.1145/1118890.1118891.

Even, A. and Shankaranarayanan, G. (**2007**). Utility-driven assessment of data quality. *SIGMIS Database*, 38: 75–93. doi:10.1145/1240616.1240623.

Even, A., Shankaranarayanan, G., and Watts, S. (**2006**). Enhancing Decision Making with Process Metadata: Theoretical Framework, Research Tool, and Exploratory Examination. In *Proceedings of the 39th Hawaii International Conference on System Sciences,* .

Fereday, J. and Muir-Cochrane, E. (**2008**). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods*, 5(1): 80–92.

Fichman, R. G., Kohli, R., and Krishnan, R. (**2011**). The role of information systems in healthcare: current research and future trends. *Information Systems Research*, 22(3): 419–428.

Finch, J. (**1987**). The vignette technique in survey research. *Sociology*, 21(1): 105–114. doi:10.1177/003803858702100100 8.

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.

Frank, A. W. (**2012**). *Phronesis As Professional Knowledge: Practical Wisdom In The Professions*, Chapter Reflective Healthcare Practice: Claims, Phronesis, and Dialogue, pp. 53 – 60. Springer.

Funk, J. D., Lee, Y. W., and Wang, R. Y. (**1998**). Institutionalizing Information Quality Practice: The S. C. Johnson Wax Case. In I. N. Chengalur-Smith and L. Pipino, Eds., *Third Conference on Information Quality (IQ 1998),* MIT Sloan School of Management, Cambridge, MA, USA.

Gil, Y. and Artz, D. (**2007**). Towards content trust of web resources. *Journal of Web Semantics*, 5: 227–239.

Google (**2013**). Google Health has been discontinued. Retrieved 2013.07.17, from `http://www.google.com/intl/en_us/health/about/`.

Greenhalgh, T., Stramer, K., Bratan, T., Byrne, E., Mohammad, Y., and Russell, J. (**2008**a). Introduction of shared electronic records: multi-site case study using diffusion of innovation theory. *British Medical Journal*.

Greenhalgh, T., Stramer, K., Bratan, T., Byrne, E., Russell, J., Hinder, S., and Potts, H. (**2010**). *The Devils in the Detail: Final report of the independent evaluation of the Summary Care Record and HealthSpace programmes.* London: University College London.

Greenhalgh, T., Stramer, K., Bratan, T., Byrne, E., Russell, J., Mohammad, Y., Wood, G., and Hinder, S. (**2008**b). *Summary Care Record Early Adopter Programme: An independent evaluation by University College London*. London: University College London.

Hasan, S. and Padman, R. (**2006**). Analyzing the Effect of Data Quality on the Accuracy of Clinical Decision Support Systems: A Computer Simulation Approach. In *AMIA Annu Symp Proc.*, Volume 2006, p. 324328. American Medical Informatics Association. Retrieved 2011.08.30, from `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1839724/`.

Health Level 7 International (**2004**). HL7 Clinical Document Architecture, Release 2.0. Internet. Retrieved 2013.10.16, from `http://www.hl7.org/implement/standards/product_section.cfm?section=1&ref=nav`.

Health Level 7 International (**2013**a). Codes defined in http://hl7.org/fhir/document-attestation-mode. Internet. Retrieved 2013.11.13, from `http://www.healthintersections.com.au/fhir/document-attestation-mode.htm`.

Health Level 7 International (2013, October). Fast Health Interoperability Resources (FHIR) Ballot Reconcilliation Master - Introducing HL7 FHIR. Internet. Retrieved 2013.11.05, from `http://www.hl7.org/implement/standards/fhir/summary.html`.

Health Level 7 International (**2013**c). HL7 Standards - Section 1: Primary Standards. Internet. Retrieved 2013.07.17, from `http://www.hl7.org/implement/standards/product_section.cfm?section=1&ref=nav`.

Health Level 7 International (2013, February). RIM Maintenance and Tooling Documentation. Internet. Retrieved 2013.10.28, from `http://wiki.hl7.org/index.php?title=%20RIM_Maintenance_and_Tooling_Documentation`.

Health Level 7 International (**2014**). FHIR Overview. Internet. Retrieved 2014.01.27, from `http://www.hl7.org/implement/standards/fhir/overview.html`.

Heinrich, B., Klier, M., and Kaiser, M. (**2009**). A Procedure to Develop Metrics for Currency and its Application in CRM. *ACM Journal of Data and Information Quality*, 1(1).

Hill, G. (**2009**). *A Framework for Valuing the Quality of Customer Information*. Ph.D. thesis, Department of Information Systems, Faculty of Science, The University of Melbourne.

HISO (**2012**a). *Health Information Exchange Clinical Data Repository Utility Services Architecture Building Block - HISO 10040.1*. Health Information Standards Organisation (HISO).

HISO (**2012**b). *Health Information Exchange Content Model Architecture Building Block - HISO 10040.2*. Health Information Standards Organisation (HISO).

HISO (**2013**). *CDA Templates for Medications and Adverse Reactions vPC*. Health Information Standards Organisation (HISO).

Horstmann, C. (**2008**). *Java Concepts*. John Wiley and Sons, Inc, 111 River Street, Hoboken, NJ, 5th ed.

Hughes, R. and Huby, M. (**2002**). The application of vignettes in social and nursing research. *Journal of Advanced Nursing*, 37: 382–386.

Institute of Medicine (1999, December). Preventing Death And Injury From Medical Errors Requires Dramatic, System-Wide Changes. Internet. Retrieved 2013.12.12, from `http://www.sciencedaily.com/releases/1999/12/991201072044.htm`.

ISO (**2005**). *ISO 20514 Health informatics Electronic health record Definition, scope and context*. International Organization for Standardization.

