The Institutionalisation of Data Quality in the New Zealand Health Sector

Karolyn Kerr

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in Information Systems
The University of Auckland, 2006
Abstract

This research began a journey towards improved maturity around data quality management in New Zealand health care, where total data quality management is ‘business as usual’, institutionalised into the daily practices of all those who work in health care. The increasingly information intensive nature of health care demands a proactive and strategic approach to data quality to ensure the right information is available to the right person at the right time in the right format, all in consideration of the rights of the patient to have his/her health data protected and used in an ethical way. The work extends and tests principles to establish good practice and overcome practical barriers. This thesis explores the issues that define and control data quality in the national health data collections and the mechanisms and frameworks that can be developed to achieve and sustain good data quality.

The research is interpretive, studying meaning within a social setting. The research provides the structure for learning and potential change through the utilisation of action research. Grounded theory provides the structure for the analysis of qualitative data through inductive coding and constant comparison in the analysis phase of the action research iterative cycle. Participatory observation provided considerable rich data as the researcher was a member of staff within the organisation. Data were also collected at workshops, focus groups, structured meetings and interviews.

The development of a Data Quality Evaluation Framework and a national Data Quality Improvement Strategy provides clear direction for a holistic and ‘whole of health sector’ way of viewing data quality, with the ability for organisations to develop and implement local innovations through locally developed strategies and data quality improvement programmes. The researcher utilised the theory of appreciative enquiry (Fry, 2002) to positively encourage change, and to encourage the utilisation of existing organisational knowledge. Simple rules, such as the TDQM process and the data quality dimensions guided the change, leaving room for innovation. The theory of ‘complex systems of adjustment’ (Champagne, 2002; Stacey, 1993) can be instilled in the organisation to encourage change through the constant interaction of people throughout the organisation.
Acknowledgements

I would like to thank my supervisor, Dr Tony Norris, sincerely, for his initial faith, ongoing patience and expert guidance. Thanks also go to other members of the faculty of the Department of Information Systems and Operations Management at the University of Auckland, who provided valuable feedback throughout the research, through PhD forum.

I would also like to thank sincerely the staff of the Ministry of Health, including:

- Dr Karen Poutasi for allowing me to undertake this research within the organisation;
- the members of the Clinical Analysis Team at NZHIS who worked hard to make change happen;
- the members of the Ministry wide data quality team for their willingness to learn and contribute to the research;
- the management of the Corporate and Information Directorate, in particular, Mr Mike Rillstone, for their ongoing belief in data quality.

I am very grateful to all study participants who actively and enthusiastically participated in this research. In particular, the Chief Information Officers of the District Health Boards and PHOs and those working in data quality in the health sector who gave me considerable time out of their busy schedules. Thanks to the entire Health Informatics community in New Zealand for their participation in surveys, discussions and presentations at the Annual Health Informatics New Zealand Conference.

Additional thanks must go to the participants in the various sessions at many international Data Quality Conferences who provided, through questions and feedback, thought provoking additional research investigations to include in the final thesis.

The researcher acknowledges the administrative research assistance of Robyn Kamira in the development of the Data Quality Evaluation Framework, and Dr Marcus Gebauer in the development of the Principles for Data Quality in National Health Data Collections.

And lastly, thanks to my friends and family, who had the courage to ask ‘how is the thesis coming along?’
# Table of Contents

Abstract...............................................................................................................................................................II
Acknowledgements...............................................................................................................................................III
Table of Contents .............................................................................................................................................. IV
List of Tables .................................................................................................................................................... IX
List of Figures ................................................................................................................................................ XI
List of Abbreviations ......................................................................................................................................... XII
Part A Introduction and Background .............................................................................................................. 1
1 Introduction .................................................................................................................................................... 2
   1.1 Scope of the Research ................................................................................................................................... 2
   1.2 Searching the Data Quality Literature ...................................................................................................... 2
   1.3 Data, Information and Knowledge ........................................................................................................... 3
   1.4 Quality ....................................................................................................................................................... 4
   1.5 Data Quality ............................................................................................................................................... 5
       1.5.1 The Context of Data Quality ................................................................................................................ 6
       1.5.2 Data Stewards, Custodians, Consumers and Collectors ......................................................................... 7
       1.5.3 Data Governance and Ownership ........................................................................................................ 7
       1.5.4 Dimensions of Data Quality ................................................................................................................ 8
       1.5.5 Total Data Quality Management (TDQM) ............................................................................................ 10
       1.5.6 Alternative Approaches to TDQM ..................................................................................................... 13
       1.5.7 The Impact of Poor Data Quality ........................................................................................................ 13
       1.5.8 Assessing and Measuring Data Quality ............................................................................................... 14
       1.5.9 Data Quality Frameworks .................................................................................................................... 15
       1.5.10 Data Quality Improvement Programmes .......................................................................................... 18
       1.5.11 A Strategic View of Data Quality .................................................................................................... 19
   1.6 Present Research ........................................................................................................................................ 20
       1.6.1 Aims of the Research .......................................................................................................................... 20
       1.6.2 Research Questions ........................................................................................................................... 21
   1.7 Chapter Conclusion .................................................................................................................................. 22
2 Data Quality and Health Care ........................................................................................................................ 23
   2.1 What are Health Data? ............................................................................................................................. 23
       2.1.1 Medical Registries and Repositories .................................................................................................. 25
       2.1.2 The Patient Record ............................................................................................................................ 25
       2.1.3 Electronic Health Records ................................................................................................................ 26
       2.1.4 Clinical Coding .................................................................................................................................... 27
       2.1.5 Privacy of Patient Data ....................................................................................................................... 28
       2.1.6 Stewardship of Health Care Data ....................................................................................................... 29
   2.2 Health Informatics ...................................................................................................................................... 30
   2.3 Why Improve Data Quality in Health Care? ............................................................................................. 30
   2.4 Improving Data Quality in Health Care .................................................................................................... 32
       2.4.1 Total Quality Management in Health Care .......................................................................................... 33
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.2.15</td>
<td>Key Concepts to Emerge from the Current State Analysis</td>
<td>131</td>
</tr>
<tr>
<td>7.3</td>
<td>Capability Maturity Assessment of New Zealand Health Care Organisations</td>
<td>132</td>
</tr>
<tr>
<td>7.4</td>
<td>Ministry of Health Data Quality Workshop</td>
<td>133</td>
</tr>
<tr>
<td>7.4.1</td>
<td>Breakout sessions</td>
<td>139</td>
</tr>
<tr>
<td>7.5</td>
<td>Stakeholder Analysis</td>
<td>141</td>
</tr>
<tr>
<td>7.5.1</td>
<td>Ministry of Health Stakeholders</td>
<td>142</td>
</tr>
<tr>
<td>7.5.2</td>
<td>Sector Stakeholders</td>
<td>143</td>
</tr>
<tr>
<td>7.6</td>
<td>Data Quality Improvement Strategy</td>
<td>144</td>
</tr>
<tr>
<td>7.6.1</td>
<td>The Scope of the Strategy</td>
<td>145</td>
</tr>
<tr>
<td>7.6.2</td>
<td>Aims of the Strategy</td>
<td>148</td>
</tr>
<tr>
<td>7.6.3</td>
<td>Roles in Data Quality in the New Zealand Health Sector</td>
<td>149</td>
</tr>
<tr>
<td>7.6.4</td>
<td>TDQM at the Ministry of Health – Institutionalising Data Quality</td>
<td>151</td>
</tr>
<tr>
<td>7.6.5</td>
<td>Defining Data Quality</td>
<td>152</td>
</tr>
<tr>
<td>7.6.6</td>
<td>Measuring Data Quality</td>
<td>153</td>
</tr>
<tr>
<td>7.6.7</td>
<td>Analysing Data Quality</td>
<td>154</td>
</tr>
<tr>
<td>7.6.8</td>
<td>Improving Data Quality</td>
<td>155</td>
</tr>
<tr>
<td>7.6.9</td>
<td>Data Quality Principles for New Collections</td>
<td>156</td>
</tr>
<tr>
<td>7.7</td>
<td>TDQM in the Wider Health Sector - the Proposed Sector Work Programme</td>
<td>158</td>
</tr>
<tr>
<td>7.7.1</td>
<td>Accreditation of Data Suppliers</td>
<td>159</td>
</tr>
<tr>
<td>7.7.2</td>
<td>Taking Part in a Sector Peer Review/Audit Process</td>
<td>159</td>
</tr>
<tr>
<td>7.7.3</td>
<td>Meeting Key Performance Indicators (KPIs)</td>
<td>160</td>
</tr>
<tr>
<td>7.7.4</td>
<td>Implementing an In-house Data Quality Education Programme</td>
<td>160</td>
</tr>
<tr>
<td>7.7.5</td>
<td>Developing and Implementing a Local Data Quality Improvement Strategy</td>
<td>160</td>
</tr>
<tr>
<td>7.7.6</td>
<td>Organising Regular Meetings with a Cross-Organisational Data Quality Group</td>
<td>161</td>
</tr>
<tr>
<td>7.7.7</td>
<td>Alignment with Best Practice Guidelines (when developed)</td>
<td>161</td>
</tr>
<tr>
<td>7.8</td>
<td>Implementation of the Strategy</td>
<td>161</td>
</tr>
<tr>
<td>7.8.1</td>
<td>Phase One</td>
<td>162</td>
</tr>
<tr>
<td>7.8.2</td>
<td>Phase Two</td>
<td>163</td>
</tr>
<tr>
<td>7.8.3</td>
<td>Phase Three</td>
<td>164</td>
</tr>
<tr>
<td>7.8.4</td>
<td>Cost Implications</td>
<td>165</td>
</tr>
<tr>
<td>7.8.5</td>
<td>Monitoring and Evaluation</td>
<td>166</td>
</tr>
<tr>
<td>7.9</td>
<td>Organisational Change Requirements</td>
<td>166</td>
</tr>
<tr>
<td>7.10</td>
<td>Chapter Summary</td>
<td>167</td>
</tr>
<tr>
<td>8</td>
<td>Detailed Discussion and Further Analysis of Results</td>
<td>169</td>
</tr>
<tr>
<td>8.1</td>
<td>The Results in Context</td>
<td>169</td>
</tr>
<tr>
<td>8.2</td>
<td>The Role of the Researcher</td>
<td>170</td>
</tr>
<tr>
<td>8.3</td>
<td>Bringing about Change Through a Paradigm Shift</td>
<td>171</td>
</tr>
<tr>
<td>8.4</td>
<td>The Complex Adaptive System in Health Care</td>
<td>172</td>
</tr>
<tr>
<td>8.5</td>
<td>Data Quality Capability Maturity of the New Zealand Health Care Sector When Compared to</td>
<td>173</td>
</tr>
<tr>
<td></td>
<td>Organisations Outside Health Care</td>
<td></td>
</tr>
<tr>
<td>8.6</td>
<td>The Impact of the Data Quality Evaluation Framework and Dimensions</td>
<td>176</td>
</tr>
<tr>
<td>8.7</td>
<td>The Data Quality Improvement Strategy</td>
<td>178</td>
</tr>
</tbody>
</table>
List of Tables

Table 1: Governance and Custodianship (Ministry of Health 2004) ................................................................................................. 7
Table 2: Data Quality Categories and Dimensions (Wang et al., 1996) ................................................................................................. 9
Table 3: Data Quality Categories and Dimensions (Gendron & D’Onofrio, 2001) ................................................................. 10
Table 5: Examples of Types and Uses of Health Care Data. Adapted from Saba and McCormick (Saba & McCormick, 2001) ............................................................................................................................................ 24
Table 6: Summary of the Development Process of the Ministry of Health Data Quality Evaluation Framework ......................................................................................................................................................................... 62
Table 7: Summary of the Number of Health Informatics New Zealand 2005 Conference Attendees by Organisation ......................................................................................................................................................................... 65
Table 8: Data Quality Management Maturity Grid (English, 1999a) ........................................................................................................ 67
Table 9: Research data collection methods, number of participants and where the participants were sourced. ........................................................................................................................................................................................................ 71
Table 10: Participant Samples used Throughout the Research Process ........................................................................................................ 71
Table 11: Summary of the Findings of the Ministry of Health Information Systems Strategic Plan (Ministry of Health, 2003c) ................................................................................................................................................. 77
Table 12: Criteria Added to Draft Framework Following Review of Ministry of Health ISSP ........................................................................................................................................................................................................ 78
Table 13: Criteria Added to Draft Framework Following Review of Ministry of Health Privacy and Security Requirements ........................................................................................................................................................................................................ 79
Table 14: Modification of the CIHI Coverage Criteria .......................................................................................................................... 87
Table 15: Modification of the CIHI Capture and Collection Criteria ........................................................................................................ 87
Table 16: Modification of the CIHI Unit Non-Response Criteria .............................................................................................................. 88
Table 17: Modification of the CIHI Edit and Imputation Criteria ............................................................................................................. 88
Table 18: Modification of the CIHI Data Currency at the Time of Release Criteria ................................................................................ 89
Table 19: Modification of the CIHI Data Dictionary Standards Criteria .................................................................................................. 89
Table 20: Modification of the CIHI Standardisation Criteria .................................................................................................................. 90
Table 21: Modification of the CIHI Linkage Criteria ................................................................................................................................. 90
Table 22: Modification of the CIHI Equivalency Criteria .......................................................................................................................... 91
Table 23: Modification of the CIHI Historical Comparability Criteria .................................................................................................... 92
Table 24: Modification of the CIHI Accessibility Criteria ......................................................................................................................... 93
Table 25: Modification of the CIHI Documentation Criteria .................................................................................................................. 93
Table 26: Modification of the CIHI Interpretability Criteria .................................................................................................................... 94
Table 27: Modification of the CIHI Value Criteria ................................................................................................................................. 94
Table 28: New Legislation Criteria ............................................................................................................................................................... 95
Table 29: New Disclosure Criteria ............................................................................................................................................................... 95
Table 30: Assessment of the New Zealand Data Quality Evaluation Framework (Eppler & Wittig, 2000) ................................................................. 99
Table 31: User Type and Report Type Provided from the Mortality Data Collection ................................................................................................. 100
Table 32: Findings of the Assessment of the Mortality Collection using the Draft Ministry of Health Data Quality Framework ........................................................................................................................................ 101
Table 33: Applicability of the DQEF with a Hospital Clinical Data Collection .................................................. 105
Table 34: Feedback to Researcher By Data Quality Dimension on the Practical Use of the DQEF in a Clinical Setting ....................................................................................................................................... 105
Table 35: Results of Data Quality Dimension Survey Rankings by Type of Respondent .................................... 115
Table 36: Overall Rankings of Data Quality Dimensions by all Respondents .................................................... 116
Table 37: High Level Grounded Theory Codes and their Definitions ...................................................................... 119
Table 38: The Range of Data Quality Capability of New Zealand Health Care Organisations .......................... 132
Table 39: Importance of Data Quality Characteristics by Role ........................................................................ 135
Table 40: Characteristics by Dimension Rated by Respondents as a Priority for Improvement on the NHI .... 135
Table 41: Perceived Current Performance of Data Quality Work for Each Characteristic of the DQEF ........... 136
Table 42: Impact, Priority and Timing of Proposed Data Quality Improvement Initiatives .............................. 138
Table 43: Feedback from Breakout Groups on Strategy Implementation Requirements .................................. 140
Table 44: Ministry of Health DQIS Stakeholders and their Information Requirements .................................... 142
Table 45: Health Sector DQIS Stakeholders and their Information Requirements ........................................... 143
Table 46: National Health Data Collections Included in the Scope of Data Quality Improvement Strategy ........ 147
Table 47: Roles and Responsibilities for Data Quality in the New Zealand Health Sector ................................ 151
Table 48: Phase One, Projects 1 – 4 of DQIS Implementation ........................................................................... 162
Table 49: Phase Two, Projects 5 - 12 of DQIS Implementation ......................................................................... 164
Table 50: Phase Three, Projects 13 - 15 of DQIS Implementation .................................................................... 165
Table 51: Comparison of Data Quality Maturity in New Zealand Health Care Organisations with International Organisations Outside of Health Care ......................................................................................................................... 174
Table 52: Specification Components for New National Health Data Collections (Ministry of Health, 2004) .... 195
Table 53: Methods to Answer the Research Questions .................................................................................. 209
List of Figures

Figure 1: Components of the TDQM Cycle (Wang, 1998) ................................................................. 11
Figure 2: Structure of the New Zealand Health Sector ........................................................................ 42
Figure 3: Structure of the Ministry of Health ........................................................................................ 43
Figure 4: Linkage Between Health Strategies and Information Strategies (Ministry of Health, 2005) .... 46
Figure 6: The Combination of Action Learning and Grounded Theory .................................................. 50
Figure 7: Action Research Structural Cycle (Canonical Action Research) (Susman & Evered, 1978) ........ 54
Figure 15: Initial Phase of the Action Research Cycle for the Data Quality Evaluation Framework ......... 69
Adapted from Susman and Evered (Susman & Evered, 1978) ................................................................. 69
Figure 16: Second Action Research Cycle Concentrating on Implementation of the Framework .......... 70
Adapted from Susman and Evered (Susman & Evered, 1978) ................................................................. 70
Figure 8: The CIHI Data Quality Framework (Version 1) Four-Level Model (Long & Seko, 2002) .......... 75
Figure 8: Extract of Discussion with Focus Group Participants ............................................................... 85
Figure 10: Strategic Alignment of KPIs (Bauer, 2004) ........................................................................... 108
Figure 11: The New Zealand Ministry of Health Data Quality Evaluation Framework Structure ............. 110
Figure 12: Example of the Breakdown of Initial High Level Grounded Theory Codes ......................... 120
Figure 13: Roles as Defined by Workshop Participants ........................................................................ 134
Figure 13: Scope of the Data Quality Improvement Strategy ................................................................. 146
Figure 17: Grounded Theory Codes Identifying the Relationship between the Presence of an Organisation Wide Data Quality Team and More Mature Capability in Data Quality Management ....................... 180
Figure 18: New Zealand’s Distributed Electronic Health Records (EHR) Model (Ministry of Health, 2005). 193
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>Action Research</td>
</tr>
<tr>
<td>CAS</td>
<td>Complex Adaptive System</td>
</tr>
<tr>
<td>CIHI</td>
<td>Canadian Institute for Health Information</td>
</tr>
<tr>
<td>CIO</td>
<td>Chief Information Officer</td>
</tr>
<tr>
<td>DHB</td>
<td>District Health Board</td>
</tr>
<tr>
<td>DQEF</td>
<td>Data Quality Evaluation Framework</td>
</tr>
<tr>
<td>DQIS</td>
<td>Data Quality Improvement Strategy</td>
</tr>
<tr>
<td>EHR</td>
<td>Electronic Health Record</td>
</tr>
<tr>
<td>HealthPAC</td>
<td>Health Payments, Agreements and Compliance</td>
</tr>
<tr>
<td>HIPPA</td>
<td>Health Information Privacy and Portability Act (USA)</td>
</tr>
<tr>
<td>HISAC</td>
<td>Health Information Strategy Action Committee</td>
</tr>
<tr>
<td>HISO</td>
<td>Health Information Standards Organisation</td>
</tr>
<tr>
<td>HIS – NZ</td>
<td>Health Information Strategy for New Zealand</td>
</tr>
<tr>
<td>IM</td>
<td>Information Management</td>
</tr>
<tr>
<td>ISSP</td>
<td>Information Systems Strategic Plan</td>
</tr>
<tr>
<td>ICD 10</td>
<td>International Classification of Diseases Version 10</td>
</tr>
<tr>
<td>IT</td>
<td>Information Technology</td>
</tr>
<tr>
<td>KPI</td>
<td>Key Performance Indicator</td>
</tr>
<tr>
<td>MIT</td>
<td>Massachusetts Institute of Technology</td>
</tr>
<tr>
<td>MDQT</td>
<td>Ministry of Health Data Quality Team</td>
</tr>
<tr>
<td>NHS</td>
<td>National Health Service (United Kingdom)</td>
</tr>
<tr>
<td>NZHIS</td>
<td>New Zealand Health Information Service</td>
</tr>
<tr>
<td>OECD</td>
<td>Organisation for Economic Co-operation and Development</td>
</tr>
<tr>
<td>PHO</td>
<td>Primary Health Organisation</td>
</tr>
<tr>
<td>SPC</td>
<td>Statistical Process Control</td>
</tr>
<tr>
<td>SSA</td>
<td>Shared Services Agency</td>
</tr>
<tr>
<td>TDQM</td>
<td>Total Data Quality Management</td>
</tr>
<tr>
<td>TQM</td>
<td>Total Quality Management</td>
</tr>
<tr>
<td>WAVE Report</td>
<td>Working to Add Value through E-Information Report</td>
</tr>
</tbody>
</table>
Part A Introduction and Background
1 Introduction

1.1 Scope of the Research

It has been estimated by an industry consultant that 1–5% of data found in organisations are of poor quality (Redman, 1996a). The average perceived cost of poor data quality is as high as 10% of organisations’ revenues, according to a 1998 survey of New Zealand and Australian organisations (Malcom, 1998; Redman, 1996a, 1996b). The survey queried 29 organisations from government, banking and financial, utilities and service organisations. These costs arise through the need to repeat work, cleanse data, fix and find errors, reduced trust in data meaning duplicate data are collected, and lost customers through poor customer relationship management.