ISO (**2008**a). *ISO 13606 Health informatics Electronic health record communication*. International Organization for Standardization.

ISO (**2008**b). *ISO/TC DIS 18308 Health informatics Requirements for an electronic health record architecture*. International Organization for Standardization, Geneva, Switzerland.

Jang, Y., Ishii, A. T., and Wang, R. Y. (**1995**). A qualitative approach to automatic data quality judgement. *Journal of Organisational Computing*, 5(2): 101–121.

Jarke, M. and Vassiliou, Y. (**1997**). Data Warehouse Quality: A Review of the DWQ Project. In D. M. Strong and B. K. Kahn, Eds., *Proceedings of 2nd Conference on Information Quality,* pp. 299–313.

Juran, J. M. and Godfrey, A. B. (**1999**). *Juran's Quality Handbook*. McGraw Hill, 5th ed.

Kerr, K. (**2006**). *The Institutionalisation of Data Quality in the New Zealand Health Sector*. Ph.D. thesis, University of Auckland.

Kerr, K., Norris, T., and Stockdale, R. (**2007**). Data Quality Information and Decision Making : A Healthcare Case Study. In *Data Quality in Healthcare, 18th Australasian Conference on Information Systems, 5-7 Dec 2007, Toowoomba,* pp. 1017 – 1026. Retrieved 2013.11.15, from `http://aisel.aisnet.org/acis2007/98/`.

Kirk, J. (**1967**). Standardisation of medical records: A means to an end. *Irish Journal of Medical Science (1926-1967)*, 42: 485–490. Retrieved 2012.06.28, from `http://dx.doi.org/10.1007/BF02954157`. 10.1007/BF02954157.

Klein, A. and Lehner, W. (**2009**). Representing Data Quality in Sensor Data Streaming Environments. *J. Data and Information Quality*, 1: 10:1–10:28. doi:http://doi.acm.org/10.1145/1577840.1577845.

Klein, B. D. and Rossin, D. F. (**1999**). Data Errors in Neural Network and Linear Regression Models: An Experimental Comparison. *Data Quality*, 5(1). Retrieved 2011.11.17, from `http://www.dataquality.com/999KR.htm`.

Lazarou, J., Pomeranz, B. H., and Corey, P. N. (**1998**). Incidence of adverse drug reactions in hospitalized patients. *JAMA: the journal of the American Medical Association*, 279(15): 1200–1205.

Lee, Y. W., Strong, D. M., Kahn, B. K., and Wang, R. Y. (**2002**). AIMQ: a methodology for information quality assessment. *Information and Management*, 40: 133–146.

Leitheiser, R. L. (**2001**). Data Quality in Health Care Data Warehouse Environments. In *34th Hawaii International Conference on System Sciences,* 34th Hawaii International Conference on System Sciences. IEEE.

Loshin, D. (**2001**). *Enterprise Knowledge management: The Data Quality Approach.* Boston: Morgan Kaufmann.

Luhmann, N. (**2000**). *Familiarity, Confidence, Trust: Problems and Alternatives*, Chapter 6, pp. 94–107. Departmant of Sociology, University of Oxford.

Madnick, S. E., Wang, R. Y., Lee, Y. W., and Zhu, H. (**2009**). Overview and Framework for Data and Information Quality Research. *J. Data and Information Quality*, 1: 2:1–2:22. doi:http://doi.acm.org/10.1145/1515693.1516680.

Marks, D. F. and Yardley, L. (**2004**). *Research methods for clinical and health psychology.* Sage Publications Ltd.

Massachusetts Institute of Technology (**2011**). Massachusetts Institute of Technology Total Data Quality Management Program. Retrieved 2013.06.19, from `http://mitiq.mit.edu/toTDQM.aspx`.

Mayer, R. C., Davis, J. H., and Schoorman, F. D. (**1995**). An Integrative Model of Organizational Trust. *The Academy of Management Review*, 20(3): 709–734.

MedTech (**2012**). Medtech statement about the Patients First PMS Report. Retrieved 2013.07.17, from `http://www.medtechglobal.com/news/398/51/Medtech-statement-about-the-Patients-First-PMS-Report.html`.

Mettler, T., Rohner, P., and Baacke, L. (**2008**). Improving Data Quality Of Health Information Systems–A Holistic Design-Oriented Approach. doi:10.1.1.193.3977.

Miller, W. L. and Crabtree, B. F., Eds. (**1999**). *Doing qualitative research*. Sage, Thousand Oaks, California, 2 ed.

Moller, J. E. and Vosegaard, H. (**2008**). Experiences with Electronic Health Records. *IT Professional*, 10: 19–23.

National Instruments (**2009**). Six Good Reasons For Prototyping. Retrieved 2013.05.23, from `http://www.ni.com/white-paper/8752/en`.

Naumann, F. and Rolker, C. (**2000**). Assessment Methods for Information Quality Criteria. In *Fifth Conference on Information Quality (IQ 2000),* pp. 148–162.

New Zealand Medical Association (**2013**). Code of Ethics. Internet. Retrieved 2014.02.04, from `http://www.nzma.org.nz/publications/other-nzma-publications/code-of-ethics`.

Nicolaou, A. I. and McKnight, D. H. (**2006**). Perceived Information Quality in Data Exchanges: Effects on Risk, Trust, and Intention to Use. *Information Systems Research*, 17(4): 332–351.

Noumeir, R. (**2012**). Requirements for Interoperability in Healthcare Information Systems. *Journal of Healthcare Engineering*, 3(2): 323–346. doi:10.1260/2040-2295.3.2.323.

NZ National Health IT Board (**2010**). *National Health IT Plan (Sep. 2010)*. National Health IT Board. Retrieved 2013.06.18, from `http://ithealthboard.health.nz/national-health-it-plan`.

NZ National Health IT Board (**2013**). *National Health IT Plan Update 2013/14*. Ministry of Health, Wellington. Retrieved 2013.12.24, from `http://www.ithealthboard.health.nz`.

OECD (**2009**). OECD Economic Surveys: New Zealand 2009. p. 148. doi:10.1787/eco_surveys-nzl-2009-en.

openEHR Foundation (**2007**). *openEHR Architecture: Architecture Overview*. openEHR Foundation. Retrieved 2011.12.21, from `http://openehr.org/ckm/`.

openEHR Foundation (**2013**a). openEHR - About Archtype Definition Language 1.5. Internet. Retrieved 2013.11.05, from `http://www.openehr.org/downloads/ADLworkbench/learning_about`.

openEHR Foundation (**2013**b). What is openEHR. Internet. Retrieved 2013.07.11, from `http://www.openehr.org/what_is_openehr`.

Orfanidis, L., Bamidis, P. D., and Eaglestone, B. (**2004**). Data quality issues in electronic health records: an adaptation framework for the Greek health system. *Health Informatics Journal*, 10(1): 23–36. doi:10.1177/146045804040665.

Parkin, D. M. and Bray, F. (**2009**). Evaluation of data quality in the cancer registry: Principles and methods Part II. Completeness. *European Journal of Cancer*, 45(5): 756 – 764. doi:DOI: 10.1016/j.ejca.2008.11.033.

Parssian, A. (**2006**). Managerial decision support with knowledge of accuracy and completeness of the relational aggregate functions. *Decision Support Systems*, 42(3): 1494 – 1502. doi:10.1016/j.dss.2005.12.005.

Peabody, J. W., Luck, J., Glassman, P., Jain, S., Hansen, J., Spell, M., and Lee, M. (**2004**a). Measuring the quality of physician practice by using clinical vignettes: a prospective validation study. *Annals of Internal Medicine*, 141(10): 771–780. doi:10.7326/0003-4819-141-10-200411160-00008.

Peabody, J. W., Tozija, F., Munoz, J. A., Nordyke, R. J., and Luck, J. (**2004**b). Using Vignettes to Compare the Quality of Clinical Care Variation in Economically Divergent Countries. *Health Services Research*, 36: 1951–1970.

Persell, S. D., Dunne, A. P., Lloyd-Jones, D. M., and Baker, D. W. (**2009**). Electronic health record-based cardiac risk assessment and identification of unmet preventive needs. *Medical care*, 47(4): 418–424.

Pipino, L. L., Lee, Y. W., and Wang, R. Y. (**2002**). Data quality assessment. *Commun. ACM*, 45: 211–218. doi:http://doi.acm.org/10.1145/505248.506010.

Pirzada, A. A. and McDonald, C. (**2004**). Establishing Trust In Pure Ad-hoc Networks. In *ACSC '04 Proceedings of the 27th Australasian conference on Computer science,* Volume 26, pp. 47–54.

Pitman, A. (**2012**). *Phronesis As Professional Knowledge: Practical Wisdom In The Professions*, Chapter Profesionalism And Professionalisation: Hostile Ground For Growing Phronesis, pp. 131 – 146. Springer.

Poeppelmann, D. and Schultewolter, C. (**2012**). Towards a Data Quality Framework for Decision Support in a Multidimensional Context. *International Journal of Business Intelligence Research*, 3(1): 17–29.

Powell, J., Fitton, R., and Fitton, C. (**2006**). Sharing electronic health records: the patient view. *Informatics in primary care*, 14(1): 55–57.

Prat, N. and Madnick, S. (**2008**). Measuring Data Believability: a Provenance Approach. In *Proceedings of the 41st Hawaii International Conference on System Sciences,* .

Pressman, R. S. (**2005**). *Software Engineering: A Practitioner's Approach.* McGraw-Hill, 1221 Avenue of the Americas, New York, NY, sixth ed.

Price, R., Neiger, D., and Shanks, G. (**2008**). Developing a Measurement Instrument for Subjective Aspects of Information Quality. *Communications of the Association for Information Systems*, 22(1).

Price, R. and Shanks, G. (**2005**). A semiotic information quality framework: development and comparative analysis. *Journal of Information Technology*, 20: 88–102. doi:10.1057/palgrave.jit.2000038.

Price, R. and Shanks, G. (**2008**). Data Quality and Decision Making. In *Handbook on Decision Support Systems 1,* International Handbooks on Information Systems, pp. 65–82. Springer Berlin Heidelberg. doi:10.1007/978-3-540-48713-5_4.

Price, R. and Shanks, G. (**2010**). DQ Tags and Decision-Making. In *Proceedings of the 43rd Hawaii International Conference on System Sciences,* .

Price, R. and Shanks, G. (**2011**). The Impact of Data Quality Tags on Decision-Making Outcomes and Process. *Journal of the Association for Information Systems*, 12(4): Art. 1. Retrieved 2011.08.30, from `http://aisel.aisnet.org/jais/vol12/iss4/1`.

Pringle, M., Wilson, T., and Grol, R. (**2002**). Measuring goodness in individuals and healthcare systems. *British Medical Journal*, 325: 704–707.

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

Raghunathan, S. (**1999**). Impact of information quality and decision-maker quality on decision quality: a theoretical model and simulation analysis. *Decis. Support Syst.*, 26(4): 275–286. doi:10.1016/S0167-9236(99)00060-3.

Ranjan, D. (2013, March). Introduction to Continuity of Care Record (CCR). Internet. Retrieved 2014.01.23, from `http://www.codeproject.com/Articles/564505/Introduction-to-Continuity-of-Care-Record-CCR`.

Redman, T. (**2001**). *Data Quality: The Field Guide*. Digital Press.