This thesis is concerned with the issues that define and control data quality and the mechanisms and frameworks that can be developed to achieve and sustain good data quality. In particular, the work focuses on the requirements to create high quality data in the New Zealand health sector and develops the fundamentals of data quality in this environment. The work extends and tests these principles to establish good practice and overcome practical barriers. Data quality is central to both health planning and delivery as it is a key factor in balancing the equation between appropriateness (i.e. quality) and cost effective health care.

1.2 Searching the Data Quality Literature

Data quality is now emerging as a discipline, with specific research programmes underway within Universities, the most significant being that of the Engineering School Information Quality Programme at the Massachusetts Institute of Technology (MIT)\(^1\). The field encompasses the well established Quality Discipline, drawing on the work of Deming (Deming, 1982), with the adaptation of the plan, do, check, act, cycle of Crosby (Crosby, 1980), through the notion that ‘quality is free’ because of the cost of doing things wrong, and Juran (Juran & Godfrey, 1999) through the utilisation of Six Sigma and Total Quality Management, adapted to Total Data Quality Management (TDQM) and the management of information as a product (Wang, Lee, Pipino, & Strong, 1998).

An extensive review of the data quality literature was undertaken by Wang et al (Wang, Storey, & Firth, 1995) in 1995, finding articles dating back to 1970. The review found that at that time research efforts focused on operation and assurance costs, research and development, and the production of data products. Much of the current literature continues to focus on solving specific data quality issues but with a move towards Customer Relationship Management (CRM) (Alshawi, Missi, & Eldabi, 2003) to justify the cost of improvement programmes. MIT publishes many articles outlining the systematic prevention of errors through the adoption of total quality management principles and much of the current research utilises these principles.
In the present study a search on electronic sources using the phrases ‘data quality’, ‘information quality’ and ‘data assurance’ provided an initial overview of the literature. The search was on databases from multiple disciplines due to the cross-disciplinary nature of data quality. For example, there is considerable research around data quality in health care that is published only in health related journals.

More specific searches on journals such as Communications of the Association of Computing Machinery (ACM), and the Institute of Electrical and Electronic Engineers (IEEE), and MIS Quarterly, elicited much of the foundation literature for the data quality discipline underpinning the current work. Further research was then elicited through a review of the references cited in the articles found. Websites provided considerable available literature through data quality organisations, such as the newly formed International Association for Information and Data Quality (IDIAQ)\(^2\), the MIT Information Quality Programme Website\(^3\), and Data Management Review Online\(^4\). The MIT website provides academic literature, with many commercial websites providing case studies and anecdotal information.

Health care related articles were located through searching the online health care databases, such as PubMed, Cinhal, and Medline. This was followed by specific journal searches. Much of the health care literature focuses on data quality in clinical research, with websites also containing a wider reference to data quality improvement for national and local clinical, administrative and management data that are required to manage health care. Key elements of this literature are discussed below.

### 1.3 Data, Information and Knowledge

Tayi and Ballou (Tayi & Ballou, 1998) define data as ‘the raw material for the information age’. A datum is a fact; a value assigned to a variable (Saba & McCormick, 2001), a single observational point that characterises a relationship (Shortliffe & Barnett, 2000). Data support managerial and professional work and are critical to all decisions at all levels of an enterprise (Tayi & Ballou, 1998), (Fuller & Redman, 1994). Data can take the form of single entities that are textual or numerical, but can also include documents, photographic images, sound, or video segments (McFadden, Hoffer, & Prescott, 1999). In particular in the health care environment data are found in many different forms than just textual (Hovenga, Kidd, & Cesnik, 1996). Unlike physical raw material, however, data are not consumed and in fact can be reused repeatedly for various purposes (Tayi & Ballou, 1998). Data models are the definitions of entities, their attributes, and the relationships between them that organisations use to structure their view of the real world (Fox & Levitin, 1996), (Rothenburg, 1996).

There is general agreement in the literature that data are now important to all organisations, regardless of their functions. Organisations are becoming more and more dependent on data; virtually everything the modern organisation does both creates and depends upon enormous quantities of data. A comprehensive

---

\(^1\) http://mitiq.mit.edu
\(^2\) www.iaidq.org
\(^3\) http://mitiq.mit.edu/
\(^4\) www.dmreview.com
data management program is therefore essential to meet the needs of the organisation (Pautke & Redman, 2001).

The management of data is becoming increasingly complex, in part through the progress of technology such as databases and telecommunications (Fuller & Redman, 1994). Pautke and Adelman (Adelman, 2001) have found that the typical organisation does not take full advantage of its data resources. There is often a poor connection between the organisation’s business strategy and the data it holds and manages. Data are of low accuracy levels, there is inadequate knowledge of what data resources are available and lack of management accountability. These issues point to the need for an organisation-wide policy on data management that actively considers data as a business requirement, that data should be of good quality and that accountability for data needs to sit at the highest level of the organisation. According to Adelman (Adelman, 2001), what is required is a ‘data strategy’.

Information is useful data that have been processed in such a way as to increase the knowledge of the person who uses the data (McFadden et al., 1999) and the term is often used interchangeably with ‘data’ in the data quality literature. High quality data and derived information are also needed to create institutional knowledge (stored information) plus reasoning processes that help an organisation extract the maximum benefit from the resources. This approach, which has recently been dubbed knowledge management (Davenport, 1998; Davidson & Voss, 2002), draws together the tangible and intangible elements of data and shares them amongst all workers.

English (English, 1999b), as a consultant in the field of data quality, builds on the idea of information being data in context, with knowledge being information in context, where you know the significance of the information. Translating information into knowledge requires experience and reflection. Knowledge emerges from the transformation of information, is derived through the formal and informal analysis and interpretation of data, but does not always result from processing information (Shortliffe & Barnett, 2000). Knowledge itself may be processed to generate decisions and new knowledge (Saba & McCormick, 2001) including the results of formal studies and also commonsense facts, assumptions, heuristics (strategic rules of thumb), and models – any of which may reflect the experience or biases of people who interpret the initial data (Shortliffe & Barnett, 2000).

1.4 Quality

Data (or information) quality is now recognised as one aspect of the Total Quality Management (TQM) movement. Perhaps the first and best known quality practitioner is W Edwards Deming. Deming initiated extensive changes in work practices that brought about a culture of quality at all levels of the organisation. The widely practised Deming Cycle for Quality Enhancement (Deming, 1982) consists of:

- **plan** – develop a plan for improving a process that produces commodities with unacceptable quality;
- **do** – implement the improvement in a controlled environment;
- **check** – assess the results to see if the plan has achieved the desired results on the level quality;
- **act** – if so, roll out the improvement to provide consistent results.
Another of the founders of the data quality movement is Joseph Juran. Juran is considered by many to be the father of the modern quality movement and developed the ‘Six Sigma’ theory of TQM (Juran & Godfrey, 1999). Juran purports that most of the possibilities for improvement lie in action on the organisation’s systems, and that contributions of production workers to improve quality are severely limited. Six Sigma is a whole operating philosophy of a customer focused methodology that drives out waste, raises levels of quality and improves the financial and time performance of organisations.

Crosby, another recognised quality practitioner, writes about the cost implications of data quality and his best known book is titled ‘Quality is Free’ (Crosby, 1980). The cost of quality is defined as ‘the expense of doing things wrong’. Crosby sees quality as an achievable, measurable, profitable entity that can be installed once you have commitment and understanding from management down to all levels of the organisation. Quality is defined as ‘conformance to requirements’ to assist in its management. Similar to Deming, Crosby notes that the system is where quality fails, not in workforce issues.

Some organisations have implemented TQM programmes with considerable success, i.e. increased efficiency and profits. There has been less success in companies where there has been incomplete buy in to the TQM philosophy, particularly amongst management (Juran & Godfrey, 1999). The health care sector is now beginning to implement quality management programmes to improve care processes in the light of poor safety records for patient care (Institute of Medicine, 2000) following recognition that more and better quality information is required to manage health care effectively (Pierce, 2004a).

1.5 Data Quality

Klein and Rossin (Klein & Rossin, 1999) note there is no single definition of data quality accepted by researchers and those working in the discipline. Data quality takes a consumer-focused view (consumers being people or groups who have experience in using organisational data to make business decisions) that quality data are ‘data that are fit for use’ (Wang, Strong, & Guarascio, 1996), (Redman, 2001), (Loshin, 2001). Data quality is ‘contextual’; the user defines what is good data quality for each proposed use of the data, within its context of use (Strong, Lee, & Wang, 1997), (Pringle, Wilson, & Grol, 2002). Therefore:

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)

Often the same data are used several times across the organisation for different purposes using different presentations. Therefore, data quality needs to be a multidimensional concept (Klein & Rossin, 1999) as data themselves are multidimensional (Fox & Levitin, 1996; Juran & Godfrey).

English (English, 1999a) makes use of the emerging discipline of ‘Enterprise Data Quality Management’ (EDQM) in his data quality consultancy practices, whereby the organisation develops and adopts a set of consistent technology processes, which institutionalise data quality as a strategic asset to make a consistent
competitive advantage. These theories have evolved from the TQM principle and provide the underpinning rigour to the academic discipline of data quality.

It is particularly important to note that data quality is not just ‘information technology’ or ‘information systems’ focused. The information system is merely the ‘enabler’, used to create, store, retrieve and manipulate data items. Although such a focus may provide insight and tools to help improve data quality, data quality improvements cannot be attained merely through information technology, the problem is one of processes and people and technology (Karr, Sanil, & Sacks, 2000). For example, the Data Warehousing Institute surveyed 647 data warehousing and business intelligence professionals (Eckerson, 2002). The survey revealed that 76% of data quality problems were caused by incorrect data entry by employees. Many studies note that people problems are most apparent when data collectors do not understand the importance of their role in the information production process (Lee, 2004), (Long, 2004), (UK Audit Commission, 2004), (Kmietowicz, 2004), (Pautke & Redman, 2001). As noted in Ward and Peppard:

… clearly, technology on its own, no matter how leading edge is not enough. (Ward & Peppard, 2002)

1.5.1 The Context of Data Quality

Context defines the type of data collected and how they are used, for example, customer data, financial data, electrocardiogram (ECG), monitoring data (Dravis, 2004a). Data are also collected within an organisational context, under defined policies and procedures. The context under which the data are collected may change over time (Lee, 2004). Data quality practitioners solve problems by understanding the contexts in which an organisation collects or creates data and how the data are stored and used. Data users may decide the quality of the data for their use depending on the context under which they were collected. This is particularly noted in the health care environment where data are collected from multiple disparate sources (Strong et al., 1997).

Problems can arise where users and practitioners are not informed of this context and may make assumptions without sufficient data quality information. Data quality information is metadata (data about data) that can be included with data to provide the user with information regarding the quality of that data (Fisher, Chengalur-Smith, & Ballou, 2003). Metadata are defined as:

all the characteristics that need to be known about data to build databases and applications and to support knowledge workers and information producers. (English, 1999a)

As data users are more and more removed from any personal experience with the data and the knowledge that would be beneficial in judging the appropriateness of the data for the intended decisions, increasing data quality information is required (Fisher et al., 2003).
1.5.2 Data Stewards, Custodians, Consumers and Collectors

The context under which data are judged for quality can be affected by the role or viewpoint of the assessor. Roles are defined by Abate, Diegert and Allen (Abate, Diegert, & Allen, 1998) and Wang, Ziad and Lee (Wang, Ziad, & Lee, 2001) as:

- **Data collectors** – those that provide initial input of data by creating, collecting, or supplying data for the information product;
- **Data custodians** – those that are responsible for storage and maintenance of data through the design, development and maintenance of information systems;
- **Data consumers** – those that utilise the data for further integration, aggregation, presentation, and interpretation of data and information products.

These roles can overlap, as a single person within an organisation may collect and utilise data, and may also have overall responsibility for the management of that data (Loshin, 2001). All roles may view the quality of the same data with considerable differences of opinion (Wang et al., 2001), particularly where data custodians are not aware of the uses of the data or do not have domain knowledge. Research on the roles in data quality has found that data collectors with knowledge about why data are collected throughout the data production process contribute to producing better data quality (Lee & Strong, 2003), (Kmietowicz, 2004). The education of the data collector is therefore one of the most effective measures to improving data quality (Kmietowicz, 2004), (English, 1999c), (Sanderson, Adams, Budden, & Hoare, 2004), (Haan, Adams, & Cook, 2004).

1.5.3 Data Governance and Ownership

Governance is the set of processes that ensure an asset is sustained for the benefit of a group of people who value that asset. Governance comprises two major processes: that of stewardship, and that of custodianship (Ministry of Health, 2004). Dravis (Dravis, 2004a) defines a data steward as ‘a person or group who manages those activities that encompass data creation, capture, maintenance, decisions, reporting, distribution, and deletion’. Data stewards have the authority to approve change and this may be their only role, where there is a data stewardship group consisting of representatives from stakeholders. The custodian then takes on the role of actively managing the data. A distinction between stewardship and custodianship in the case of national health data collections is provided in Table 1 below.

<table>
<thead>
<tr>
<th>Governance comprises:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>stewardship, which is:</strong></td>
<td><strong>and custodianship, which is:</strong></td>
</tr>
<tr>
<td>representation of stakeholder interests (both data providers and data users) in national collections requirements, definition and maintenance; monitoring of the delivery of collections to meet these requirements.</td>
<td>day to day management of national collections development or operations; operational decision-making on allocation of IS resources or funds; management of IS or business projects.</td>
</tr>
</tbody>
</table>

Table 1: Governance and Custodianship (Ministry of Health 2004)
Identifying data ‘ownership’ is considered paramount in data quality, as this ownership helps to define the roles and responsibilities throughout the data flow (Loshin, 2001). This can be complex in a distributed, national environment where there are many suppliers to one collection, all with varying uses of the data and the sometimes considerable cost to suppliers in collecting and submitting data to the national collection. Loshin (Loshin, 2001) discusses the ‘creator as owner’ concept, whereby a consortium creates the information, and data from all members are required for the data to be of use; therefore the consortium claims ownership of the data. What is required is a ‘data ownership policy’ whereby the stakeholders, data sets, responsibilities, and dispute resolutions are all clearly defined and to which stakeholders agree to subscribe (Loshin, 2001).

1.5.4 Dimensions of Data Quality

The concept of data quality is defined by a set of dimensions, usually considered in data quality literature as quality properties or characteristics of data. A data quality dimension is defined by Wang et al (Wang et al., 2001) as ‘a set of data quality attributes that most data consumers react to in a fairly consistent way’. This treatment of dimension is consistent with previous empirical research (Zmud, Lind, & Young, 1990). Modern definitions of data quality have a wider scope and many more attributes than the traditional characteristics of accuracy. Ballou and Tayi (Ballou & Tayi, 1999) identified four dimensions of data quality as being accuracy, completeness, consistency, and timeliness. In 1996, Wand and Wang (Wand & Wang, 1996) noted there was no general agreement on data quality dimensions and this is still the case today. However, research is beginning to show consistency, for example, Lee (Lee, Strong, Kahn, & Wang, 2002) concurred with Zmud et al (Zmud et al., 1990) that accessibility issues are becoming increasingly important to organisations. Each data quality practitioner needs to determine the dimensions of data quality applicable to the consumers of their data. Lee et al (Lee et al., 2002) note that dimensions employed by data quality practitioners are driven by the context in which they are delivering data quality, more so than academic research. Loshin (Loshin, 2001) notes that dimensions take on different levels of importance to different organisations.

Dimensions break down data quality into practical constructs that can be defined and measured. In defining appropriate dimensions for the context of their use the developer must take care to include all relevant dimensions of data quality to ensure subsequent appropriate measurements of data quality. For example, measuring timeliness is often considered important, as is completeness. By measuring both and prioritising their importance the organisation has appropriate data quality information on which to base decisions around a timeliness versus completeness trade off (Ballou & Pazer, 1995). Trade-offs are sometimes required where, for example, complete data are not available in a timeframe acceptable to the data user, and therefore the user could decide to use the incomplete data set. Data consumers may need to decide which is more important for the context in which they use the data.

These dimensions need to be clearly understood by everyone as having the same meaning to be effective (Wand & Wang, 1996). Without consistent definitions for the naming of dimensions, it will be difficult to develop generic data quality frameworks. At present there are no consistent definitions used and further research is required in this area to develop these definitions.
Wang & Strong (Wang et al., 1996) analysed the various attributes of data quality from the perspective of data consumers. Dimensions were then grouped into four broad categories: intrinsic, contextual, representational and accessibility. These dimensions and categories are detailed in Table 2. Intrinsic data quality captures the fact that data has quality in its own right. Contextual data quality highlights the requirement that data quality must be considered within the context of the task at hand. Representational and accessibility data quality emphasise the importance of the role of information systems (Wang et al., 2001) and reveal the need for ease of use and convenience if ‘quality’ data are to be of value.