Redman, T. C. (**1997**). *Data Quality for the Information Age*. Artech House, Inc., Norwood, MA, USA, 1st ed.

Saigh, O., Triola, M. M., and Link, R. N. (**2006**). Brief report: failure of an electronic medical record tool to improve pain assessment documentation. *Journal of general internal medicine*, 21(2): 185–188.

Scannapieco, M., Virgillito, A., Marchetti, C., Mecella, M., and Baldoni, R. (**2004**). The DaQuinCIS architecture: a platform for exchanging and improving data quality in cooperative information systems. *Inf. Syst.*, 29(7): 551–582. doi:10.1016/j.is.2003.12.004.

Shankaranarayanan, G. and Cai, Y. (**2006**). Supporting data quality management in decision-making. *Decision Support Systems*, 42: 302–317.

Shankaranarayanan, G., Even, A., and Watts, S. (**2006**). The Role of Process Metadata and Data Quality Perceptions in Decision Making: An Empirical Framework and Investigation. *Journal of Information Technology Management*, 17(1).

Shanks, G. (**2001**). The Impact of Data Quality Tagging on Decision Outcomes. In *ACIS 2001 Proceedings,* p. Paper 72.. Retrieved 2011.08.30, from `http://aisel.aisnet.org/acis2001/72`.

Skaner, Y., Bring, J., and Strender, L.-E. (**2004**). Selecting Representative Case Vignettes for Clinical Judgement Studies: Examples from Two Heart Failure Studies. *Quality & Quantity*, 38: 621–635.

Strong, D. M., Lee, Y. W., and Wang, R. Y. (**1997**). Data quality in context. *Commun. ACM*, 40: 103–110. doi:http://doi.acm.org/10.1145/253769.253804.

Stvilia, B., Gasser, L., Twidale, M. B., and Smith, L. C. (**2007**). A framework for information quality assessment. *Journal of the American Society for Information Science and Technology*, 58(12): 1720–1733. doi:10.1002/asi.20652.

Tang, P. C., Ralston, M., Arrigotti, M. F., Qureshi, L., and Graham, J. (**2007**). Comparison of methodologies for calculating quality measures based on administrative data versus clinical data from an electronic health record system: implications for performance measures. *Journal of the American Medical Informatics Association*, 14(1): 10–15. doi:10.1197/jamia.M2198.

Tate, A. R., Williams, T., Puri, S., Beloff, N., and Van Staa, T. (**2011**). Developing quality scores for electronic health records for clinical research: a study using the general practice research database. In *Proceedings of the first international workshop on Managing interoperability and complexity in health systems,* MIXHS '11, pp. 35–42. ACM, New York, NY, USA. doi:10.1145/2064747.2064756.

Taylor, G. W. and Ussher, J. M. (**2001**). Making sense of S&M: A discourse analytic account. *Sexualities*, 4(3): 293–314. doi:10.1177/136346001004003002.

van Deursen, T., Koster, P., and Petkovic, M. (**2008**). Hedaquin: A Reputation-based Health Data Quality Indicator. *Electronic Notes in Theoretical Computer Science*, 197(2): 159 – 167. doi:10.1016/j.entcs.2007.12.025.

Vassiliadis, P., Bouzeghoub, M., and Quix, C. (**2000**). Towards Quality-oriented data warehouse useage and evolution. *Information Systems*, 25: 89–115.

Wagner, M. M. and Hogan, W. R. (**1996**). The accuracy of medication data in an outpatient electronic medical record. *Journal of the American Medical Informatics Association*, 3(3): 234–244. doi:10.1136/jamia.1996.96310637.

Wand, Y. and Wang, R. Y. (**1996**). Anchoring Data Quality Dimensions in Ontological Foundations. *Communications of the ACM*, 39(11): 86–95.

Wang, R., Allen, T., Harris, W., and Madnick, S. (**2003**). An Information Product Approach For Total Information Awareness. Technical report, MIT Sloan School of Management. Retrieved 2011.08.30, from `http://hdl.handle.net/1721.1/1823`.

Wang, R., Kon, H., and Madnick, S. (**1993**). Data quality requirements analysis and modeling. In *Data Engineering, 1993. Proceedings. Ninth International Conference on,* pp. 670 –677. doi:10.1109/ICDE.1993.344012.

Wang, R. Y. (**1998**). A product perspective on total data quality management. *Commun. ACM*, 41: 58–65. doi:http://doi.acm.org/10.1145/269012.269022.

Wang, R. Y. and Strong, D. M. (**1996**). Beyond accuracy: what data quality means to data consumers. *J. Manage. Inf. Syst.*, 12: 5–33. Retrieved 2011.08.30, from `http://dl.acm.org/citation.cfm?id=1189570.1189572`.

Watts, S., Shankaranarayanan, G., and Even, A. (**2009**). Data quality assessment in context: A cognitive perspective. *Decision Support Systems*, 48(1): 202 – 211. doi:10.1016/j.dss.2009.07.012. ¡ce:title¿Information product markets¡/ce:title¿.

Wengraf, T. (**2001**). *Qualitative research interviewing: Biographic narrative and semi-structured methods*. Sage.

Wilson, P. (**2002**). How to find the good and avoid the bad or ugly: a short guide to tools for rating quality of health information on the internet. *BMJ*, 324(7337): 598. doi:10.1136/bmj.324.7337.598.

World Wide Web Consortium (W3C) (**2013**). W3C XML Schema Definition Language (XSD) 1.1 Part 1: Structures. Internet. Retrieved 2013.12.03, from `http://www.w3.org/TR/2012/REC-xmlschema11-1-20120405/structures.html#key-anyType`.

Woudstra, L. and van den Hooff, B. (**2008**). Inside the source selection process: Selection criteria for human information sources. *Information Processing and Management*, 44: 1267–1278. doi:10.1016/j.ipm.2007.07.004.