<table>
<thead>
<tr>
<th>Category</th>
<th>Dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intrinsic</td>
<td>Accuracy</td>
</tr>
<tr>
<td></td>
<td>Objectivity</td>
</tr>
<tr>
<td></td>
<td>Believability</td>
</tr>
<tr>
<td></td>
<td>Reputation</td>
</tr>
<tr>
<td>Accessibility</td>
<td>Accessibility</td>
</tr>
<tr>
<td></td>
<td>Access security</td>
</tr>
<tr>
<td>Contextual</td>
<td>Relevancy</td>
</tr>
<tr>
<td></td>
<td>Value-added</td>
</tr>
<tr>
<td></td>
<td>Timeliness</td>
</tr>
<tr>
<td></td>
<td>Completeness</td>
</tr>
<tr>
<td></td>
<td>Amount of data</td>
</tr>
<tr>
<td>Representational</td>
<td>Interoperability</td>
</tr>
<tr>
<td></td>
<td>Ease of understanding</td>
</tr>
<tr>
<td></td>
<td>Concise representation</td>
</tr>
<tr>
<td></td>
<td>Consistent representation</td>
</tr>
</tbody>
</table>

Table 2: Data Quality Categories and Dimensions (Wang et al., 1996)

Gendron and D’Onofrio (Gendron & D’Onofrio, 2001) examined the data quality dimensions developed by Wang et al, (Wang et al., 1996) for three sectors of the health care industry, eliminated five dimensions and analysed the efficacy of the remaining 15. They found the dimensions, as noted in Table 3, to be sufficient to define data quality in all sectors of the health care industry, but that each segment of the health care industry must develop a set of domain specific dimensions to supplement the generic 15.
<table>
<thead>
<tr>
<th>Category</th>
<th>Dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy of Data</td>
<td>Believability</td>
</tr>
<tr>
<td></td>
<td>Accuracy</td>
</tr>
<tr>
<td></td>
<td>Objectivity</td>
</tr>
<tr>
<td></td>
<td>Reputation</td>
</tr>
<tr>
<td>Relevancy of Data</td>
<td>Value Added</td>
</tr>
<tr>
<td></td>
<td>Relevancy</td>
</tr>
<tr>
<td></td>
<td>Timeliness</td>
</tr>
<tr>
<td></td>
<td>Completeness</td>
</tr>
<tr>
<td></td>
<td>Appropriate amount of data</td>
</tr>
<tr>
<td>Representation of Data</td>
<td>Interpretability</td>
</tr>
<tr>
<td></td>
<td>Ease of Understanding</td>
</tr>
<tr>
<td></td>
<td>Representational Consistency</td>
</tr>
<tr>
<td></td>
<td>Concise Representation</td>
</tr>
<tr>
<td>Accessibility of Data</td>
<td>Accessibility</td>
</tr>
<tr>
<td></td>
<td>Access Security</td>
</tr>
<tr>
<td>Eliminated Dimensions</td>
<td>Traceability</td>
</tr>
<tr>
<td></td>
<td>Variety of Data Sources</td>
</tr>
<tr>
<td></td>
<td>Ease of Operation</td>
</tr>
<tr>
<td></td>
<td>Flexibility</td>
</tr>
<tr>
<td></td>
<td>Cost-Effectiveness</td>
</tr>
</tbody>
</table>

Table 3: Data Quality Categories and Dimensions (Gendron & D’Onofrio, 2001)

By defining the dimensions important to the health care industry, Gendron and D’Onofrio (Gendron & D’Onofrio, 2001) provide the data quality field with clear structure for further research into data quality specifically in health care.

### 1.5.5 Total Data Quality Management (TDQM)

Total data quality management (TDQM) is based on the traditional Total Quality Management (TQM) discipline and adapts the widely used Deming Quality Cycle (Deming, 1982), pictured in Figure 1, to encompass a continual cycle:

- **define**: e.g. what does data quality mean to the user?;
- **measure**: to provide data quality information through the measurement of data collections;
- **analyse**: what level of data quality do we need and where should our priorities lie?;
- **improve**: implement improvement initiatives. (Kovac, Lee, & Pipino, 1997), (Wang et al., 2001), (English, 1999a).
Figure 1: Components of the TDQM Cycle (Wang, 1998)

The definition component of the TDQM cycle identifies important data quality dimensions and the corresponding data quality requirements. Defining and measuring data quality should include both objective (metric based) and subjective (opinion based) views. The measurement component produces data quality metrics. The analysis component identifies root causes for data quality problems and calculates the impact of poor quality data. The improvement component provides techniques for improving data quality. The components are applied along data quality dimensions according to requirements specified by the consumer.

TDQM has been shown to be particularly effective in improving information management in organisations where top management has a strong commitment to a data quality policy (Wang et al., 2001).

Defining what data quality means to an organisation the level of data quality required by those who use the information can be a difficult task. Wang et al (Wang et al., 2001) provide some guidance.

- Define the functionalities of the information product. This is the data needed by the data consumers.
- Define the characteristics of the information product. Information product mapping shows us the ‘critical path’ that data takes as it develops into an information product and helps define the information product.
- Define the basic unit, components and their relationships for the information product through an entity relationship (ER) model. The ER model outlines the relationships between different data entities. For new data collections, Quality Entity Relationships (QER) can define where data quality can be incorporated into the database design (Wang et al 2001).
- Define the data requirements from the perspectives of information product suppliers, manufacturers, consumers and managers.
- Prioritise the data quality dimensions for the information product. This could include ranking, weighting or applying a trade-off to decide what dimensions are of the highest priority for an individual or group of information products.
Fundamental to the TDQM process is that data are seen as product, a valuable asset that should be produced as part of a well defined production process rather than the traditional view of data as a by product (Wang et al., 1998), (Ballou, Wang, Pazer, & Tayi, 1998), (Shankaranarayan, Ziad, & Wang, 2003). A process is ‘a structured, measured set of activities designed to produce a specified output for a particular customer or market (such as a bank statement or a hospital discharge summary). It implies a strong emphasis on how work is done within an organisation’ (Wu, 2004). TDQM looks at the processes that data flow through before ending in an information product. Whilst human, random error may lead to the entry of incorrect data, it is paramount that none of the processes themselves should change the initial meaning of the data leading to systematic errors and repeated data quality problems.

In order to manage data as a product, organisations should:

- know their customers/consumers of the data and their data needs;
- manage the data as if it were the product (rather than a by product) of a well defined data process, which includes considering the technology and the organisational culture;
- manage the entire life cycle of their information products;
- make management accountable for managing their data processes and resulting products (Wang et al., 1998).

The multidimensional nature of data and therefore data quality requirements means that TDQM encompasses the existing practices of ‘find and fix’ and adds to this the dimension of prevention. Whilst prevention of all errors is the aim, it is likely that some errors will still occur. Further, traditional TQM practices encourage building on existing practices and knowledge, rather than a complete change (Plsek & Greenhalgh, 2001). There is a paradigm or culture shift towards adding preventative measures and process management to a structured data quality programme.

Process management requires the mapping of each step along the data continuum from collection to warehousing. To meet this need, the MIT Information Quality research group has developed the concept of Data Production Maps (Ballou et al., 1998). Production Maps use the concept of the Data Flow Diagram (Shankaranarayan & Cai, 2005), (Wang et al., 1998), already understood by many information systems professionals. In addition a ‘quality block’ provides the analyst with the ability to pinpoint specifically in the data flow where data quality issues are likely to arise, what data quality initiatives are already underway, and where data quality checks need to be made. The data quality block enhances data quality so that the output stream has a higher quality level than the input stream; the nature of the activities performed by the quality control block is context dependent. Davidson (Davidson, 2004) has used this process to successfully map patient discharge reporting requirements in a health care environment. The Process Map could also be used in the development phase of a new data collection, in much the same way as data modelling techniques such as Entity Relationship Modelling (Moody & Shanks, 1998). This would provide the developer with a clear outline of where data quality needs to be instilled into the new system, allowing for the prevention of errors caused by poorly developed systems (Storey & Wang, 1994).
1.5.6 Alternative Approaches to TDQM

The alternative to TDQM, whereby issues are managed through single initiatives such as a ‘database cleanup project’, is compared with TDQM methodologies by Redman (Redman, 1994). Cleanup involves inventorying each record by comparing its fields with corresponding properties of objects in the real world, discarding inaccurate data, and entering new correct values. Many organisations employ outside consultants to provide one off database cleanups. This may be an expensive outlay, but provides the organisation with specialist services and a ‘quick fix’. Redman (Redman, 1994) does note that cleanup is often necessary, and could be used in conjunction with process management where specific deficiencies require immediate rather than long term improvement. Existing data management practices within the organisation are unlikely to change as a result, however, and in time further expensive cleanups could be required.

By comparison, management of the data processes would mean that all processes that can enter or change data be identified, the level of data quality measured, and the quality compared with the requirements. Processes that are found to be deficient can then be improved.

Data quality software exists that is able to test data collections for compliance with Codd’s database Integrity Rules (Codd, 1986), business rules compliance and data accuracy, and basic completeness of fields. Where data are more complex than names and addresses, however, this software can be of limited use and is often expensive to buy and support. Many organisations develop in-house software much more applicable to their needs. Much of the proprietary software available does not meet the needs of health care, in particular for clinical data, largely due to the complexity of the data. Whilst the software does not provide data quality information for all data quality dimensions or attributes that may be important to customers, it can be used as one method of data quality assessment. In the area of customer relationship management, data quality software is prevalent and often provides sufficient data quality information for organisations to improve their customer demographic data.

1.5.7 The Impact of Poor Data Quality

As noted earlier, it has been estimated that around 1-5% of data found in organisations are of poor quality (Redman, 1996a). The average perceived cost of poor data quality can be as high as 10% of organisations’ revenues, according to a 1998 survey of New Zealand and Australian organisations (Malcom, 1998). Gartner have found that ‘by 2005, Fortune 1000 companies will have lost more money in operational inefficiency due to data quality issues than they will spend on data warehouse and customer relationship management initiatives’ (Dubois, 2003).

The cost to private organisations is also far more than merely financial. Trust is lost from valuable customers (both internal and external), potential customers and sales are missed, operational costs increased, workers lose motivation, long-term business strategy is hindered (Leitheiser, 2001) and business re-engineering is impeded (Bowen, Fuhrer, & Guess, 1998; Redman, 1996a; United States Department of Defence, 2003), (Loshin, 2001). Redman also details how poor data quality affects operational, tactical and strategic decisions (Redman, 1996a). Strong et al note (Strong et al., 1997) the social impact of poor quality data when governmental organisations fail to ensure their data have sufficient quality to make effective decisions.
Organisations generally struggle with the issue of estimating the actual cost of poor data quality and the return on investment (ROI) of data quality initiatives (Loshin, 2001), (Ballou & Tayi, 1989). There is no formal structure that would be applicable to all organisations; priorities for improvement differ from one organisation to the next. The literature suggests that the cost of poor quality is often underestimated as not all impacting factors are considered (English, 2003), (Loshin, 2003), (Olson, 2003). In calculating the cost of poor data quality organisations need to:

- assess the business problems that create 'data scrap and rework';
- calculate direct (lost and missed opportunity) versus indirect costs (soft measures);
- identify the organisation’s key performance indicators (KPIs) and business drivers;
- identify critical data that has a significant impact on the business;
- meet with subject matter experts to assess the impact of poor quality (English, 1999a).

The rapid expansion of information technology only increases the problem through the ability to collect large amounts of disparate data and to integrate that data. Organisations are only now realising the implications of not addressing poor data quality in a systematic fashion (Stackpole, 2001). An example is the progress on e-business, where businesses need to change their methods of delivering products and services. Where web services have become the foundation of service delivery new and improved ways to manage data quality are required. This method of business has increased the complexity of data quality management due to the increase in a company’s interaction with its environment and new levels of data integration (English, 2003; Redman, 2001; Segev, 1999). Web based companies will have to establish trust with their customers. Those which provide high quality data have a better chance of doing so and can expect to achieve competitive advantages (Redman, 2001).

Poor data quality in health care can impact on patient safety. A significant review of the safety and quality of health care in the United States (Institute of Medicine, 2000) found that between 44000 and 98000 deaths in the US each year can be attributed to medical errors. While not all of these errors are attributable to data quality issues, the report recommends ‘better access to accurate, timely information’ and to ‘make relevant patient information available at the point of care’ in an effort to improve patient safety. The cost of compliance for health care providers to meet national reporting requirements can be considerable when systems do not provide a process for data collection and submission that enables high quality data management. A detailed review of data quality in health care is provided in Chapter Two.

### 1.5.8 Assessing and Measuring Data Quality

Pipino et al (Pipino, Lee, & Wang, 2002) conclude that assessing data quality is an ongoing effort that requires awareness of the fundamental principles underlying the development of subjective and objective data quality metrics. Currently many organisations’ data quality assessments are developed on an ad hoc basis to solve specific problems. Data quality is both the subjective perception of those involved in either the collection or usage of an organisation’s data; and an objective, measurable metrics-based analysis of the data sets. Subjective data quality reflects the needs of the stakeholders. However, because the evaluation is subjective it is important to be cognisant of the environment; to be aware that what represents poor data
quality to one stakeholder might be deemed more than sufficient for another, depending on their perception, usage of that data and requirements (Wand & Wang, 1996).

When performing objective data quality analyses, metrics must be established and a decision made as to whether to test the data in the context of the application and business rules in which it is utilised, or independently. In the latter case, the data quality test is often a test of the integrity, validity and quality of the data, whereas in the former, it is contextualised to extend the concept of quality into the operational environment that created, sustains and utilises those data (Pipino et al., 2002). Data quality problems are more likely to arise as data become more complex, as they are shared across multiple systems and as their volume increases.

The quality of data can be accessed at an internal level (the data conform to commonly accepted validity criteria), a relative level (they comply with the requirements of the user) and at an absolute level, where observation and sampling can be used to confirm that data closely resemble reality (Agmon & Ahituv, 1987). Strong et al (Strong et al., 1997) define high-quality data as ‘data that are fit for use by data consumers’. This concept of fitness for use places an emphasis on both the usefulness and the usability of data, leading to the definition of a data quality problem as ‘any difficulty encountered along one or more of the quality dimensions that renders data completely or largely unfit for use’ (Strong et al., 1997). As noted earlier the complex nature of data has led to a long standing belief that data quality itself has many attributes (Agmon & Ahituv, 1987), (Wang & Strong, 1996) leading to the introduction of data quality frameworks to capture this complexity.

1.5.9 Data Quality Frameworks

Seminal works (Wang & Strong, 1996), (Willshire & Meyen, 1997), (Eppler & Wittig, 2000) in the area of data quality have defined various extensive frameworks to review systems within organisations. The frameworks all seek to assess areas where poor quality processes or inefficiencies may reduce the profitability of an organisation. At its most basic, a data quality framework is a tool for the assessment of data quality within an organisation (Wang & Strong, 1996). The framework can go beyond the individual elements of data quality assessment, becoming integrated within the processes of the organisation. Willshire and Meyen (Willshire & Meyen, 1997) describe data quality frameworks as ‘a vehicle that an organisation can use to define a model of its data environment, identify relevant data quality attributes, analyse data quality attributes in their current or future context, and provide guidance for data quality improvement’. Eppler and Wittig (Eppler & Wittig, 2000) add that a framework should not only evaluate, but also provide a scheme to analyse and solve data quality problems by proactive management.

Porter’s (Porter, 1998) conception of theory development adds that a framework:

*identifies the relevant variables and the questions which the user must answer in order to develop conclusions tailored to a particular industry and company. Frameworks seek to help the analyst to better think through the problem by understanding the firm and its environment and defining and selecting among strategic alternatives available.*
Wang and Strong (Wang & Strong, 1996) argue that any conceptual model for data quality, such as a framework, should ensure that the consumer:

- is able to access the data;
- can interpret the data;
- sees relevance in the data;
- finds that the data are accurate.

The final step is to identify, evaluate and select effective remedies. These should be based on the current and future functional context or environment, type of data, processes affecting the data, recommendations from the literature, experience and best judgement. The framework is a step-by-step process. Steps include:

a) modelling existing data;

b) defining data quality attributes;

c) determining data quality priorities;

d) evaluating existing data quality levels;

e) identifying remedies and

f) re-measuring and reiterating (Gendron & D'Onofrio, 2001).

Rather than defining a framework, Willshire and Meyen (Willshire & Meyen, 1997) provide a methodology for the development of domain specific frameworks. The methodology prescribed is:

- develop an appropriate functional and data modelling paradigm;
- define data quality attributes;
- collect, measure and analyse data quality attributes;
- identify, evaluate, select, apply and analyse results.

Similar to the Wang & Strong framework (Wang & Strong 1996), the Willshire and Meyen (Willshire & Meyen, 1997) framework uses a modelling phase before the traditional define, analyse and improve quality cycle. The model would include a description of the environment which the data supports, to provide context for data users, as well as describing their needs. Data requirements flow and business rules would be defined.

In developing a data quality framework for the International Monetary Fund, Carson (Carson, 2000) notes that an assessment tool for data quality needs to have the following characteristics:

- comprehensive coverage of the dimensions of quality and characteristics that might represent quality;
- balance between rigour desired by an expert and the bird’s eye view desired by a general data user;
- structure but enough flexibility to be applicable across a broad range of data collections;
- a lead to transparent results;
- a conclusion, arrived at by drawing on best practice.
Data quality practitioners agree (Willshire & Meyen, 1997), (Carson, 2000) that a framework needs both objective and subjective attributes to be considered, using both objective and subjective measurement metrics to reflect the contextual nature of data quality and the many users of that data.

An extensive review of data quality frameworks from the last ten years (Eppler & Wittig, 2000) analyses seven conceptual frameworks, identifying common elements, differences, and missing components of such frameworks and outlines future direction in the development of data quality frameworks. The review found that existing data quality frameworks are often domain specific and either strong on objective or subjective measurements, but not strong on both types of measurements at the same time. Frameworks also often fail to analyse the interdependencies between the various criteria within the framework. Therefore, Eppler and Wittig (Eppler & Wittig, 2000) suggest new developments in data quality frameworks provide:

- a generic framework, not specific to a single application such as data warehouses or corporate communications;
- a framework that shows interdependencies between the different quality criteria;
- a framework that includes a list of problem areas and indicators, therefore going beyond a simple quality criteria list;
- the development of tools which are based on information quality frameworks;
- a framework that is at the same time theoretical and practical.

Included in the Eppler and Wittig (Eppler & Wittig, 2000) assessment of frameworks is the Wang & Strong (Wang & Strong, 1996) data quality framework. This framework was assessed using the following criteria – clear definitions, contextual positioning, mutually exclusive and collectively exhaustive criteria, concise, provides examples, and provides tools. Table 4 shows the conclusions made by Eppler and Wittig (Eppler & Wittig, 2000) when assessing the Wang & Strong (Wang & Strong, 1996) framework. They found that overall the framework is generic and balanced. The Wang & Strong (Wang & Strong, 1996) framework was the only one out the seven assessed that included a means of measurement, tools to assist with using the framework and offering both a solid foundation in existing literature and practical applications. The framework also stands out as being the only one to strike a balance between theoretical consistency and practical applicability.
Criteria: | Wang & Strong 1996
---|---
1. Definitions | All criteria and dimensions are defined.
2. Positioning | Clearly positioned within existing information quality literature in the information technology context.
3. Consistency | Overall concise. Some criteria are quite similar such as interpretability and ease-of-understanding.
4. Examples | Many examples are provided.
5. Conciseness | 16 criteria in 4 dimensions
6. Tools | A comprehensive tool (questionnaire with software support) provided.
Conclusion | Generic, balanced


The Eppler and Wittig (Eppler & Wittig, 2000) review of frameworks provides empirical support for the development of the Canadian Data Quality Framework and this is discussed further in Chapter Five.