Appendix A

Ethics Approval

This appendix contains the ethics approval letter received from the Central Regional Ethics Committee.

A.1 Ethics Approval Letter

Dear Stephen Lean

Ethics ref: CEN/09/57/EXP
Study title: A Study to Examine Factors for Confidence in Health Information
Investigators: Dr Stephen Lean

The above study has been given ethical approval by the **Central Regional** Ethics Committee.

Approved Documents:
• A Study to Examine Factors for Confidence in Health Information: Interview Information Sheet, Version 3, dated 3 December 2009
• A Study to Examine Factors for Confidence in Health Information: Participant Consent Form, dated 3 December 2009
• A Study to Examine Factors for Confidence in Health Information: Question List

Certification
The Committee is satisfied that this study is not being conducted principally for the benefit of the manufacturer or distributor of the medicine or item in respect of which the trial is being carried out.

Accreditation
The Committee involved in the approval of this study is accredited by the Health Research Council and is constituted and operates in accordance with the Operational Standard for Ethics Committees, April 2006.

Final Report
The study is approved until **30 July 2010**. A final report is required at the end of the study. The report form is available on http://www.ethicscommittees.health.govt.nz and should be forwarded along with a summary of the results. If the study will not be completed as advised, please forward a progress report and an application for extension of ethical approval one month before the above date.

Amendments
It is also a condition of approval that the Committee is advised if the study does not commence, or is altered including all documentation eg advertisements, letters to prospective participants

Please quote the above ethics committee reference number in all correspondence.

The Principal Investigator is responsible for advising any other study sites of approvals and all other correspondence with the Ethics Committee.

It should be noted that Ethics Committee approval does not imply any resource commitment or administrative facilitation by any healthcare provider within whose facility the research is to be carried out. Where applicable, authority for this must be obtained separately from the appropriate manager within the organisation.
We wish you well with your study.

Yours sincerely

Sonia Scott
Administrator
Central Regional Ethics Committee
Email: sonia_scott@moh.govt.nz

Sonia Scott
Ethics Committee Administrator
Ethics Committees
Health & Disability Services Policy Group
Population Health Directorate
Ministry of Health
DDI: 04 8162405
Fax: 04 496 2191

http://www.ethicscommittees.health.govt.nz
Sonia_Scott@moh.govt.nz

Figure A.1: Ethics approval letter.

Appendix B

Materials For Interviews

This appendix presents materials that received ethical approval and were used for interviews with GPs. These were:

- Participant Information Sheet
- Participant Consent Form
- Participant Demographic Information collection form
- Interview Questions List
- Interview Rating Exercise form

B.1 Participant Information Sheet

A Study to Examine Factors for Confidence in Health Information

Interview Information Sheet

Version: 3
Date: 3rd December 2009
Researcher: Stephen Lean
Supervisors: Prof. Hans Guesgen, Dr. Inga Hunter and Dr. Kudakwashe Dube

I am a student at Massey University, Palmerston North, studying toward a PhD in Computer Science. I am conducting research into the use of health information that is being used outside of the context in which it was gathered. The City Doctors organisation has suggested that you would be a good candidate to participate in my study. Prof. Hans Guesgen, Chair of Computer Science, Massey University, Dr. Inga Hunter, a member of the Royal College of General Practitioners and senior lecturer in Health Informatics at Massey University and Dr. Kudakwashe Dube, lecturer in Computer Science, are supervising my research.

As part of my research I am conducting interviews with GP's who have a requirement to use patient health information in clinical decision making. I would very much appreciate it if you would assist in this study by agreeing to be interviewed. The interview will be to discover what you would like to know and see about the information such that you have confidence in using it meaningfully and safely in clinical decision making.

To achieve this I will be presenting you with a series of short scenarios and then asking open ended questions about them. This will take approximately 50 minutes, at a place and time of your choosing. This is an opportunity to take part in and provide input for, the development of systems that hope to make better use of electronic health information in clinical environments. Please indicate your agreement to participate by filling in and signing the accompanying participation form and returning it to me using the supplied pre paid envelope. If you agree, I will contact you to make arrangements for the interview.

The raw data will be analysed to draw conclusions. Confidentiality of the data from the interview is assured. The data will only be held by myself, stored at Massey University, with additional access only to my supervisors named above. The data will be destroyed by Massey University's approved disposal methods ten years after the completion of my study.

You will be asked at the end of the interview if you would like a summary of the findings. You may also contact me at a later date, using the details below, for a summary of the findings.

You are under no obligation to accept this invitation. If you decide to participate, you have the right to:

- decline to participate.
- refuse to answer any particular question.
- ask any question of myself or my supervisors about the study at any time during participation.
- provide information on the understanding that your name will not be used unless you give permission to the researcher.
- be given access to a summary of the findings of the study when it is concluded.
- ask for the recorder to be turned off at any time during the interview.

Figure B.1: Participant Information Sheet.

There is no penalty from declining to participate or from withdrawing from the research project at any stage. Please do not hesitate to contact myself or my supervisors if you have any questions regarding this study and thank you for taking the time to read this information. I have included contact details below.

Stephen Lean
School of Engineering and Advanced Technology
Massey University
Private Bag 11222
Palmerston North
Phone: 06 350 5799 extn 2625
Email: s.lean@clear.net.nz

Prof. Hans Guesgen
School of Engineering and Advanced Technology
Massey University
Private Bag 11222
Palmerston North
06 3505799 extn 7364
h.w.guesgen@massey.ac.nz

Dr. Inga Hunter
Department of Management
Massey University
Private Bag 11222
Palmerston North
06 3505799 extn 2686
i.hunter@massey.ac.nz

Dr. Kudakwashe Dube
School of Engineering and Advanced Technology
Massey University
Private Bag 11222
Palmerston North
06 3505799 extn 2456
k.dube@massey.ac.nz

This project has been reviewed and approved by the Central Region Ethics Committee. If you have any questions or concerns about your rights as a participant in this research study, you can contact an independent health and disability advocate. This is a free service provided under the Health and Disability Commissioner Act.