It can be said from the literature then, that a data quality framework is:

* a point-in-time assessment and measurement tool, integrated into organisational processes, providing a benchmark for the effectiveness of any future data quality improvement initiatives and a standardised template for information on data quality both for internal and external users.*

Data quality frameworks can provide guidelines and structure to an organisation-wide data quality improvement programme.

1.5.10 Data Quality Improvement Programmes

Olson (Olson, 2003) notes that any data quality improvement programme has the goal to reach high levels of data quality within the critical data stores. It must encompass all existing important databases and, more importantly, is a part of every project that creates new data stores or that migrates, replicates, or integrates existing data stores. Its mission is three-fold; to improve, prevent, and monitor.

An analysis of the requirements for a data quality improvement programme finds that the data quality practitioners, including English (English, 1999a), Wang (Wang et al., 2001), Olson (Olson, 2003) and Loshin (Loshin, 2001), agree that to make effective improvements to data quality the organisation needs to:

- take an organisational view of data management and data quality;
- have high level buy-in from management to support change and projects across business units;
- instil a continuous process of improvement and assessment, such as TDQM, with effective measurements (data quality metrics) to assess whether improvement has taken place;
- utilise total quality management principles, such as a customer focused view, to define quality user requirements;
• use statistical process control (Carey & Lloyd, 2001) to highlight process management, process measurement and process improvement;
• instil, in the organisation, a culture of producing information products that contain high quality data;
• clearly define accountability and ownership of data through data stewards and stewardship policies;
• provide contextual information and clear standards on data definitions through a metadata repository that is maintained and available to all stakeholders;
• treat data as a product with a life-cycle, map the processes the data flows through and where quality can be impacted in that flow;
• prevent errors and expensive re-work through root cause analysis of known process problems;
• instil data quality into new systems at the development phase or at the time of system change/upgrade;
• profile and model data and systems to highlight problems with data quality;
• understand that data quality is not just an information technology (IT) problem; IT enables the movement of information but is reliant on the data input to be of high quality. The cause of poor quality data is often found to be human or process error.

A programme of work is required by many participants in an organisation and often across business units to implement the above initiatives and such a programme requires long term commitment (Olson, 2003).

1.5.11 A Strategic View of Data Quality

There are currently few published ‘data quality strategies’, although the components of many data quality improvement programmes could be considered strategic. A review of the literature shows there is no one definition for what a strategy is. Robson (Robson, 1997) defines strategy as ‘the pattern of resource allocation decisions made throughout an organisation. These encapsulate both desired goals and beliefs about what are acceptable and, most critically, acceptable means to achieving them’. In discussing data quality strategies, Dravis (Dravis, 2004a) came to an appropriate definition of a data quality strategy:

*a cluster of decisions centred on organisational data quality goals that determine the data processes to improve, solutions to implement, and people to engage.*

According to Dravis (Dravis, 2003) a data quality strategy should include the following:

1. a statement of the goals. What is driving the project?
2. a description of the primary organisational processes impacted by the goals
3. a high-level list of the major data groups and types that support the operations
4. a description of the data systems where the data groups are stored
5. a statement of the type of data and how they are used
6. discussion of cleansing solutions matching them to the types of data
7. inventory of the existing points where data are accessed
8. a plan for how, where, and when the data can be accessed for cleansing
9. a plan for how often the cleansing activity will occur and on what systems
10. a detailed list of the individual data elements.

While Dravis's (Dravis, 2003) list is a useful starting point, it does include aspects that could be considered 'sub-strategies', such as point numbers 6 to 10, where operational teams may develop applicable strategies to meet these requirements. These sub-strategies may not be applicable across all of the organisation’s business units. Further, a data quality strategy must consider the needs of the customer and define and document these (Wang et al., 1998). While it may not be possible to meet all the needs of all customers, a practical strategy will define the most important customers and meet their most important needs. This may be difficult to do, particularly as customers often do not know what their needs are (Redman, 2001).

Fuller and Redman (Fuller & Redman, 1994) found when improving data quality in a telecommunications company that, similar to Dravis (Dravis, 2003), strategies must feature the process approach to ensure long-term sustainable improvement. The greater the rate of turnover in data, the greater the relative effort applied to the process approach and that data clean-ups should be targeted at data that turn over relatively slowly.

Redman (Redman, 2001) discusses ‘second generation data quality systems’. The difference between first and second-generation systems (systems here means the totality of the organisation’s efforts on data quality) is the inclusion of the prevention of errors, reducing the amount of clean up and correction of data required. The second-generation system includes ‘strategic data quality management’, whereby the data quality program aims to ensure that the organisation’s business strategy is ‘data enabled’ so the organisation has the data and information assets required to effect its strategy. One benefit of a strategic view is that data are related to strategy and not to information technology. A third generation data quality system would be based on the philosophy of making it virtually impossible to make errors, similar to six-sigma theory in manufacturing. Redman (Redman, 2001) notes most organisations currently focus on the cleaning and correction of data. When a data quality problem is found, organisations increase the effort of clean up and correction and are often reticent to change from this paradigm (Lee, 2004). This approach will not continue to be effective in the absence of preventative measures due to the increasing amount of data collected and could become increasingly cost ineffective. If clean up of data is required, then root cause analysis should be performed and changes made to the process or system to ensure the clean up does not need to be repeated (Redman, 2001), (Loshin, 2001), (English, 1999a). Changing the processes by which data are collected is the fourth step in the TDQM process (Lee, 2004).

1.6 Present Research

1.6.1 Aims of the Research

This research aims to develop a data quality evaluation framework and improvement strategy for the New Zealand Health sector. The research elicits the dimensions of data quality that data consumers, collectors and custodians in the New Zealand health care sector consider important to measure and improve. These dimensions form the basis of the development of a national strategy for data quality improvement that provides guidance initially to data custodians and suppliers of the national health collections, but that can also be used by all stakeholders of data in the health sector. The research was limited to the improvement of
data quality on the national health data collections held at NZHIS, with data supplied by multiple public health care providers. A national health data collection is a long term collection of nationwide data or reference data set, of which NZHIS is the custodian on behalf of the sector, and which is used for analysis and information provision to achieve improvement in the services and capabilities of the publicly funded health sector (Ministry of Health, 2004). A comparison with other domains highlights different driving forces, such as a need for cost effective and safe health care rather than improving profit margins.

The research explores the progress of the learning and development of stakeholders through action research methodology that enables the changes in philosophy required to institutionalise data quality, first within the Ministry of Health and then out into the national public health sector. Institutionalisation in this context is defined as:

> fundamental changes in daily behaviours, attitudes, and practices that make ‘changes’ permanent. The cultural adoption of changes made by process improvement, design or redesign including complete business systems such as HR, MIS etc (One Six Sigma Glossary).

### 1.6.2 Research Questions

There was considerable impetus from within the organisation for improvement in data quality management, and any improvement needed to be applicable to the local environment. The organisation’s management felt that improvements needed to be based on the existing literature, where available and appropriate. The framework developed by the CIHI appeared to provide an applicable initial guideline for the organisation to further explore through action research methods. The absence of literature around strategic data quality management in general, and in particular in the context of a national health sector, provided justification for undertaking the research for the purpose of a PhD thesis.

The research questions were developed following an initial review of the literature and discussions with the research participants and stakeholders of data in the health sector. This helped to elicit the high level needs of the health sector around the improvement of data quality.

1. What existing data quality theories assist in the decision making around defining and measuring data quality requirements in national health data collections?
2. What are the data quality dimensions considered important to measure and improve by data consumers, collectors and custodians in the New Zealand health sector?
3. What initiatives assist to raise the profile of data quality improvement in the New Zealand health sector?
4. What are the data quality roles found and by which stakeholders, in the New Zealand Health sector?
5. Does action research methodology provide appropriate research methodologies for understanding learning and change initiation in a data quality improvement programme?
6. What steps would be required to initiate a structured data quality improvement programme within the New Zealand Health Information Service?
7. What are the components of a national health care data quality improvement strategy that initiates effective ‘institutionalisation’ of total data quality management across the New Zealand health sector?

These research questions have evolved through the learning process enabled by action research and grounded theory methods for the analysis of research data. The initial action research cycle focused on questions one to five. Following initial data collection and analysis it was evident that further work was required to prevent increasing data quality problems through the development and implementation of strategic management initiatives and the research was expanded to answer research questions numbers six and seven.

1.7 Chapter Conclusion

A review of the current data quality literature finds that there are increasing levels of research from an emerging discipline that seeks to find ways to improve data quality. This is achieved through the adaptation and application of TQM processes and the implementation of TDQM. Much of the literature can be found in information technology journals, and also in journals relating to specific domains, where the research outlines the application of data quality practices within an organisation or domain. The impetus for research is the increasing complexity of data management through improved information technology and telecommunications, allowing for the movement, integration and subsequent mining of data. Organisations are beginning to realise the importance of data as an asset to the organisation, the potential loss of profits through poor customer relationship management, and the cost of having to fix poor data once it is in the information systems.

The research thus far has provided the discipline with the theoretical underpinnings required to develop practical structured programmes to address data quality from a holistic perspective, whereby all aspects of data management are addressed with prioritisation for improvements that meet the needs of customers, as defined by customers. The roles of customer, collector and custodian have been defined and research has noted the differing data quality needs and perceptions for each of these roles.

Research is now developing ways to combine TDQM into the strategic direction of the organisation, aligning the data quality requirements with overall goals of the organisation. At present, there is little research published in this area, although some organisations do have data quality programmes with some strategic alignment to the business requirements.

This thesis seeks to develop a national data quality strategy in the domain of health care, where there is now considerable interest in general quality improvement in health care delivery and management that requires alongside it an improvement in access to high quality clinical and administrative data. The strategy is aimed at the initial ‘institutionalisation’ of data quality practices within the New Zealand health care sector. It is important to note that the New Zealand health care system is not competitive, but collaborative. For this reason the strategy development does not seek to provide financial competitive advantage, but does seek to provide the New Zealand health sector with world leading data quality management tools.
2 Data Quality and Health Care

Health care planning and delivery rely heavily on data from management, administrative and clinical sources; nearly all health care activities involve gathering, analysing, or using data. Quality data can lead to quality and cost-effective health care delivery by improving patient outcomes through better decision making. Data quality is inextricably linked to the use of information systems and the health sector is increasingly an information-driven service (Hovenga et al., 1996), where information held in databases and other electronic repositories delivered in a reliable and timely manner, is critical to the health and well being of patients, the wider population, and to the management of health care organisations (Long & Seko, 2002). Along with an increase in information complexity, there has been a parallel increase in the complex nature of organisations and organisational relationships within the health sector (Gendron & D’Onofrio, 2001).

2.1 What are Health Data?

Health data deal with a continuum that ranges from the clinical records of individual patients, detailing their interactions with health services, through aggregated data and data warehouses, knowledge-based data for planning and decision support, to comparative and community data that can be used at a policy development level (Al-Shorbaji, 2001). Health care data include the administrative data required to manage the complex business of health care, but also data captured from the patient’s medical record.

Data are often collected at the individual patient level and referred to as ‘unit level data’. Aggregated data are the sum of unit level data. Data are used to support the proper care of the patient from whom the data were obtained, and they may also contribute to the good of society through the aggregation and analysis of data concerning populations of individuals. In an aggregated form, health data within the health care organisation provides information of the prevalence and distribution of disease in a population, and at the government level informs policy makers.

Diagnosis relies on data gathered through observations. Gathering data and interpreting their meaning is central to the health care process because these steps are central to the process of decision-making. Increasingly wide ranges of data are collected from the multidisciplinary team to eliminate or confirm the possibility of the presence of a specific disease. Data in health care are found in many different formats and are used in many different ways. One item of datum may be used several times over:

- in the patient’s electronic health record;
- in expert systems for artificial intelligence;
- by a decision support system;
- by the hospital administration system;
- for claiming payment from central government or insurance companies for services provided;
- for the purpose of reporting to a national health collection.
Health data can also be made up of varying types of information. Clinical information is generally based on patient related data and often found in patient records. The data describes the characteristics of the patient, of the illness, and of the health care process. Medical knowledge abstracts from the individual patient and describes general insights, for example, about a disease, diagnosis or procedure (Liener, Gaus, Haux, & Knaup-Gregori, 2003) and administrative data provide information on the cost or services required for delivering care.

Within these types the data format also varies. Data may be narrative, textual, numerical measurements, recorded signals, still pictures and videos. In some fields of medicine, data in the form of continuous signals are particularly important, such as the ECG, a tracing of the electrical activity from a patient’s heart (Shortliffe & Barnett, 2000). Data collection in health care provides the ability to:

- anticipate future health problems;
- record standard preventative measures;
- identify deviations from expected trends;
- provide a legal record;
- support clinical research (Shortliffe & Barnett, 2000).

Table 5 describes the various types of health care users for the types of data found in health care, and the scope of the data.

<table>
<thead>
<tr>
<th>Users</th>
<th>Data/Information</th>
<th>Scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>World health officials</td>
<td>General health status and health-related needs of individual nations</td>
<td>World wide data</td>
</tr>
<tr>
<td>Policy makers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Researchers</td>
<td></td>
<td>Abstracted, Summarised, aggregated</td>
</tr>
<tr>
<td>Lawmakers</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trend in incidence, prevalence, outcomes, and costs by region, by diagnosis, by type of provider</td>
<td>Nationwide data</td>
</tr>
<tr>
<td>Policy makers</td>
<td></td>
<td>Abstracted, Summarised, aggregated</td>
</tr>
<tr>
<td>Researchers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lawmakers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insurers</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Comparison of treatments, outcomes, and costs by locality and by provider. Incidence and prevalence of diagnosis by region</td>
<td>Community/region wide data</td>
</tr>
<tr>
<td>Analysts, Researchers, Quality management</td>
<td></td>
<td>Abstracted, Summarised, aggregated</td>
</tr>
<tr>
<td>Public health officials</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Administrators, Researchers, Accreditors, Quality managers</td>
<td>Costs of care by category of patient. Number of patients admitted with specific diagnosis, volume of tests, procedures and interventions, outcomes for patients grouped by diagnosis</td>
<td>Funder/Provider organisation wide data</td>
</tr>
<tr>
<td>Care givers, Provider organisation departments, Insurers, QA personnel</td>
<td>Unit level patient specific data e.g. assessments, diagnosis, interventions, diagnostic test results, procedures, treatments, outcomes. Used to provide most appropriate care</td>
<td>Individual patient data</td>
</tr>
</tbody>
</table>

Table 5: Examples of Types and Uses of Health Care Data. Adapted from Saba and McCormick (Saba & McCormick, 2001).
Health data are collected by physicians, nurses, office staff, admissions personnel, physical and respiratory therapists, laboratory personnel, radiology technicians, and pharmacists. Therefore, all those who work in healthcare collect data through the nature of their occupations, as these data are required in the performance of their jobs.

Recording medical data electronically can be difficult when the data can be viewed and perceived in many different ways. A blood pressure is a good example – a reading of 120/80 can be read as a single datum, but also provides decomposed information as a dual view of a systolic blood pressure of 120 and a diastolic pressure of 80. Whilst humans are easily able to make this pluralistic distinction, a computer would need this allowed for in the design process for data storage and processing. The contextual requirements of such information in health means that computers can have a limited use – any subjective or intuitive knowledge of the clinician is difficult to measure and record on a one-level, non-relational database.

### 2.1.1 Medical Registries and Repositories

Progress in information technology and increasing demands for accountability have led to an increase in the number of medical registries over recent years. A medical registry is defined by Arts, De Keizer, and Scheffer as;

>a systematic collection of a clearly defined set of health and demographic data for patients with specific health characteristics, held in a central database for a predefined purpose. (Arts, De Keizer, & Scheffer, 2002).

These registries can serve different purposes, such as a tool to monitor and improve quality of care or as a resource for epidemiological research. To be useful data in a medical registry must be of sufficient quality to meet its purpose. However, several studies note the often poor quality of data held in medical registries (Arts et al., 2002; Davidson, Melinkovich, Beatty, Chandramouli, Hambridge, Phibbs, Braun, LeBaron, & Steiner, 2003). Arts defines data quality within a medical registry as ‘the totality of features and characteristics of a data set, that bear on its ability to satisfy the needs that result from the intended use of the data’ (Arts et al., 2002).

A case study by Broder (Broder, 2004) found that the implementation of a medical registry to monitor quality of care improved patient mortality rates by 28% and morbidity by up to 55%. Using data from the repository, treatment algorithms were developed for patients who are at risk for certain conditions. In some cases, this has involved doing something as simple as putting important information about patient risk factors on laminated pocket cards that doctors can easily reference. Such a case study indicates the impact of access and use of health care data on the delivery of care.

### 2.1.2 The Patient Record

The patient record is composed of all data and documents generated or received during the care of a patient at a health care institution. Every patient record contains information on demographics, General Practitioner
details, the patients’ health history, complaints and reason for seeking treatment, information and results about diagnostic examinations, outcomes to treatment and interventions, adverse events, and discharge summary information. These records are archived for future reference, such as the patients’ next hospital visit and are increasingly being computerised. It has been estimated that the average patient generates up to 50,000 data items during their life (Weed, 1989) and computers are able to assist with the management of this information.

Individual health care organisations generally have electronic patient management systems that capture demographic and sometimes clinical data and provide for appointment scheduling. These systems may link to other information systems, such as those holding laboratory and radiology data. In primary care, patient management systems are often advanced in their clinical applications and provide the general practice with many of the tools required to manage a busy practice and significant patient workloads. In New Zealand at least 85% of GPs now use computers, some for administration purposes only but many use them to record all clinical notes, prescribing and laboratory test ordering and diagnosis coding (Gribben, Coster, Pringle, & Simon, 2001). However, most of these systems do not integrate in any way to systems outside the individual general practice.

A clinical data repository is an aggregate of all computer based clinical data for patients. Electronic records support the integration of care across different health care providers by:

- supporting the capture of and access to a patient’s outpatient clinical data as he/she moves through the system;
- capturing data for quality measurement and research;
- supporting best care practices by reminding physicians to perform certain tests and complete documentation (clinical decision support);
- structuring data collection to enhance the quality of care by guiding, for example, regular testing of certain parameters for patients with diabetes (Glaser, 2002).