Telephone: (NZ) wide): 0800 555 050
Free Fax (NZ wide): 0800 2787 7678 (0800 2 SUPPORT)
Email (NZ wide): advocacy@hdc.org.nz

Figure B.2: Participant Information Sheet.

B.2 Participant Consent Form

A Study to Examine Factors for Confidence in Health Information

Participant Consent Form

Version: 2
Date: 3rd December 2009
Researcher: Stephen Lean

- I have read the Information Sheet for the proposed study and understand the details for the proposed study.
- I have had the opportunity to consider the information, ask questions and have had these answered to my satisfaction, and I understand that I may ask further questions at any time.
- I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason and without any penalty.
- I agree to the interview being sound recorded.
- I agree to participate in this study under the conditions set out in the Information Sheet.

Signature: _____ Date: _____

Full Name (please print): _____

Please circle the appropriate:
I would like to review the transcript of my interview. Yes / No
I would like a copy of the summary of findings for this study. Yes / No

Date: 3rd December, 2009 Consent Form Ver: 2

Figure B.3: Participant Consent Form.

B.3 Participant Demographic Information Collection Form

Year completed training: _____

Age Bracket you fall into (please circle one):

20 – 29
30 – 39
40 – 49
50 – 59
60 – 69

How many years in practice: _____

Vocationally registered? (please circle one) Yes / No

How comfortable are you in using Information Computing Technology (ICT) in your work (please circle one, where 10 is extremely comfortable and 1 is not at all comfortable):

1 2 3 4 5 6 7 8 9 10

How competent do you consider yourself to be in using ICT in your work (please circle one):

1 2 3 4 5 6 7 8 9 10

Figure B.4: Participant Demographic Information Collection Form.

B.4 Interview Questions List

A Study to Examine Factors for Confidence in Health Information

These are the questions to be asked of participants in the semi-structured interviews. Interviews will take place at a place and time of the participant's choosing and will be up to 50 minutes in duration. Interviews will be recorded and later transcribed. This study is to be piloted on 2 to 3 members from the target group of participants (see also the brief description for the study, Section 3 of the application form).

Vignette 1

A patient presents with symptoms indicating an ear infection. His medical record shows an allergy to penicillin.

V1.Q1

How do you use this information when deciding on a potential treatment?

Variation

When questioned regarding the allergic reaction, he describes symptoms that, while a reaction, make you think that it was not an allergic reaction i.e. more likely a side affect.

V1.Q2

How do you consider this information?

Variation

A patient presents with symptoms indicating an ear infection. Her medical record shows that she was prescribed Amoxicillin 3 years ago and an allergic reaction was recorded for it. The record shows she was again prescribed Amoxicillin, 1 year ago with no record of an allergic reaction. The patient is unable to clarify.

V1.Q3

What do you think when you see information that seems to imply that either one entry or another is not correct?

V1.Q4

How does this affect your use of the information?

V1.Q5

If contradictory entries are found, does it make you trust the rest of the record less?

Variation

Patient presents with a badly infected finger from a cut. Their medical record states they have no known drug allergies to any medications.

Figure B.5: Interview Questions List.

V1.Q6

How is the statement of "no known drug allergies" used when deciding on treatment?
When you see no known drug allergies what do you think?

V1.Q7

What is it about recorded information on a patient's drug allergy status that would provide you with confidence that the information can be safely relied upon and used in clinical decisions?

Vignette 2

Patient presents describing symptoms indicating a level of depression. An entry on the patient's records dated 7 years previously has a diagnosis of schizophrenia. On questioning the patient they indicate they were unaware of this diagnosis and that this may relate to an incident at Accident and Emergency where they were seen after having an adverse reaction to medication at the time.

V2.Q1

From a perspective of the entry on their medical history, what would you want to know about the diagnosis of Schizophrenia that would help in the current consultation?

V2.Q2

How would knowing that the person who made the diagnosis of schizophrenia had a qualification relevant to diagnosing mental illness affect your use of the recorded entry?

Variation

A patient presents to you who has a previous diagnosis of mild depression. This diagnosis was made 3 months ago. At that time they were prescribed an antidepressant. The patient indicates they are feeling no better. The recorded dosage is high for initial prescription of this drug for depression, especially as the patient says they feel no different/better. The patient is unable to clarify what dosage they were taking.

V2.Q3

How do you treat the recorded dosage information?

V2.Q4

Can you tell me of an instance when you felt dosage information was not correct?

V2.Q5

How did you resolve this?

V2.Q6

What accompanying information would have helped you in that situation?

Figure B.6: Interview Questions List.

Vignette 3

A patient presents for a regular checkup and medications check. When questioned on one of the medications he is taking he states he is not taking the medication that is recorded in his record. You want to write a refill script for what he is taking as he has stated that it seems to be working and he is happy to continue taking it as it has minimal side effects for him.

V3.Q1

Can you tell me of an instance where you have had a similar encounter?

V3.Q2

How did you reconcile what medication the patient was actually taking?

V3.Q3

Was this different from what had been prescribed?

V3.Q4

What information, which could have potentially been recorded in the patient's record, would have helped you in this situation

Variation

You are seeing the patient after the previously described encounter. The last encounter was now not with you but was made elsewhere. At this last visit the medications entry was changed from one prescription to another for the same medical condition. You are looking at the changed medical record.