### 2.1.3 Electronic Health Records

The ‘electronic health record’ (EHR) is increasingly being used to deliver health information to the point of care. Considerable work is now underway in many countries to develop the components of effective and comprehensive EHRs, and EHR development is central to many national health information strategies (Orfanidis, Bamidis, & Eaglestone, 2004). EHRs consist of components that enable health care providers to access a patient’s health information regardless of geographical location. Other components refer patients to various providers and enable providers to access clinical decision support and electronic prescribing, physician order entry, and integrated communication with laboratories, imaging centres, colleagues and patients; and would include population health management with the overall goal of assisting providers to give better quality health care (Kerr, 2004).
EHRs could simply be software applications that provide integrated, longitudinal views of patient data; providing several software applications in a networked environment. The Australian Health Online project provides a definition of an EHR that includes all the components of a comprehensive EHR model:

*An electronic health record is an electronic longitudinal collection of personal health information, usually based on the individual, entered or accepted by health care providers, which can be distributed over a number of sites or aggregated at a particular source. The information is organised primarily to support continuing, efficient and quality health care. The record is under the control of the consumer and is to be stored and transmitted securely.* (Australian Department of Health and Aging, 2003)

Orfanidis (Orfanidis et al., 2004) found evidence that gaining the acceptance of EHR systems by physicians, nurses and other health care professionals is often problematic. A primary reason cited is a general lack of trust in EHR systems with respect to the critical issue of data quality, since possible errors may cause serious problems to patient health. A further barrier to effective understanding and use is the poor computer literacy of some health care personnel. Also, EHR systems can be perceived as a distraction, which impedes the physician’s focus on the patient.

It may be that health care is missing out on the potential benefits of the EHR through these perceived data quality issues. There is a building body of evidence that suggests capturing and providing health care data in an electronic form improves the safety and quality of care provided and reduces the clinician’s time spent on administrative tasks (Menke, Broner, Campbell, McKissick, & Edwards, 2001), (Patel, Kushniruk, Yang, & Yale, 2000), (Bates, 2002), (Ball & Douglas, 2002), (Institute of Medicine, 2000). In particular, components of the EHR, such as electronic prescribing, reduce likelihood of human error with as much as a 55% reduction in serious medication error rates (Bates, 2002). This is achieved through changing the ‘systems of care’ rather than merely computerising the existing workflow.

### 2.1.4 Clinical Coding

In clinical coding a short and formal code is recorded against clinical data to enable more accurate data analysis, making clinical data shorter and easier to record, but restricting the variability of expression. A coding system is defined as a documentary language that is based on a concept system (Shortliffe & Barnett, 2000). A concept system is a system of labels that are defined in an ordered way. This means, therefore, operations can be divided into operations on the nervous system, chest or blood vessels, then further divided into operations on the chest that can be broken down into operations on the heart, lung etc, ending up with further distinctions such as valvotomy of the aortic valve (Liener et al., 2003), (Shortliffe & Barnett, 2000). The health status and medical condition of a patient is difficult to describe using comprehensive, coded data. Factors that contribute to this problem include the following:

- accepted methods for formally decomposing many key components of the patient record, such as admission and physical history, into coded concepts have not yet been developed;
• where available, vocabularies to represent the terms within the model in a standard way are difficult to develop. The condition of the patient is often complex, probabilistic, and highly nuanced. Multifactorial and temporal relationships can exist between pieces of data;
• there is no single way to organise automated medical data. There are often idiosyncratic ways developed to code data within and across organisations;
• data definitions vary across settings. These variations exist in problem lists, laboratory tests and procedures, for example;
• care processes vary, such as referrals, medication ordering, care documentation;
• definitions of optimal care vary, such as frequency of visits or appropriate tests;
• differing technologies, vendors, and levels of IT sophistication are found across settings, as well as different levels of computerisation (Glaser, 2002).

Therefore, establishing clinical meaning, measuring care, determining the health status of a patient or population, and developing clinical information systems that interoperate can be difficult. The complexity of boundaries compounds the complexities of the medical process and health data (Glaser, 2002), (Sanderson et al., 2004).

2.1.5 Privacy of Patient Data

Patients have a right to have personal health data kept private. The outcome of indiscreet record keeping could be that patients no longer disclose intimate details that are necessary for their proper medical care, as they do not trust the clinician to keep this information private. The patient owns his or her data and no one has the right to access this without the patient’s authorisation.

In the era of electronic communication, this is somewhat more complicated to ensure. To address the changing environment and the changing attitudes of health consumers, most OECD countries now have privacy regulations, with New Zealand being one of the founding developers of such regulations. The Privacy Act of 1993 has provided guidelines for the collection, use and storage of health information in New Zealand.

In the United States, HIPPA, or the Health Insurance Portability and Accountability Act of 1996, has further developed the privacy regulations for health data in which the patient is identifiable. The regulations refute the view of computer experts that personal control of private information is an illusion in the computer age and require that all patients receive a notice of privacy practices around their health care record. Privacy regulation needs to take into account legitimate users of the data for public good. These include:

• health care providers;
• patients;
• health care organisations;
• research institutions;
• public health departments;
• insurance companies;
• law enforcement;
• accrediting bodies (such as for health care organisations) (Saba & McCormick, 2001).

Adequate protection of health information depends on both technological and organisational policies.

2.1.6 Stewardship of Health Care Data

Stewardship involves taking responsibility for the good care of a set of resources on behalf of others (Ministry of Health, 2004). This involves:

• protection and development of the resource;
• ensuring management of the resource in a sustainable manner; and
• representing the interests of stakeholders in the management of that resource.

When multiple agencies, organisations or individuals share common needs and interests in a set of resources, there is a need to have input and representation from these group(s) in the management of the data. Stewardship arrangements should provide processes and mechanisms for these groups to act collectively in the interest of the effective and sustainable management of data (Ministry of Health, 2004).

Without appropriate and effective stewardship arrangements it is not clear how stakeholder input is gained, or who is responsible and accountable for defining requirements. Both duplication of activities and gaps in activities can occur where no-one assumes responsibility for decisions (Seiner, 1997). The result is poor definition and understanding of stakeholder needs and the development and use of resources is less than ideal (Loshin, 2004). Health strategies are key drivers of national collection establishment and change. Stewardship guides the alignment of national collections within the health strategy, but has no role in defining those strategies. Stewardship also recognises consumer interest in data, and provides opportunity for consumer representation in the stewardship processes (Ministry of Health, 2004). The key design principles to achieve effective stewardship are:

• stakeholder representation – stewardship relies on the involvement of stakeholder groups including both the providers of data and the users of that data;
• link to organisation directions and priorities – the underlying driver of all stewardship actions is support of the organisation’s strategies and priorities;
• structure alignment with organisation culture – in order to be effective the stewardship roles and structures must align with existing organisational culture and accountability arrangements;
• management commitment – stewardship must be sponsored and supported by senior management in all the organisations affected;
• sustainability – any implementation of stewardship needs to be as practical and streamlined as possible, so that the role is sustainable (Ministry of Health, 2004).
2.2 Health Informatics

Health informatics has emerged as a discipline to develop appropriate computer systems to assist with the management of health care information. The World Health Organisation defines health informatics as:

*the combination of technology and methodology which makes possible the computer-assisted collection, storage, processing, retrieval, distribution and management of health information* (World Health Organisation, 1988).

Ranson (in (Hovenga et al., 1996)) notes that $28 billion could be saved in health care costs through the efficient use of telecommunications and the movement of patient management information. Health care costs are increasing; in western countries this ranges from about 6% to more than 12% of gross domestic product (GDP) (van Bemmel & Musen, 1997). Now that technology has progressed to the point where all types of data, with the exception of smell, may be produced in digital form it has become feasible to develop fully integrated health information systems (Hovenga et al., 1996). However, most of the evidence needed to make more informed decisions from the clinician at the bedside up to the formulation of national strategy remains embedded in fragmented, irretrievable, and often illegible paper-based patient records (Bates, 2002). Consumers are now coming to expect good quality, co-ordinated care and interoperable information systems provide the communication means required to deliver this kind of care (Rigby, 1998).

However, as Effken (Effken, 2002) notes, health care systems are complex sociotechnical systems in which many information system innovations fail because of problems in planning or design. Designers with an understanding of this complexity are required to provide appropriate analytical approaches to development. Ribiere, LaSalle, Khorraramshahgol and Gousty (Ribiere, LaSalle, Khorraramshahgol, & Gousty, 1999) found, when researching hospital information systems quality, that there is a significant lag behind other industries in the health care sector’s ability to deliver good quality information. Further, at present there are no current accreditation or regulatory requirements for health care software or hardware to provide quality assurance around such things as the privacy of health information through the appropriateness of design and integrity of systems (Rigby, 1998).

Health informatics research, in its position across the disciplines of health care and information systems, provides for the significant developments needed to manage health care in the future. Health informatics research may enable the development of systems that increase flow of information from the health care recipient and the provider to various organisations for fund and resource allocation and for management.

2.3 Why Improve Data Quality in Health Care?

Examples of the implications of poor quality data in health care provide considerable impetus for professionals and managers to ensure data quality management is a priority. Single errors can have significant effects on patients and are highlighted by the examples below:

- errors in a reference database calculation of Down’s Syndrome screening, giving false negatives (Rigby, 1998);
• an age cohort of women omitted from call up for cervical screening (Rigby, 1998);
• duplicate unique identifiers leading to patient clinical data being sent to the wrong patient.

Improving data quality in the health care industry is linked to better overall management of health plans (Henderson in (Leitheiser, 2001)). Further, health care delivery has been under scrutiny recently following an extensive review of the quality of care provided in the United States (Institute of Medicine, 2000). Error in care provision has been found to cause significant ‘adverse events’, where the patient suffers an injury resulting from medical intervention. The report estimates that between 44,000 and 98,000 people die each year in the United States as a result of medical errors. This caused ‘shock waves’ around the western world as other nations calculated their possible death rates. There is now considerable interest in health care to improve the safety of patients through the improvement of the system of delivering care (Hagland, 2003). Whilst human error is considered a factor, often this human error is a direct result of a system or process failure (Institute of Medicine, 2000).

The practice of evidence-based medicine requires access to significant research data, collated and presented in a way that the clinician can use at the time of decision making. There are now overwhelming amounts of health research results available and continually being published, so much so that no clinician could ever remain completely up to date. In order to provide clinicians with this information, organisations such as the Cochrane Collaboration\(^5\), an international non-profit and independent organisation, provide up-to-date information about the effects of health care readily available worldwide, produce and disseminate systematic reviews of health care interventions and promote the search for evidence in the form of clinical trials and other studies of interventions. Guidelines Groups throughout the world develop evidence-based guidelines that assess policy, purchasing and practise options, and make recommendations for best practice. If appropriately implemented, an evidence-based approach will improve quality and outcomes by introducing effective care and services, while reducing unnecessary costs and ineffective care. Effectiveness takes into account:

• international and national clinical research-based evidence;
• improved patient/consumer outcomes (i.e. patient safety) and satisfaction;
• service delivery improvements;
• rational resource use.

Many of these guidelines are integrated into electronic clinical decision support tools (Institute of Medicine, 2000) and linked to the patient’s EHR. By providing access to interactive clinical decision support tools, the clinician, for example, can be alerted to potential adverse drug reactions when prescribing for a particular patient. These sophisticated tools provide information on the latest research, applicable to the patient in question i.e. what co-morbidities and current drugs is the patient on that would make the drug being prescribed unsafe? Data of high quality are required for such tools to be effective.

\(^5\) www.cochrane.org.nz
2.4 Improving Data Quality in Health Care

The health care sector’s widely decentralised and largely autonomous data collection efforts make data quality a significant challenge (Berndt, Fisher, Hevner, & Studnicki, 2001). Attempts have been made for some time to manage and control the quality of health data including data within health-based information systems. For example, Article 6 of the European Union data protection directive (Council of the European Union, 1995) relates specifically to data quality and observes that data must be:

- processed fairly and lawfully;
- collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes;
- adequate, relevant and not excessive;
- accurate and, where necessary, kept up to date;
- kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data were collected or for which they are further processed.

Given the complexity of health care and the data required, as outlined above, it is not surprising that improving data quality in health care is a complex issue. The multiple uses of the same data provide considerable difficulty in ensuring and measuring for adequate data quality in health care. Olson (Olson, 2003) talks about the ‘principles if unintended uses’, whereby you cannot separate data from its uses. To assess data quality you must undertake a thorough requirements analysis and design an assessment tool that meets these requirements. But often data collections remain in place and the environment changes. Considerable change to the management of health care has taken place over the last 15 years, particularly in New Zealand, as well as changes to legislation and the move towards evidence-based medicine. Many data collections and/or registries continue to collect data in the same format they did when developed many years previously and the use of the data has expanded into ways not thought of at the time of the database’s conception. Without good metadata, the understanding of the context of data collection is likely to be unclear (Olson, 2003).

All medical records and registries are necessarily incomplete because they reflect the selective collection and recording of data, known as the hypothetico-deductive approach (Shortliffe & Barnett, 2000). This approach is one of sequential, staged data collection, followed by data interpretation and the generation of hypotheses around possible diagnoses (differential diagnoses), leading to hypothesis directed selection of the next most appropriate data to be collected. Therefore, any data quality improvement in health care needs to consider all dimensions of data quality, for example, the relevance of missing data, or the timeliness of data depending on the context of its use. Data required by health care providers may not be required to be complete to be relevant to the delivery of care, but may be insufficient to meet the needs of government reporting or hospital management requirements. Clinician involvement in the validation and use of data has been found to be significant in ensuring and improving data quality (Croft & Williams, 2005). Whilst some data collections are managed with teams that include doctors, many do not involve nurses or allied health professionals (Black, Barker, & Payne, 2004).
A review of data quality management practices in US health care organisations (Lorence & Jameson, 2001) found that there was a technology lag and regional variation in practices, making it difficult to make comparisons with broader health sector data such as for evidence based health care. Where data quality assessment did occur within organisations, traditional audits were still preferred and system based audits were uncommon. Lorence and Jameson (Lorence & Jameson, 2001) feel the national comparison of health care data is difficult when there are no common standardised practices for data quality assessment.

### 2.4.1 Total Quality Management in Health Care

In response to the recent highlighting of poor quality patient safety (Institute of Medicine, 2000) and the drive for improved efficiency, many health care providers are implementing TQM practices to improve the organisational processes for care delivery and administrative management. For example, the National Health Service (NHS) in the United Kingdom, through the ‘NHS Modernisation Programme’[^6^], has implemented considerable changes at the local, regional and national level to the culture of health care provision and management. Similar efforts are also taking place in Health Management Organisations in the United States. TDQM (Lee, Pipino, Strong, & Wang, 2004) aligns closely with this philosophy of providing a context for data management improvements within the overall management philosophy.

‘Clinical care pathways’ are an example of quality management practices being implemented in many health care organisations. The pathways are developed through the analysis of evidence to ascertain the best health outcomes for patients following a care pathway devised for a specific diagnosis. The data obtained when documenting the patient’s progress or ‘pathway’ through the devised treatment plan provide health care service planners with clear expected outcomes, for example, average length of stay and cost of treatment. Patient data that vary from the expected outcomes are analysed to assess the reason for the variation. The variation may be unavoidable co-morbidities, related or not to the diagnosis or treatment. Data may also provide valuable information on the quality of care provided or clinical decisions made. Aggregated data from several patients on the same clinical pathway may provide information on errors in processes within the organisation that consistently cause variation from the clinical pathway expected outcomes.

Throughout the data collection and analysis process, the quality of data is paramount to providing the clinician and the organisation with a true picture of the quality of care provided. With the movement towards more customer focused health care provision and increasing consumer responsibility for health, data related to patient outcomes is sometimes being made public now (Carey & Lloyd, 2001). This enables health care consumers to make decisions about which health care organisation and even which clinician to choose when requiring treatment, giving clinicians greater impetus to collect good quality data, or at the very least to consider these data more carefully.

Given the considerable importance of high quality data to the management of health care, any TQM programme would by necessity include data quality improvement. TDQM provides a structure to prioritise

data quality improvement in a complex environment where cost containment can considerably restrict the choices available for improvement projects.

**2.4.2 Improving Data Quality in the EHR**

Given the considerable interest in the development and implementation of EHRs, it is not surprising that studies are being undertaken into ensuring and improving data quality in the EHR. De Lusignan, Stephens, Adal and Majeed, and Fine, Keogh, Cretin, Orlando and Guild, and Lorenzoni, Da Cas and Aparo (De Lusignan, Stephens, Adal, & Majeed, 2002; Fine, Keogh, Cretin, Orlando, & Gould, 2003; Lorenzoni, Da Cas, & Aparo, 1999) all found that continuous monitoring and feedback comparing a clinician’s performance with other clinicians proved to be successful in improving the quality of data on medical records. More detailed feedback appears to have a greater effect (Gribben et al., 2001). This is supported by the findings of the UK Audit Commission (UK Audit Commission, 2004) who noted data quality in health care was hampered by a lack of understanding of the importance of the role of the data collector to the downstream quality and the usefulness of the data.

Data capture, manipulation and retrieval at the locations where they are generated and/or utilised is also necessary if an EHR system is to be integrated into the environments within which it is used and into the working practices of its users (Haan et al., 2004). Mobile technology (palmtop, pocket PC, mobile phones, etc.) is the key to achieving this (Orfanidis et al., 2004). Preventing the need to write data on paper based forms to be entered later means data are timelier and input errors are less likely to occur. Clear data definitions are also required to ensure all are interpreting fields correctly (Thiru, Hassey, & Sullivan, 2003).

Giving patients access to their records is a key part of EHR management strategies. For example, the NHS plans that every person in the United Kingdom will have his or her own record. Patient access to records is important, as it allows them to take control over their own health and make decisions about their health care, and also improves data quality (NHS Information Authority, 2003). In particular, this provides patients with chronic illnesses the opportunity to get involved with the management and treatment of long-term conditions (Orfanidis et al., 2004). A research project that allowed parents of paediatric patients to view and change their child’s records found ‘improved sensitivity for detection of historical risk factors for illness can be achieved by augmenting the paediatric EHR with a section for direct parental data input’ and significantly improved data quality (Porter & Mandl, 1999).

However, a study in New Zealand revealed that in general the public are unaware that health information is held about them and that they can view and make amendments to that information (Ryan, 2004), indicating considerable public awareness education is required if patients are to have understanding about input into their medical records.

Orfandis et al (Orfanidis et al., 2004) and Dugas, Hoffman, Janko, Hahnewald, Matis and Uberla (Dugas, Hoffman, Janko, Hahnewald, Matis, & Uberla, 2001) in their studies of data quality in EHRs, found that analysis and design of data quality issues are an integral part of the development of an EHR and should be
addressed from the inception of the project, but that there are currently technological limitations to some of the issues, such as storage requirements, which still need to be resolved.

### 2.4.3 Clinical Research

Clinical research often utilises the results of data collected from international clinical trials, for example from double-blind drug trials using ‘gold standard’ research methodologies, usually with many thousands of participants from several countries taking part over 3-5 years. Increasingly, high quality data are expected by authorities and consumers before new drugs or new uses for existing drugs are approved. These trials are now utilising the World Wide Web to enable investigators to directly input clinical trial data at the point of observation. Marks, Conlon and Ruberg (Marks, Conlon, & Ruberg, 2001) found the accuracy and timeliness of the data improved considerably with a simplified data collection process when using the web to collect clinical data. Arts et al (Arts et al., 2002) also note the requirement for good study design, with extensive training for data collectors to ensure the prevention of errors. In addition detection through routine monitoring and action on errors is required throughout the study. Measures need to be taken at the local data collection source and standard definitions of data quality and data quality attributes are necessary to be able to compare data quality among research registries or within a registry at different points in time (Thiru et al., 2003).

### 2.4.4 Health Information Standards

Within health care systems data can be documented in many different ways. A particular level of detail may be chosen; synonymous labels may be selected, spelt differently and structured according to the author’s preference. This causes considerable problems in subsequent data analysis. Retrieval is harder and less reliable; the use of homonymous terms may lead to the selection of irrelevant data objects when searching data collections. It may be nearly impossible to count the frequency of certain similar objects (such as diagnostic categories for administrative or scientific purposes), because the terms used do not indicate the degree of similarity between them (Liener et al., 2003).

Standards facilitate the sharing and exchange of information between departments, health agencies and health workers (Hovenga et al., 1996), providing a defined set of rules, conditions or requirements concerning definition of terms, classification of components, specifications of materials, performance of operations, and delineation of procedures. International and national standards developments are providing the benchmarks for data access and storage. Health Level 7 (HL7) is emerging as a significant international health data and messaging standard that includes the development of a standard format for the content of the EHR.

Providing international standards allows for the movement of health data across any geographical boundary. Murray, Lopez and Wibulpolprasert (Murray, Lopez, & Wibulpolprasert, 2004) note the importance of internationally agreed data standards to enable improved global health monitoring. The increasing importance of monitoring epidemics globally was recently highlighted by the SARS virus epidemic and its spread to distant continents within a very short space of time requiring appropriate international responses.
The benefits of standards development in improving data quality are considerable. The ability to transfer data throughout the health sector is greatly increased (Kenney & Macfarlane, 1999), reducing the requirement for subsequent mapping of data to different coding or classification standards when reaching its destination (Shortliffe & Barnett, 2000). All who send and receive data have a consistent understanding of its meaning. Agreed standards can then be included into a ‘metadata’ repository, providing the format for storage of metadata which may be applicable within the organisation, the region or nationally.

2.4.5 National Data Quality Improvement Strategies in Health Care

Government departments of health are beginning to recognise the need for national strategies that provide for the consistent improvement of data across the national health sector. A review of international data quality improvement strategies in health care, including the NHS in the United Kingdom (United Kingdom Department of Health, 2004), the Canadian Institute for Health Information (CIHI) (Canadian Institute for Health Information, 2003b), HealthConnect Australia (Australian Department of Health and Aging, 2003) and the United States Department of Health and Human Services (United States Department of Health and Human Services, 2002) identifies similarities between the strategies. All the reviewed international strategies note the complexity of data quality improvement in their multi-level, multi-dimensional initiatives. Principles are based on the need to manage data proactively to ensure ongoing integrity by preventing data quality problems, using a systematic TDQM approach. There is also commonality of role expectations – the data suppliers are responsible for the quality of the data they provide to central government, while central government is required to provide leadership and assistance to data suppliers by developing sector-wide standards and best practice guidelines. A common theme underlying many of their quality issues has been previous low levels of investment in information technology in health care.

The UK National Health Service Information Authority (NHSIA) views data quality as an intrinsic part of the Information for Health Strategy (NHS Information Authority, 1998). Data quality is seen as critical to ensure the accuracy and timeliness of data that ‘supports information used for improving health and more effective management’ (NHS Information Authority, 1998). A vital aspect of the UK’s strategy revolves around the ability of the health professional to access consistently recorded and coded clinical information. Standards are clearly defined and applied to ensure consistency such that comparisons can be made both across time and between data sources. Despite data quality underpinning the strategic use of information in health, the UK approach to data quality is focused on data entry and consistency of initial coding. There is little consideration of data abstraction, relying instead on the extension of the World Health Organisation’s ICD-10 coding classification through the use of internationally agreed metadata standards. ICD-10 has also been adopted as a standard coding model used in Australia (Australian Department of Health and Aging, 2003) and New Zealand.

The NHSIA (NHS Information Authority, 1998) sees high quality data being produced as part of the routine daily activity within a hospital or general practice surgery. The NHSIA data quality framework is then seen as providing confidence in their findings, providing necessary information to make relevant changes and
achieve continuous quality improvement. The framework outlines clearly, and in detail, substantial work that is required by health care providers to ensure good data quality. The NHS developed an accreditation scheme that was initially thought to be all that would be required to ensure the supply of good quality data. The scheme is extensive and was found to be very successful but did not sufficiently identify the responsibilities of the data supplier; central government was still monitoring more than it was leading. This led to the more extensive guideline development on the principles and theories of data quality supported within the NHS. Several NHS Trusts have developed and published, on their websites, data quality strategies that align to the central NHS core strategy requirements.

A recent review of the programme by the UK Audit Commission (UK Audit Commission, 2004) found significant improvements to levels of data quality in the NHS. However similar issues are still apparent after five years of targeted improvements and the report recommended:

- developing a more co-ordinated and strategic approach to data quality;
- developing a NHS wide strategy for specifying, obtaining and using both national and local information;
- making more and better use of patient-based information;
- involving NHS Trust board members;
- training and developing staff;
- keeping systems up to date.

The CIHI also briefly discussed accreditation for enhancing collaboration with data suppliers (Canadian Institute for Health Information, 2003b). They have undertaken extensive work on data quality through collaborative work with experienced statisticians from Statistics Canada. They base their theories on research by the Massachusetts Institute of Technology in the United States. The CIHI Data Quality Strategy focuses on the following areas:

- fostering a data quality culture within CIHI and the health sector;
- strengthening data quality infrastructure and capacity;
- cultivating the data supply chain;
- enhancing external collaboration;
- establishing dedicated funds for fast track priority projects;
- communicating strategies and action plans, importance of data quality, opportunities for collaborative action and progress achieved.

The HealthConnect model (Australian Department of Health and Aging, 2003) is essentially the requirements for the receipt, storage and retrieval of clinical and demographic data. It is recognised that in order to be a success, HealthConnect will address a number of key data issues such as:

- validity and integrity of the data;
- completeness of data;
- responsibility for data
• controlling access to data;
• errors in data;
• corrections to data;
• data consistency and standardisation;
• data definitions and descriptors;
• ownership of data.

The series of projects related to HealthConnect are still in the pilot phase. The researcher was unable to find any evidence of the development of a specific strategic data quality plan for ensuring good data quality for the HealthConnect model.

The Department of Health and Human Services in the United States (United States Department of Health and Human Services, 2002) disseminated very high level guidelines to those who manage health care data, no specific data quality strategy was found by the researcher. The guidelines pertained to the quality for information disseminated to the public and all health care providers were requested to develop their own guidelines by the Office of Management and Budget (OMB). The guidelines noted that information should be accurate, reliable, clear, complete, unbiased, and useful. Guidelines were provided on the policies and practices that support information dissemination. Data quality is to be treated as an integral step in every aspect of the information development process, but organisations also need to weigh the costs and benefits of quality. Organisations should:

• make data management methods transparent by providing documentation;
• ensure quality by reviewing methods, consulting with experts and users, notifying users about corrections and revisions;
• make data and information supported with public funds available to the public consistent with confidentiality concerns and resource availability;
• review all data before dissemination.

2.5 Chapter Conclusion

Data quality is now becoming an increasingly important issue for health care providers, managers and government departments. The movement towards total quality management in health care to improve patient safety and health care efficiency is demanding high quality information. Further, evidenced-based care requires the assimilation of large amounts of relevant research data to be available at the point of clinical decision making.

The increasing movement of data through electronic innovations is highlighting the contextual nature of data in health care. There is a need for good metadata regarding the context of its collection to ensure applicable use of the data and to prolong the usefulness of the data. The development and implementation of national EHRs is providing clinical decision support and integrated care that demands high levels of data quality. Consumers themselves are demanding greater access to their health data. Whilst this is likely to improve the
quality of the data it also places greater demands on the data manager to maintain high quality data to ensure consumer confidence in their health care provider.

Methods engaged outside of health care are proving valuable in improving health care data. Many of the issues found in other industries are found in health care, but the complexity of health care provision means that data quality improvement requires a multi-faceted approach. Strategic prevention, national consistency of improvement practices and data standards, and targeted improvements with increasing consumer involvement are moving health care towards a TDQM approach to data quality management.

Much of the current literature on data quality in health care focuses on data quality improvement for single data collections or one-off projects for multiple collections. Further academic research is required to elicit the theoretical underpinnings on which to understand the needs of health care in the management of data quality on a systemic and strategic level. This will enable increasing data mining and assessment of data across not only multiple organisations, but also nationally and internationally.

It can be said then, that ensuring sufficient data quality in health care can impact over all individual health care management and decision making in the areas of:

- national and regional service delivery decision making, impacting on health economics;
- patient safety whilst undergoing care;
- evidence to support clinical decision making through health care research;
- information provided to patients on their illness and care;
- effectiveness of clinical care pathways;
- movement of data across government departments to provide a holistic view of health care requirements.
3 The Structure of the New Zealand Health Sector

This brief Chapter provides a summary of the structure of the New Zealand health sector to enable the reader to better understand the results of the research, provided in Chapters Five and Six.

3.1 Health Care in New Zealand

New Zealand has a population of four million people spread over 270,000 square kilometres. Eighty five percent of the population are concentrated in the urban areas. The main ethnic groups are European 71%, indigenous Maori 14.5%, Pacific Islanders 4.8%, and Asian 4.5%. The cornerstone of New Zealand’s health system is public finance through taxes with access to health services based upon need. In New Zealand in 2000/01 $9.884 billion, including private funding, was spent on health and disability support services (NZ$2601 per capita). Of this, $2.3 billion (23.3 %) was privately funded and $7.584 billion (76.7 %) publicly funded (Ministry of Health, 2003b). Since 1999 the New Zealand Government has been moving away from market-based structures in the health sector, by combining the health care purchaser and provider functions into community-focused District Health Boards (DHBs) (Kerr, 2004). An illustration of the structure of the health sector is provided in Figure 2 below.

Central government provides broad guidelines on what services the DHBs must provide, and national priorities have been identified in the New Zealand Health Strategy (Ministry of Health, 2000). A range of providers including public hospitals, non-profit health agencies, or private organisations can deliver services. Funding is allocated to DHBs using a weighted population based funding formula.

DHBs are responsible for planning, funding and ensuring the provision of health and disability services to a geographically defined population. This reflects a move away from the previous purchaser/provider split, as DHBs provide hospital (and some community-based) services. DHBs are responsible for improving, promoting and protecting the health and independence of their populations. Boards must assess the health and disability support needs of the people in their regions, and manage their resources appropriately in addressing those needs. DHBs vary considerably in size. There are six tertiary DHBs and 14 secondary DHBs. The largest DHB provides services for 477,000 patients, with the smallest providing services for 30,300 (Ministry of Health, 2003a).

It is important to note that the management and institutional knowledge within DHBs is rapidly evolving. DHBs are relatively new organisations, having been established in 2001 with considerably different roles and culture to their previous entities, where competition for health dollars was actively encouraged. This had the flow on effect of discouraging collaboration with other health care providers.

Under the New Zealand Public Health and Disability Act (2000), the Minister of Health is required to determine a New Zealand Health Strategy to provide the framework for the Government’s overall direction of the health sector in improving the health of people and communities.
The seven principles of the New Zealand Health Strategy (Ministry of Health, 2000) are:

1. acknowledging the special relationship between Maori and the Crown under the Treaty of Waitangi;

2. good health and well being for all New Zealanders throughout their lives;

3. an improvement in health status of those currently disadvantaged;

4. collaborative health promotion and disease and injury prevention by all sectors;

5. timely and equitable access for all New Zealanders to a comprehensive range of health and disability services, regardless of ability to pay;

6. a high-performing system in which people have confidence;

7. active involvement of consumers and communities at all levels.
Figure 2: Structure of the New Zealand Health Sector

http://www.moh.govt.nz/moh.nsf/e00eda991ab5e3704c256670004079ba/d4cd4c98ad5e8994cc256be2007eb883?OpenDocument
3.2 The Structure of the Ministry of Health

The Ministry of Health is the Government's principal agent and advisor on health and disability. It develops policy advice for the Government on health and disability issues, administers health regulations and legislation, funds health and disability support services, plans and maintains nationwide frameworks and specifications of services, monitors sector performance and provides information to the wider health and disability sector and the public. Figure 3 below outlines the structure of the Ministry of Health and its relationship with the health sector as funders and providers of health care services and suppliers of data to the national health data collections. The New Zealand Health Information Service (NZHIS) and the Health Payments, Agreements and Compliance unit (HealthPAC) are business units within the Corporate and Information Directorate. Both NZHIS and HealthPAC provide and receive a considerable amount of data from and to the health sector and the Ministry of Health. This research pertains to the national health collections held at NZHIS, therefore the function of NZHIS is provided in detail below.

![Diagram of the Ministry of Health structure]

**Figure 3: Structure of the Ministry of Health**

3.2.1 The New Zealand Health Information Service

The New Zealand Health Information Service (NZHIS) is a specialised group, within the Ministry of Health, responsible for the collection and dissemination of health related data. NZHIS has as its foundation the goal of making 'fit-for-purpose' information readily available and accessible in a timely manner throughout the health sector. The vision of NZHIS is to be a leader in the provision of health information services in New
Zealand, and to be recognised and respected as a leading organisation internationally. NZHIS has responsibility for:

- the collection, processing, maintenance, and dissemination of health data, health statistics and health information;
- the ongoing quality improvement of data entering the national health information systems;
- the continuing maintenance and development of the national health and disability information systems;
- the provision of appropriate databases, systems and information products;
- the development and provision of health and disability information standards and quality audit programmes for data;
- coordination of ongoing national health and disability information collections and proposals for their development;
- analysis of health information, performance monitoring, benchmarking, and advice on the use of information obtained from NZHIS.

NZHIS and the sector supplies and receives data from HealthPAC, which provides services to health funders that include:

- establishment and administration of agreements;
- payments to health providers for contracted services;
- payment and clinical data collection from health provider claims;
- provision of information and reports relating to payment and other health data;
- audit and counter fraud methodologies to ensure health funds are applied legitimately and appropriately;
- patient eligibility administration.

### 3.2.2 Health Information Management in the New Zealand Health Sector

Historically, there has been no sector wide approach or consistent approach to developing health information systems in New Zealand. However, since 2001 DHBs have been replacing isolated departmental and clinical systems with more integrated and dynamic web-based technologies that support a more connected information delivery network. All of the 21 DHBs have entered into some form of shared service arrangements for information systems (e.g., finance and/or patient management systems), corporate support (e.g., health provider contract management) and contracting or clinical data analysis. Such arrangements have reduced duplication and contributed to more effective and efficient management of infrastructure with greater interoperability. Information technology has been essential for implementing the population health care initiatives in New Zealand (Kerr, 2004).

The Health Information Standards Organisation\(^8\) (HISO), a Ministerial Committee formed in 2003, has a key coordination role in leading the development and implementation of information management and technology

\(^8\) [http://www.hiso.govt.nz/](http://www.hiso.govt.nz/)
standards for the New Zealand health sector. The HISO ensures that relevant standards are identified for development and that, once defined, are implemented effectively for the overall benefit of the health sector (Kerr, 2004).

The Health Information Strategy for New Zealand (HIS-NZ) was published in 2005 and provides a context to support New Zealand health and disability strategies to make innovative use of information. It builds on previous national health information strategies, including the 1996 Health Information Strategy for the Year 2000 (Ministry of Health, 1996) and the WAVE (Working to Add Value through E-information) (WAVE Advisory Board, 2001) Project. Figure 4 below provides a summary of the linkages between health strategies and information strategies, and highlights the importance given to information to support strategic goals. The importance of information is also noted in the overarching New Zealand Health Strategy (Ministry of Health, 2000) with requirements of the health sector to:

- exchange high quality information;
- have better access to timely and relevant clinical information;
- have a nationally coherent and consistent approach to health information infrastructure, based on improving access to information and the consolidation of appropriate standards.

HIS-NZ provides direction to the health and disability sector in making better decisions about how to improve the quality and availability of health information and is intended to develop a single and co-ordinated strategy of information systems, including current and emerging information, communications and technology within the health sector and provides a context for increasing information systems capability. The focus on improved information requires improving data quality management to address all of the strategy’s 12 action zones, particularly to improve data quality on the National Health Index (NHI), and subsequently reduce duplication of data collection.

Government has an obligation that the health status of Māori is comparable with non Māori, as well as a commitment to self-determination such that Māori can determine what health care services are appropriate and how they are delivered (WAVE Advisory Board, 2001). For Māori, data are embodied with significant spiritual and cultural significance. This is the case regardless of whether data are personal and identifiable or not. The governance process surrounding the collection, storage and use of data pertaining to Māori requires additional attention and processes can and should differ from those traditionally employed in the health sector. It is, for example, possible to apply the concept of kaitiakitanga, meaning guardianship, protection, care and vigilance to data management. This approach offers the potential to locate Māori as the primary beneficiaries of data relating to themselves, to ensure that data collected are appropriate and relevant and that issues of collective ownership, and collective privacy are addressed (Kamira, 2003).
3.3 Data Quality at NZHIS

The manager of the Clinical Analysis Team at NZHIS (responsible for data quality) reports directly to the group manager, as do managers from the Business Intelligence Unit, the Clinical Coding Team, the Operations Team and the Chief Advisor, Health Information Strategy and Policy. Each of these teams within NZHIS has considerable impact on the quality of data management for the health sector, and these teams work closely together. However, all require expert advice if the organisation and the sector are to make significant and effective improvements in data quality levels. Figure 5 is a diagram of the data flows within NZHIS, highlighting the complexity of data management processes.

The NZHIS Clinical Analysis Team has nine staff working on Data Quality, two work specifically on the Mental Health Information Collection (MHINC). According to management ‘as many as 15 people work on data quality within NZHIS, but they are working on the wrong things’. A detailed analysis of data quality in the New Zealand health sector is provided in Chapter Six.
Figure 5: NZHIS Core Business Flowchart

- **DHB Primary**
- **PHOs**
- **IPAs**
- **Public Hospitals**
- **DHB Secondary/Tertiary**
- **Transactions**
- **Warehouses**
  - **NMDS WH**
  - **MHINC WH**
  - **MNIS WH**
  - **NRBS WH**
  - **GMS WH**
  - **Lab WH**

- **National Health Index**
- **GHE**
- **HI & x25**
- **HI & H/Link**
- **HI & Disk**
- **Paper/Disks**
- **Paper & Disks**
- **Disk, Paper**
- **Disk, e-mail, paper**
- **Disk, Paper**
- **E-mail, Paper**
- **Paper**
- **Paper**
- **Paper**
- **Paper**

- **Other Primary Health**
- **Community Pharmas**
- **Community Labs**
- **Histology**
- **Cytology**
- **Private Hospitals**
- **National Screening Unit**
- **Coroner**
- **BDM**
- **LTSA**
- **Registered Health Professionals Survey**

- **HealthPac**
- **NGO**
- **DHB**
- **FTP/Disk**

- **Public Hospitals**
- **Hospital Labs**
- **Health Workforce**
- **Stand Alone Databases (13)**

- **Medical Warning System**
- **Mortality Collection**
- **National Minimum Data Set**
- **National Health Index**
- **FTP/CD**
- **FTP/CD**
- **FTP/CD**
- **FTP/CD**

- **Transactions**
  - **SID**
  - **ASE**
  - **NMDS**
  - **MHIN**
  - **MNIS**
  - **NBRS**
  - **NZCR**
  - **Cancer**

- **Health Events**

- **Figure 5: NZHIS Core Business Flowchart**
3.3.1 The National Health Index

The National Health Index (NHI) is a central data set, the function of which is to uniquely record every individual who is either eligible for health and disability services or has received care and treatment in the New Zealand health system. The NHI is a single point of identification of individual-level data on health service delivery and as such it forms a critical backbone of any health information system. Unfortunately, while the concept is sound in principle, in reality the quality of data in the NHI is often poor such that:

- data are duplicated;
- address and geo-coding information is incorrect, missing or inadequate;
- ethnicity information is missing or incorrect.