V3.Q5

How would having an accompanying narrative of the reason for the change to the patient's medication, affect your belief in, or use of, the information?

V3.Q6

How would past interaction with the person who had made the change to the patient's medication affect your belief in, or use of, the information?

General

G.Q1

When thinking about believing recorded medical information, what background information on the entries would persuade you that using this information would be safe?

Figure B.7: Interview Questions List.

G.Q2
What thoughts come to mind when thinking about how happy you feel in trusting that the information you wish to rely in is ok for the purpose you want to use it for.

G.Q3
When recording information into a patient's record yourself, what are some details, currently not collected about what you are recording, that you feel would make it easier for the next user of that information in their decision making?

G.Q4
Without identifying specific people involved, have you had experience yourself where you have wanted to use information from a patient's medical record, but have felt unable to do so reliably?

G.Q5
If so, what information did you want to use?

G.Q6
Why did you feel uncomfortable about relying on that information?

G.Q7
What extra information, had that been available to you, would have made you feel confident enough to use the information safely?

G.Q8
How does knowing what organisation produced information affect your use of that information, if at all?

Figure B.8: Interview Questions List.

B.5 Interview Rating Exercise Form

Rating Scale

Rating scale is from 1 to 4, where 4 is very important, 3 is somewhat important, 2 is slightly important and 1 is not important.

Attribute	Rating
Knowing the qualifications of the clinician involved in healthcare event	
Knowing the qualifications of the recorder of the information	
The years of experience of the clinician	
The years of experience of the recorder of the information	
The position held by the clinician	
The position held by the recorder of the information	
Attestation of entries recorded	
Qualifications, experience and position of person attesting to information	
Where the information was captured	
Who has been responsible for storing the information	
Knowing information was produced at an organisation known for its expertise in that field	
Knowing the information was produced at an organisation that has a good reputation	
Age of the data	
Indication of contradictory entries	
Knowing length of time from information captured to information recorded	
Knowing testing methodology for test results	
Positive/negative personal experience with provider	
Contextual narrative for event for entry	
Reason for change of entry e.g. change of diagnosis or change in prescribed medication	
Indication of where to go for further information	
Internal consistency/cohesion of record	

Figure B.9: Interview Rating Exercise.

Appendix C

XML Schema for Quality Criteria Model: QCSchema.xsd

C.1 Schema for Quality Criteria Model

Listing C.1: XML schema formal definition for Quality Criteria Model.

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema targetNamespace="urn:qc-models:QC"
  elementFormDefault="qualified"
  xmlns="urn:qc-models:QC"
  xmlns:ccr="urn:astm-org:CCR"
  xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:import namespace="urn:astm-org:CCR" schemaLocation="CCR.xsd"/>

  <!-- Quality Criteria Model. -->

  <xs:element name="quality-criteria">
  <xs:complexType>
      <xs:sequence>
        <xs:element name="label" type="xs:string" minOccurs="1" maxOccurs="1"/>
        <xs:element name="qc-id" type="xs:string" minOccurs="1" maxOccurs="1"/>
        <xs:element name="date" type="ccr:DateTimeType" minOccurs="1" maxOccurs="1"/>
        <xs:element name="evaluation-type" type="eval-type" minOccurs="0" maxOccurs="1"/>

        <xs:element name="status">
          <xs:complexType>
            <xs:sequence>
              <xs:element name="age" type="ccr:DurationType" minOccurs="0" maxOccurs="1"/>
              <xs:element name="active" type="xs:boolean" minOccurs="1" maxOccurs="1"/>
            </xs:sequence>
          </xs:complexType>
        </xs:element>
```

```
    <xs:element name="entry" minOccurs="1" maxOccurs="1">
      <xs:complexType>
        <xs:sequence>
          <xs:element ref="ccr:InternalCCRLink" minOccurs="1" maxOccurs="1"></xs:element>
          <xs:element ref="quality-criteria" minOccurs="0" maxOccurs="unbounded"/>
        </xs:sequence>
      </xs:complexType>
    </xs:element>

    <xs:element name="actors" minOccurs="1" maxOccurs="1">
      <xs:complexType>
        <xs:sequence>
          <xs:element name="actor" type="actor-type" minOccurs="1" maxOccurs="unbounded"/>
        </xs:sequence>
      </xs:complexType>
    </xs:element>

    <xs:element name="document" minOccurs="1" maxOccurs="1">
      <xs:complexType>
        <xs:sequence>
          <xs:element name="type" type="xs:string" minOccurs="1" maxOccurs="1"/>
          <xs:element ref="ccr:InternalCCRLink" minOccurs="1" maxOccurs="unbounded"/>
          <xs:element ref="quality-criteria" minOccurs="0" maxOccurs="unbounded"/>
        </xs:sequence>
      </xs:complexType>
    </xs:element>

    <xs:element name="qualifications" minOccurs="0" maxOccurs="1">
      <xs:complexType>
        <xs:sequence>
          <xs:element name="qualification" type="qc-data" minOccurs="1" maxOccurs="unbounded"/>
        </xs:sequence>
      </xs:complexType>
    </xs:element>

    <xs:element name="role" type="qc-data" minOccurs="0" maxOccurs="1"/>
    <xs:element name="narrative" type="qc-data" minOccurs="0" maxOccurs="1"/>
    <xs:element name="experience" type="qc-data" minOccurs="0" maxOccurs="1"/>
    <xs:element name="expertise" type="qc-data" minOccurs="0" maxOccurs="1"/>
    <xs:element name="reputation" type="qc-data" minOccurs="0" maxOccurs="1"/>

    <xs:element name="further-information" minOccurs="0" maxOccurs="1">
      <xs:complexType>
        <xs:sequence>
          <xs:element name="further-info" type="qc-data" minOccurs="1" maxOccurs="unbounded"/>
        </xs:sequence>
      </xs:complexType>
    </xs:element>