NZHIS is currently involved in a programme to remedy the above problems and upgrade the NHI such that:

- access to the NHI, especially for primary health care, is improved;
- public understanding and knowledge of the NHI is increased;
- everyone using health services in New Zealand has an NHI number.

An extract from the Health Information Strategy for New Zealand (Ministry of Health, 2005) highlights the importance of good data quality on the NHI:

As the sector needs to rely more on the NHI and the associated national clinical data collections, it is important that programmes to improve NHI data continue. The NHI can be more effective as a unique identifier for tracking service delivery across a continuum of care, but requires additional work to resolve duplicates and improve access by key parts of the sector who either do not have access to the NHI or are currently not able to register and update clients on the NHI. As the NHI was not originally designed to be used as an identifier for population-based health boundaries (e.g., for enrolment of patients in Primary Health Organisations or for particular disease groups or risk groups), some remedial work needs to be carried out. (Pg 23)

This Chapter provides a description of the New Zealand health sector. In particular, how health information is managed within the sector and the strategies that guide this management are noted. The aim of this Chapter is to provide context to the reader for the research results and discussion Chapters that follow.
Part B Research Methodology
4 Research Methodology

This chapter details the research methods used throughout the research. Further detail of the methodology used for the analysis of the data is provided in Chapters Five and Six. The research utilises action research to provide structure to the research design. The qualitative data collected was analysed using grounded theory. Pauleen and Yoong (Pauleen & Yoong, 2004) also used this methodology in action learning research and note:

while action learning allows for a creative and flexible approach to gathering field data, the grounded theory methods promote analytical rigour and validity. The result, grounded action learning, is a symmetrical and harmonious melding of two research methods that has great potential in information systems research to promote theory development.

Figure 6: The Combination of Action Learning and Grounded Theory (Pauleen & Yoong, 2004)

The researcher found the context of this research lent itself to the combination of these two methodologies, with the aim of developing new theory, whilst bringing about organisational learning and process change. Data Generation was provided through focus groups and participant observation, rather than the training programmes used by Pauleen and Yoong (Pauleen & Yoong, 2004). Figure 6 above illustrates the process of data generation and theory development through using a grounded theory approach to analyse data.

4.1 Research Paradigm

The research sits within the paradigm (a way to explain the different ways that people see and make sense of the world) of interpretive research. Interpretive research attempts to understand phenomena through the meanings that people assign to them (Plummer, 2001). This research also uses some positivist methodologies within survey design and analysis. The combination of both positivist and interpretive methodologies is thought to address the criticisms brought upon both methodologies (Plummer, 2001). Interpretive research is perceived to lack methodological rigour and generalisability and positivist methods are unable to account for the complexity and subjectivity of the social world.
Miles and Huberman (Miles & Huberman, 1994) have found there are recurring themes in interpretive research. The researcher is present in the field for a prolonged and/or intense period, the researcher’s role is to gain an holistic view of the context of the study, thereby capturing the perceptions of the participants and analysing the data (usually words) for themes and expressions used by the participants and relatively little standardised instrumentation is used, with the researcher as the main ‘measurement’ device (Miles & Huberman, 1994). The main intent is not to identify independent or dependent variables, rather to make meaning out of the situation as the analysis of the data unfolds (Plummer, 2001). The research uses a hermeneutic approach, the theory and practice of interpretation, to understand the spoken text through the lived experience of the participants (van Manen, 1997).

In this research the discovery of regularities or the discerning of socialised patterns is found in the data through a process of critical inquiry. The philosophical underpinnings of personal development for all research participants, including the researcher, empower all to make change. Further, the research reveals ways to help participants see a better way of doing things. Health informatics is an applied discipline, therefore, health informatics research should be designed to improve practice and should be practice driven (Plummer, 2001).

Ethnography, a research method utilised in this research, relies on first-hand observations made by a researcher immersed over a period of time, typically in a culture with which the researcher is familiar (Schultze, 2001). Researchers ask such questions as ‘what do people do here?’ in contexts where people hold a certain kind of knowledge, ways of doing things, and perceptions (van Manen, 1997). Through the extended observation, immersion, and participation in the culture, the researcher seeks to discover and understand rules of behaviour (Spradley, 1980). Ethnography is a research design that is capable of moving beyond description to reveal complex relationships and theories. Shultze provides a clear guideline for researchers of information systems wishing to use reflexive ethnology:

*Ethnography can be used in a reflexive context, whereby the researcher considers their interaction with the object of study, openly acknowledging the role of the ethnographer throughout the fieldwork. Therefore the researchers’ reactions to situations are considered legitimate data.* (Schultze, 2001)

Data were collected by the ethnographic methods of focus groups, workshops, semi structured face-to-face interviews, written survey questionnaires, email and organisational documentation, participant observation, and personal discussions to study culturally shared perceptions of everyday experiences (van Manen, 1997). Mumford (Mumford, 2001) suggests all of these methods of data collection provide valuable comparative data that can help to confirm or challenge research questions. The use of several methods of data collection allowed the researcher ‘data triangulation’. Different kinds of data give the researcher different views or vantage points from which to understand a concept and develop its properties and enhances the construct validity of the research (Pandit, 1996). The predominant research methodologies used in this study are qualitative. The methodologies were chosen for their abilities to provide the appropriate answers to the research questions, having been previously utilised by researchers in similar studies (Avison & Myers, 1995), (Lee, 2004), (Pauleen & Yoong, 2004). Further, the methodologies were utilised as change agents, whereby the process of the research brought about increased awareness of data quality theories and techniques amongst the participants.
The research initially used selective sampling, whereby the sampling was guided by the purpose of the study (Bowers, 1988). The Clinical Analysis Team and the Ministry-wide Data Quality Team were chosen as the initial sample due to their knowledge and roles within the Ministry. The research then used theoretical sampling (Strauss & Corbin, 1998), (Strauss, 1987) within the Ministry to elicit data from specific role perspectives, for example, Managers of NZHIS were interviewed to understand their level of support for change. The development of the strategy used the same sampling methodology, with the initial selective sampling of DHB CIOs and the subsequent theoretical sampling of DHB Data Quality Teams, PHOs and Shared Services Agencies, as the importance of their roles emerged from the initial coding of data.

A checklist matrix was utilised by the researcher. This is a format for analysing field data on a major variable or general domain of interest (Miles & Huberman, 1994). This helped to ensure that the findings made by the researcher were in fact ‘grounded’ in the data, and supported by the triangulation of evidence where possible. The methodologies used in this research are discussed below.

4.2 Research Context

This research studies learning and change in a real organisational setting (the New Zealand Ministry of Health), with the awareness of potential for change across the entire New Zealand health sector. In order to satisfy the aim of the research and to answer the research questions, the researcher utilised action research methodology to provide a structured and iterative cycle of learning and improving through participatory observation. The research took place over a two-year period, with the researcher embedded in the process of development and the initiation of change. Longevity and participatory action are fundamental to action research (Mumford, 2001), (Baskerville & Wood-Harper, 1996). The data gathered provided for inductive analysis and interpretation leading to iterative cycles of research as learning took place and initial theories were tested against the existing literature (Urquhart, 2001) and within the research organisation.

The researcher’s role within the organisation was one of health information strategy and policy development. This meant the researcher was limited to the development of the strategy and was not expected to implement the strategy; this was the role of the Clinical Analysis Team at NZHIS.

Ethics approval was sought and obtained from the University of Auckland Human Subjects Ethics Committee. All participants were provided with an information sheet about the study prior to data collection and signed consent forms agreeing to take part. Approval was sought and obtained from management at the Ministry of Health to undertake research within the organisation and to solicit members of the organisation as research participants. Approval was also sought and gained from all participants to record interviews and focus groups using audiotapes to enable accurate transcribing later.

4.2.1 Action Research

Action research provided the most applicable research methodology to answer the initial and subsequent research questions and to bring about the beginnings of institutional change across the health sector. Action research is a form of collaborative social research (Miles & Huberman, 1994). As noted earlier, data quality
improvement is most often a ‘people’ issue. In this particular study, the researcher, along with the study participants, took part in an educational and change initiation process through two cycles of action research.

The essence of action research lies in its objectives of both advancing theories in research as well as facilitating organisational change (Lee, 2004), (Mumford, 2001). Action research methods are highly pragmatic in nature, and place information systems researchers in a helping-role within the organisations that are being studied (Baskerville & Wood-Harper, 1996). Baskerville and Wood-Harper (Baskerville & Wood-Harper, 1998) note that action research characteristics are orientated toward the research process rather than merely the outcome of the research. The social interaction that takes place throughout the action research process was considered important in this research, as the research needed to bring about organisational learning and change. Action research has been found to be highly suitable for the research of data quality improvement in a real organisation and has been used in several research projects by Lee (Lee, 2004), (Lee et al., 2004), and Avison and Myers (Avison & Myers, 1995). Lee found the methodology merges research and practice, which in turn produces significantly relevant information systems research findings (Lee, 2004).

Action research introduces changes into complex social processes and observes the effects of these changes. The outcome is increased understanding of the social situation, practical problem solving and an expansion of scientific knowledge. The knowledge obtained by the researcher needs to be applied immediately, providing research that links theory with practice (Baskerville & Wood-Harper, 1996). Much of the work of the data quality practitioner is one of change leader, facilitator (for example, of cross organisational data quality groups), and teacher. Clark notes ‘for convenience it is useful to think of the practitioner as part of a set of actors who are oriented to solution of practical problems, who are essentially organisational scientists rather than academic scientists’ (Clark, 1972, p. 65 in (Baskerville & Wood-Harper, 1996). The action researcher must agree to a set of rules to ensure a collaborative framework for action with those already working in the organisation. In this case, the researcher was an employee of the organisation and therefore bound by policies of the organisation and the expectations of management.

There are several forms of action research, outlined by Baskerville and Wood-Harper (Baskerville & Wood-Harper, 1998):

- Canonical;
- IS Prototyping;
- Soft Systems Methodology;
- ETHICS;
- Multiview;
- Action Science;
- Participant Observation;
- Action Learning;
- Clinical Field Work;
- Process Consultation.

Baskerville and Wood-Harper (Baskerville & Wood-Harper, 1998) also outline the characteristics of IS action research:
The most prevalent description of action research is from Susman and Evered (Susman & Evered, 1978), and details a five-phase, cyclical process that can be described as an ‘ideal’ exemplar of the original formulation of action research, and a form of Canonical action research. In practice such methods often vary depending on the application. This ideal approach first requires the establishment of a client-system infrastructure or research environment. Then, five identifiable phases are iterated:

1. diagnosing (the organisation's reason for wanting change, providing a working hypothesis);
2. action planning (to relieve the found problems guided by the theoretical framework);
3. action taking (active intervention by the researcher and members of the organisation);
4. evaluating (by the researcher and participants);
5. specifying learning (an ongoing process).

Figure 7: Action Research Structural Cycle (Canonical Action Research) (Susman & Evered, 1978)
This research utilises Canonical action research. The process model is iterative, whereby the research utilises a repeating sequence of problem diagnosis and action. The client-system infrastructure is the agreement between the researcher and the New Zealand Ministry of Health, providing the authority to undertake the research. The structure of the research is relatively structured, characterised by delineated stages that are carried out in sequence, close to that of the Susman & Evered (Susman & Evered, 1978) cycle noted in Figure 7. The nature of involvement of the researcher was collaborative, working alongside the research participants as a co-worker. The goals of this research are to facilitate organisational development, through improving the management of data quality within NZHIS and the health sector, and to provide understanding that increases scientific knowledge in the field of strategic data quality management in health care.

Argyris and Schön (Argyris & Schön, 1978) call step five double-loop learning, where the restructure of organisational norms reflects the new knowledge gained by the organisation during the research. Second, where the change was unsuccessful, the additional knowledge may provide foundations for diagnosing in preparation for further action research intervention. Finally, the success or failure of the theoretical framework will provide important knowledge to the scientific community faced with future research settings.

Baskerville and Wood-Harper (Baskerville & Wood-Harper, 1996) note that action research is one of the few research methodologies where the researcher legitimately actively participates and intervenes in the change process, ensuring relevance to the real world, requiring rigorous intervention and providing for the study of human methods. Figure 6 is a diagram of this action research structural cycle as described by Susman and Evered (Susman & Evered, 1978). Action research enables both qualitative and quantitative data collection methods. Participant observation that allows the researcher to deliberately join with the activities of a group while observing them (Langley, 1994), can be accompanied by in-depth interviews, the collection of statistics and the analysis of relevant documents (Mumford, 2001). The cyclic nature of action research means that the initial cycle of research produces considerable raw data that require detailed analysis to inform the next cycle of research. The analysis phase within action research lends itself to the application of grounded theory techniques.

### 4.2.2 Grounded Theory

The researcher found that grounded theory techniques provided a structured method for the analysis of the data. Emerging theories became clear through the coding process and required that the researcher be highly cognisant of the data through the detailed transcribing and coding process. Topics and questions used in subsequent interviews, workshops or focus groups were drawn from initial data collected. Initial coding provided for the raising of questions and the giving of provisional answers (hypothesis) about coding categories and their relationships.

Grounded theory as a research methodology was developed to provide researchers with a methodology to systematically derive theories of human behaviour from empirical data. The researcher is required to set aside theoretical ideas to allow a substantive theory to emerge (Dey, 1993). The methodology allows for the inductive development of theory, using the current literature to substantiate his/her own findings throughout
the research process (Urquhart, 2001). Grounded theory is a method that can be used in most research paradigms and in information systems research has been used extensively in an interpretive context (Urquhart, 2001). Grounded theory methods sit within the theoretical framework of symbolic interactionism, where the actions of the individual in his/her roles are socially constructed by the context of their environment. Symbolic interactionism discovers the reality of the way an individual or group defines and experiences their world; you cannot extract the individual from their social context (Bowers, 1988). This research utilised the methodologies commonly used in grounded theory to understand the processes of development and change the research participants journeyed through, as their level of understanding of data quality theories increased.

Grounded theory has developed considerably since its inception in 1967 by Glaser and Strauss (Glaser & Strauss, 1967). Since then there has been a division between the two authors, with Strauss attempting to develop a more structured methodology of a coding paradigm that enables the researcher to make connections between data categories (Strauss & Corbin, 1998). Glaser felt this no longer allowed the analysis of the data to emerge naturally and this was reiterated practically through the experience of Urquhart (Urquhart, 2001) when attempting to categorise coded data. This research utilises the guidance of Glaser and Strauss (Glaser & Strauss, 1967), Strauss and Corbin (Strauss & Corbin, 1998) and Dey (Dey, 1993) in initial coding as they provide practical guidance for the initial coding phase. Further, Strauss’s 1987 book notes grounded theory techniques should be used as guidelines and should by no means be regarded as hard and fixed rules for converting data into effective theory (Strauss, 1987). Modifications include the use of some existing theories to provide a basis for the initial engagement of the organisation to embark on some form of structured data quality improvement programme, the development of high level research questions and the initial research design.

Grounded theory is detailed grounding by systematically and intensively analysing data, sentence-by-sentence, or phrase-by-phrase (Strauss & Corbin, 1998). By constant comparison data are extensively collected and coded (Strauss & Corbin, 1998) and then recoded or not as applicable, by constantly comparing with new or other data that have been labelled as a particular category. When no new categories or changes to existing coding are required, ‘theoretical saturation’ is reached (Strauss & Corbin, 1998). Grounded theory techniques were chosen to develop theory inductively and therefore inform the next action research cycle through the continual analytical coding, classifying, categorising and eventual axial coding, where the relationship between codes/categories/concepts emerge (Urquhart, 2001). Systematic coding of raw data helps to identify, develop, and relate the concepts that are the building blocks of theory (Strauss & Corbin, 1998).

Data are broken up into ‘data bytes’ of individual meaning that can be categorised. This was done initially at a high level, to begin to understand the data, then later on a more detailed level, where sub categories were then developed. Axial coding, where data are analysed for high level themes, then allows for combining of categories and sub categories. Dey (Dey, 1993) notes the analysis of empirical data using coding techniques similar to that of grounded theory are also of use. The coding and analysis remains iterative in nature but differs in that pre-determined categories, derived from the literature and the researcher’s existing knowledge, may be used. As Dey states we want to use existing ideas, but not prejudice the data (Dey, 1993). Dey’s
Concern is that practical reality for the researcher undertaking analysis is not as rigid as grounded theory suggests.

There is a considerable body of work available on the theory of data quality. However, little research is available that pertains to a national health sector, nor around the development of a national data quality strategy. Grounded theory attempts to allow the researcher to start the research project without a preconceived overall theory in mind, letting the theory emerge from the data. Theory derived from data is more likely to resemble the participant's reality than is theory derived from putting together a series of concepts based on the researcher's experience and previously held theoretical perspectives. Because the theory comes directly from the data it is likely to offer insight, enhance understanding, and provide a meaningful guide to action (Strauss & Corbin, 1998). The initial literature search provided for a foundational theory from the body of work available. Subsequent analysis of the data developed new theory around the application of the existing theory from the literature, from a strategic perspective. The data collected are predominantly unstructured text, obtained through the methodologies outlined below.

4.2.3 Focus Groups

Focus groups can yield useful data on complex behaviour and explore attitudes towards new and unfamiliar topics, combining both interviewing and observing skills to allow for the observation of a great deal of discussion on a topic in a short time. Focus groups are also known as group depth interviews. A group is 'a number of interacting individuals having a community of interest'; depth involves 'seeking information that is more profound than is usually accessible at the level of inter-personal relationships', and interview 'implies the presence of a moderator who uses the group as a device for eliciting information'. The term focus implies that the interview is limited to a small number of issues (Stewart & Shamdasani, 1990). Focus groups generally involve at least five but no more than 15 participants who discuss a particular topic under the direction of a moderator who promotes interaction and assures that the discussion remains on the topic of interest (Stewart & Shamdasani, 1990). Too small a group does not give the researcher the participant interaction that the focus group aims to achieve and too many can lead to the exclusion of quieter members.

Two sets of data can be collected from focus groups; first, the direct comments provided by the participants and second, the interaction of the participants. The researcher used the information provided by participant interaction to obtain a clearer understanding of cross-organisational issues with data quality and the maturity of the organisation as a whole to make change.