    <xs:element name="attestation" type="qc-data" minOccurs="0" maxOccurs="1"/>
    <xs:element name="data-age" type="ccr:DurationType" minOccurs="0" maxOccurs="1"/>
    <xs:element name="info-captured" type="qc-data" minOccurs="0" maxOccurs="1"/>
  </xs:sequence>
</xs:complexType>
```

```
  </xs:element>

  <xs:complexType name="actor-type">
    <xs:sequence>
      <xs:element name="link" type="ccr:ActorReferenceType"/>
      <xs:element name="type" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element ref="quality-criteria" minOccurs="0" maxOccurs="unbounded"/>
    </xs:sequence>
  </xs:complexType>

  <xs:complexType name="qc-data">
    <xs:choice>
      <xs:sequence>
        <xs:element ref="ccr:InternalCCRLink"/>
        <xs:element name="description" type="xs:string"/>
        <xs:element name="true" type="xs:boolean"/>
      </xs:sequence>
    </xs:choice>
  </xs:complexType>

  <xs:simpleType name="eval-type">
    <xs:restriction base="xs:string">
      <xs:enumeration value="objective"/>
      <xs:enumeration value="subjective"/>
    </xs:restriction>
  </xs:simpleType>

  <xs:element name="rel-type">
    <xs:simpleType>
      <xs:restriction base="xs:string">
        <xs:enumeration value="inf-inf"/>
        <xs:enumeration value="actor-inf"/>
        <xs:enumeration value="actor-actor"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:element>

</xs:schema>
```

C.2 QCM Instance Example

Listing C.2: QCM instance example from Chapter 5, Figure 5.19.

```
<?xml version="1.0" encoding="utf-8" ?>
<qc:quality-criteria
  xmlns:qc="urn:qc-models:QC"
  xmlns:ccr = "urn:astm-org:CCR"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="urn:qc-models:QC QCSchema.xsd urn:astm-org:CCR CCR.xsd">

  <qc:label>C4</qc:label>
  <qc:qc-id>QCM1</qc:qc-id>
  <qc:date></qc:date>
  <qc:evaluation-type>objective</qc:evaluation-type>
  <qc:status>
    <qc:age></qc:age>
    <qc:active>true</qc:active>
  </qc:status>

  <qc:entry>
    <ccr:InternalCCRLink>
      <ccr:LinkID>E01</ccr:LinkID>
    </ccr:InternalCCRLink>

    <qc:quality-criteria>
      <qc:label>C8</qc:label>
      <qc:qc-id>QCM2</qc:qc-id>
      <qc:date></qc:date>
      <qc:evaluation-type>objective</qc:evaluation-type>
      <qc:status>
        <qc:age></qc:age>
        <qc:active>true</qc:active>
      </qc:status>
      <qc:entry>
        <ccr:InternalCCRLink>
          <ccr:LinkID>E01</ccr:LinkID>
        </ccr:InternalCCRLink>
      </qc:entry>
      <qc:document>
        <qc:type>CCR</qc:type>
        <ccr:InternalCCRLink>
          <ccr:LinkID>D01</ccr:LinkID>
        </ccr:InternalCCRLink>
      </qc:document>
      <qc:narrative>
        <qc:description>contextual narrative</qc:description>
      </qc:narrative>
    </qc:quality-criteria>

    <qc:quality-criteria>
      <qc:label>C12</qc:label>
      <qc:qc-id>QCM4</qc:qc-id>
      <qc:date></qc:date>
```

```
    <qc:evaluation-type>objective</qc:evaluation-type>
    <qc:status>
      <qc:age></qc:age>
      <qc:active>true</qc:active>
    </qc:status>
    <qc:entry>
      <ccr:InternalCCRLink>
        <ccr:LinkID>E01</ccr:LinkID>
      </ccr:InternalCCRLink>
    </qc:entry>
    <qc:document>
      <qc:type>CCR</qc:type>
      <ccr:InternalCCRLink>
        <ccr:LinkID>D01</ccr:LinkID>
      </ccr:InternalCCRLink>
    </qc:document>
    <qc:data-age>
      <ccr:Value>1 months</ccr:Value>
    </qc:data-age>
  </qc:quality-criteria>

</qc:entry>
<qc:actors>
  <qc:actor>
    <qc:link>
      <ccr:ActorID>A01</ccr:ActorID>
    </qc:link>
    <qc:type>person</qc:type>

    <qc:quality-criteria>
      <qc:label>C11</qc:label>
      <qc:qc-id>QCM3</qc:qc-id>
      <qc:date></qc:date>
      <qc:evaluation-type>objective</qc:evaluation-type>
      <qc:status>
        <qc:age></qc:age>
        <qc:active>true</qc:active>
      </qc:status>
      <qc:entry>
        <ccr:InternalCCRLink>
          <ccr:LinkID>E01</ccr:LinkID>
        </ccr:InternalCCRLink>
      </qc:entry>
      <qc:document>
        <qc:type>CCR</qc:type>
        <ccr:InternalCCRLink>
          <ccr:LinkID>D01</ccr:LinkID>
        </ccr:InternalCCRLink>
      </qc:document>
      <qc:experience>
        <qc:description>3 years</qc:description>
      </qc:experience>
    </qc:quality-criteria>

  </qc:actor>
```

```
  </qc:actors>

  <qc:document>
    <qc:type>CCR</qc:type>
    <ccr:InternalCCRLink>
      <ccr:LinkID>D01</ccr:LinkID>
    </ccr:InternalCCRLink>
  </qc:document>
  <qc:attestation>
    <qc:isTrue>true</qc:isTrue>
  </qc:attestation>

</qc:quality-criteria>
```