4.2.4 Participant Observation

Observation is used as a method of data collection for both qualitative and quantitative research. Much information can be gained from mundane events, which the social actors may not be consciously aware of and, due to this lack of awareness, may not articulate this information (Porter, 1991b), (Langley, 1994). Participant observation is an essential data collection method in action research, as collaborative teams work alongside the researcher to bring about learning and change. By joining with those who were under observation, i.e. participating, the researcher was better able to understand the situation from the research participants’ perspective (Langley, 1994). The organisation where the study took place is a complex
environment with multiple and sometimes conflicting goals, with significant outside drivers impacting on work priorities and change limitations. This environment would be extremely difficult to understand and take into account if the researcher were not ‘one of the team’ as a member of staff within the organisation. A more in-depth review of the organisational environment is provided in Chapter Three.

The observation in this study was overt and unstructured. Lynes (Lynes, 1998) reinforces Pretzlik’s (Pretzlik, 1994) view that structure can inhibit creativity and flexibility in observational research, as it is logistically difficult to categorise phenomena at the time of the event. Participant observation in this study was heavily utilised, as the researcher was able to observe events and categorise them later, meaning that no interruption during the observation period was required. The researcher later recorded field notes regarding the observations made. Study of the organisation and the research participant’s knowledge of data quality was able to be observed over time, providing a study of the change process. Participant observation allows for this type of longitudinal study, whereas a one-off survey, for example, would merely provide a point-in-time snapshot (Dey, 1993).

4.3 Organisational Recognition of the Need for Change

Documentation around difficulties with data quality within the organisation provided evidence that a review of the current work practices was required. The researcher was new to the organisation and began work on two steering groups for the implementation of two proposed new national data collections. Both steering groups requested advice on how to ensure the best data quality in the new collections. This prompted the researcher to review the literature on data quality. Modern data quality theory was new to the researcher as it was to the research participants. This allowed the development and learning to begin with an open view, reviewing the literature and talking to those who manage and use data within the organisation and the health sector. Gaining agreement from management that improvement in data quality practices was necessary and supported by a review of current data quality practices (the Current State Analysis Report outlined below) and the feedback provided by staff for the development of the Ministry of Health Information Systems Strategic Plan (ISSP).

4.3.1 Current State Analysis of Data Quality at the Ministry of Health

The ‘Current State Analysis Report’ systematically records data quality measures of 33 national health data collections. The data collections are managed by HealthPAC Wellington, HealthPAC Dunedin, the National Screening Unit and/or the New Zealand Health Information Service. The methods for gathering information included interviewing staff of the collection manager groups, reviewing Ministry of Health website information, scanning organisational data quality documentation for relevant information and a review of ISSP background documents. A template was completed for each collection to provide comparative and complete data.

An analysis of the data highlighted the issues found by gathering information on the national health data collections and added to the knowledge gained through the ISSP working groups. The ISSP provided considerable information on the management of information within the Ministry of Health and had been
developed through the extensive involvement of Ministry of Health staff in 2003. The Current State Analysis looked at the data quality measures undertaken on a regular and ad hoc basis, and their perceived effectiveness on data quality. Recommendations were made, and agreed to by management, as to the next steps in the strategy and framework development process. Management at this stage also agreed to send the researcher to the Massachusetts Institute of Technology in Boston to attend an international data quality conference and a five-day foundation course (IQ1) to gain further knowledge in the field and to interact with other data quality practitioners. The researcher returned the next year to attend the conference, present the research in progress (Kerr & Norris, 2004) and complete the more advanced five-day course.

4.4 Development of the Data Quality Evaluation Framework

A search on Google by the researcher highlighted the work of the Canadian Institute for Health Information (CIHI), which has undertaken pioneer work on health care data quality. The CIHI Data Quality Framework (Long & Seko, 2002) is based on Statistics Canada guidelines and methods, information quality literature and the principles of Total Quality Management. The CIHI is similar in function to NZHIS in that they are the custodians of national health data collections. The development of the New Zealand Framework was preceded by assessing the suitability of the CIHI Framework for the New Zealand environment and the potential to modify it to ensure local applicability.

The CIHI Framework was assessed for completeness and relevance against current Ministry information technology and information management strategy documents. These include regional information strategy plans and the WAVE Report (Working to Add Value through E-information) (WAVE Advisory Board, 2001) which, at that time, was the national information management strategy for health. Compliance with New Zealand legislation was also considered.

The proposed Ministry of Health framework was then discussed at two focus groups of internal Ministry staff. Focus groups were used in an effort to bring together business units who appeared to have similar issues with data quality, but at that time no formal infrastructure to coordinate quality initiatives. The focus group participants were derived from a ‘Ministry Data Quality Team' (MDQT) formed specifically to look at ways of improving data quality consistently across the organisation. Membership of the MDQT was selected from across the Ministry and its separate business units. Members were mostly ‘information users' such as Information Analysts and Business Intelligence staff, some were also members of the already existing operational Clinical Analysis Team at NZHIS, whose function is to improve the data quality of the national health data collections. All regularly use data for different purposes.

Membership of the MDQT totalled 11 and included:

- Public Health Intelligence Analyst;
- HealthPac Business Analyst;
- Senior Advisor, Clinical Analysis;
- Advisor, Health Information Strategy and Policy (the researcher);
- Senior Advisor, Health Information Strategy and Policy;
• Team Leader, Data Quality Team;
• Team Leader, Clinical Analysis Team;
• Manager, Clinical Analysis and Data Quality Team;
• Senior Advisor, National Screening Unit (& Register);
• Two Senior Advisors, Funding and Planning Unit, Clinical Services Directorate.

The proposed framework was sent to all group participants. A presentation to the MDQT was made prior to the focus groups to ensure all participants had a common understanding of the purpose of the framework and the desired outcome goals. The group participated in two focus groups of two hours each. A member of the Ministry’s Health Information Strategy and Policy Group (the researcher) led the focus groups and an administrator was present to make audio recordings and to later transcribe the recordings, noting also the interaction between group members on discussion points. The subsequent changes made to the framework were posted on a collaborative workspace (password controlled website) to ensure the participants agreed with the researcher’s interpretation of the feedback.

Following the focus group sessions, a second review of the framework was made using meta criteria defined by Eppler and Wittig (Eppler & Wittig, 2000), to ensure that it remained robust according to the data quality literature following localised changes. The meta-criteria consist of analytic and pragmatic criteria. The analytic criteria are:

• academic standards requiring clear definitions of terms;
• position within the literature;
• consistent and systematic structure.

The pragmatic criteria are based on the applicability of the framework and are:

• conciseness;
• examples provided for illustration of meanings;
• inclusion of tools.

The criteria chosen by Eppler and Wittig (Eppler & Wittig, 2000) are based on existing literature, common sense and interviews with practitioners. Eppler and Wittig (Eppler & Wittig, 2000) note that there are currently no defined meta criteria for the evaluation of frameworks and the methodology they used in their development was also used by Huang, Lee and Wang (Huang, Lee, & Wang, 1999) in their analysis of data quality frameworks. The researcher therefore considered the meta criteria to be the most appropriately developed tool to provide assurance that the framework remained robust.

Review of the literature on data quality metrics and methodologies for their appropriate development was undertaken. Refinement of the metrics associated with each data quality characteristic was required. To this end, a generic process for defining metrics provided an outline to the Ministry and the sector on how to develop appropriate metrics related to the data being assessed. The literature and discussions with data managers highlighted the current lack of clear measurements at NZHIS. Further refinement of the metrics used at NZHIS is still required as the lack of staff to develop appropriate metrics meant this work could not be undertaken. This completed the first cycle of action research.
The framework then went through a pilot evaluation process using two health data collections. Initial assessment was made on a national health collection considered to have good data quality in relation to other collections, the Mortality Data Collection. The Mortality Collection has been established to provide data for public health research, policy formulation, development and monitoring, and cancer survival studies. A complete data set of each year’s mortality data is sent to the World Health Organisation to be used in international comparisons of mortality statistics.

The second data collection consisted of clinical data held in a local hospital setting. These data are used to determine best health outcomes for clinical care pathways and they are consequently stored at a more granular level than the national health data.

An 'issues register' was kept to ensure that any problems found with the use of the framework were captured and changes made where required. The research concentrated on eliciting the opinions of the participants on areas outlined below and specifically provided insight into answers for research questions one and two (Section 1.6.2):

- the applicability of the dimensions, characteristics and criteria for the assessed collection;
- proposal of other dimensions that may be applicable;
- the language used in the framework;
- the language and examples provided in the accompanying user manual;
- the length of time required to complete the assessment using the framework;
- the value to users of the information provided from using the framework.

The framework will be further developed through an iterative process of refinement. As NZHIS and the health sector undertake further analysis and improvement to data quality, better data quality information may highlight areas for improvement in the framework. Table 6 provides a summary of the research process undertaken to develop the Ministry of Health Data Quality Evaluation Framework and the research question addressed by each process.
<table>
<thead>
<tr>
<th>Process</th>
<th>Outcome</th>
<th>Research Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Search</td>
<td>CIHI Framework found, data quality literature found</td>
<td>1, 6</td>
</tr>
<tr>
<td>Discussions with stakeholders within Ministry of Health around the applicability of the CIHI framework, background paper provided to management team</td>
<td>Approval from management for further investigation of the CIHI framework and its applicability to New Zealand</td>
<td>2, 3</td>
</tr>
<tr>
<td>CIHI Framework assessed for relevance and completeness, applicability to the local environment</td>
<td>CIHI Framework confirmed as good basis for development of a New Zealand Framework</td>
<td>2</td>
</tr>
<tr>
<td>Presentations made to MDQT on data quality literature and the CIHI Framework</td>
<td>All focus group participants shared an understanding of current data quality literature and the structure and function of the CIHI Framework</td>
<td>1, 3</td>
</tr>
<tr>
<td>Proposed Ministry of Health Framework provided to MDQT for detailed examination in two focus groups</td>
<td>Confirmation of applicable data quality dimensions with addition of one further dimension. The development of a draft Ministry of Health Data Quality Evaluation Framework</td>
<td>2, 3</td>
</tr>
<tr>
<td>Draft Ministry of Health Data Quality Evaluation Framework provided via web site for approval from MDQT</td>
<td>Approval from MDQT that researcher correctly interpreted the feedback from the focus groups</td>
<td>2</td>
</tr>
<tr>
<td>Review of the literature on data quality metrics and methodologies for their appropriate development</td>
<td>Refinement of the metrics associated with each data quality characteristic. Further refinement of these measures is required.</td>
<td>1, 2</td>
</tr>
<tr>
<td>Pilot evaluation of the draft Ministry of Health framework on two collections</td>
<td>An issues register was provided to the researcher to ensure all required changes were noted</td>
<td>2, 3</td>
</tr>
<tr>
<td>Workshop held with Ministry of Health wide staff (outlined in section 3.6)</td>
<td>Further confirmation that the data quality dimensions included in the framework are applicable and complete</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 6: Summary of the Development Process of the Ministry of Health Data Quality Evaluation Framework

This concluded the second ‘cycle’ of action research. The results of both cycles of action research are detailed in Chapter Five. Analysis of the learning within NZHIS on data quality improvement provided strong evidence to management and staff of the need for a strategy to support the findings from the frameworks implementation.

4.5 Workshop

The overall objective of the workshop was to develop a targeted roadmap and portfolio of data quality improvement ‘initiatives’ at NZHIS as a basis for practically informing the development of the draft Data Quality Improvement Strategy. The intention was to qualify the approach thus developed on the NHI Register. The NHI has a fundamental role to play in many of the national collections by providing a unique patient identifier. Therefore initial improvement efforts concentrated on the NHI collection, with subsequent rollout to all other national health data collections. The uses of the NHI will change with the changing needs of the sector and duplicate entries on the register were beginning to impact on its usefulness. Therefore, there was a perceived need to ‘future proof’ the NHI data quality improvement initiatives.

The goals of the workshop were to:
• socialise the initial work of the Data Quality Improvement Strategy to date with internal stakeholders;
• link the existing Data Quality work and the NHI upgrade programme to the strategy;
• develop an indicative data quality work programme for the NHI;
• develop an ongoing direction for the development of the Data Quality Improvement Strategy.

All workshop participants were asked to familiarise themselves with the following documents relevant to the topic and the workshop objectives:

• Ministry of Health’s 2003 Information Services Strategic Plan (Ministry of Health, 2003c);
• Current State Analysis of Data Quality at the Ministry of Health Report (New Zealand Health Information Service, 2003);
• Draft Data Quality Improvement Strategy.

Workshop participants, who had a role to play in the management or use of data, were invited to attend; again using theoretical sampling, to ensure attendees came from areas across the entire Ministry of Health. Participants included:

• members of the NHI Upgrade Programme;
• the management team of NZHIS (excluding the Group Manager);
• Analyst Public Health Intelligence;
• HealthPac Business Analyst;
• Senior Advisor Clinical Coding;
• Senior Advisor Mental Health Information;
• Senior Advisor Clinical Analysis;
• Senior Advisor Health Information Strategy and Policy;
• Senior Advisor, Knowledge Management (Sector Policy);
• Team Leader, Data Quality Team;
• Team Leader, Clinical Analysis Team.

Participants were given a Workbook (Appendix 1) to provide them with sufficient information to come prepared to participate in the workshop, as some were not familiar with the data quality programme. Participants were also given a presentation from the NHI Upgrade Programme team providing an outline of the project. The researcher provided a summary of the progress of the ‘Data Quality Improvement Programme’, including the development of the Data Quality Evaluation Framework and the dimensions of data quality included in the Framework. A summary of an initial draft of the Data Quality Improvement Strategy was also provided. Participants were asked to identify themselves as a data custodian, a data consumer, or a data collector. Lee (Lee & Strong, 2003), notes the potentially differing knowledge requirements for the each role, and that overall the knowledge of data collectors is more critical than that of data custodians.

Minor changes were made to the draft Data Quality Improvement Strategy following the workshop. The second draft was then circulated, providing a discussion document of the draft Data Quality Evaluation
Framework and Data Quality Improvement Strategy for internal Ministry of Health staff and the Executive Team.

4.6 Survey of International Organisational Data Quality Practices

A survey was undertaken of organisations based in New Zealand and overseas. The survey elicited information on data quality practices of organisations, particularly around strategies and programmes in place. The survey questionnaire was piloted by the 12 attendees of the MIT IQ1 foundation course held in Boston in November 2003 for the applicability of responses. Feedback on the questions was sought and obtained from the attendees and the course lecturers, Richard Wang and Yang Lee. Most attendees of the course were employed full time in large organisations as data quality practitioners with several years’ experience in the field and extensive knowledge of the data quality programmes in place within their organisation.

The revised questionnaire was then used to survey attendees at a data quality course held in New Zealand in July 2004 by Richard Wang and also at a Business Intelligence conference in Sydney in November 2004. A total of 15 responses in free text was analysed through grounded theory methodology, by coding content into categories and eventually themes. Responses came from organisations in New Zealand, Australia, USA, Germany and Brazil. A review of the coding methodology and results is provided in Chapter 6.

4.7 Current State Analysis of Data Quality in the New Zealand Health Sector

The researcher undertook an extensive review of current data quality practices within the New Zealand health sector through stakeholder interviews and a survey of opinion on the data quality dimensions used in the Data Quality Evaluation Framework.

4.7.1 Semi Structured Stakeholder Interviews

The researcher was informed through face to face semi structured interviews with:

- three organisational data quality teams within health care provider organisations;
- Group Manager, NZHIS;
- Manager, Clinical Analysis Team, NZHIS;
- ten (of a possible 18) DHB Chief Information Officers;
- staff from two Shared Services Agencies;
- one PHO.

A questionnaire was developed that provided the researcher with a general guideline for semi structured interviews (Appendix 2) and was designed to elicit information on areas such as current data quality programmes or initiatives in place, the level and type of support provided from management, the roles found in data quality in health sector organisations and the constraints faced in improving data quality. Research
question four was specifically addressed in the interviews. The free text data provided by transcribing the audio taped interviews that were coded using grounded theory techniques and elicited key concepts around data quality management in the New Zealand health sector. The coding and analysis processes used and results found are detailed in Chapter Six.

4.7.2 Survey of Data Quality Dimensions and Priorities for Improvement

To provide further information for the analysis of data quality work in the New Zealand health sector, and to address research question two, a survey of data quality dimensions and their priority with health sector stakeholders was undertaken at the largest annual Health Informatics Conference in New Zealand. The aim of the survey was to elicit the opinions of the health sector as to which applicable data quality dimensions to measure and improve and which dimensions are considered to be a priority for improvement. The Data Quality Evaluation Framework was used for initial guidance in defining what data quality means to the New Zealand health sector, through the discussions around the dimensions of data quality found in the CIHI Framework, thus initiating the first step in the TDQM process. Data quality dimensions were defined, using an empirical approach that analyses the data collected from data consumers, to determine the characteristics they consider appropriate to assess data quality. Dimensions cannot be theoretically determined or intuitively selected by the researcher. Empirical research captures the views of the customers and may reveal dimensions that the researcher may not have considered. Wang & Strong (Wang et al., 1996) note however that the limitation of such empirical research is that the results cannot be proven via fundamental principles. A total of 259 people registered to attend the Conference. Conference attendees come from a wide range of backgrounds, as summarised in Table 7 below.

<table>
<thead>
<tr>
<th>Type of Organisation</th>
<th>Number of Attendees</th>
</tr>
</thead>
<tbody>
<tr>
<td>District Health Board</td>
<td>78</td>
</tr>
<tr>
<td>Ministry of Health</td>
<td>19</td>
</tr>
<tr>
<td>Health IT Vendor</td>
<td>99</td>
</tr>
<tr>
<td>University</td>
<td>19</td>
</tr>
<tr>
<td>Primary Care Provider or NGO</td>
<td>7</td>
</tr>
<tr>
<td>Private Health Care Provider</td>
<td>4</td>
</tr>
<tr>
<td>Health Consultant</td>
<td>6</td>
</tr>
<tr>
<td>Accident Compensation Corporation</td>
<td>5</td>
</tr>
<tr>
<td>Health Care Standards Organisations</td>
<td>4</td>
</tr>
<tr>
<td>Other</td>
<td>19</td>
</tr>
</tbody>
</table>

Table 7: Summary of the Number of Health Informatics New Zealand 2005 Conference Attendees by Organisation
All conference attendees were provided with a survey form (Appendix 3) on registering at the conference. Throughout the conference, attendees were reminded by the organisers and the researcher to complete the questionnaire. The researcher attended the conference and further data were provided through informal discussions with those completing the questionnaire.

The questionnaire was anonymous, only asking respondents to define themselves as a data custodian, collector or user. This allowed for analysis of the results by role definition. The respondents were asked to rank the dimensions of data quality used in the Ministry of Health Data Quality Evaluation Framework, giving their opinion of the most important dimension for measurement and improvement in the national health collections. Respondents were also asked if any further dimensions should be added to the list, with space for free text comments on the existing data quality dimensions. Seventy-two completed surveys were returned to the researcher, providing for a 36% response rate.

4.8 Analysis of Capability Maturity of New Zealand Health Care Organisations and International Organisations

The results of the Current State Analysis of Data Quality in the New Zealand Health Sector and the international organisations surveyed provided information that could be used to assess the data quality capability maturity of both types of organisation. This data was then used to compare these two types of organisation to ascertain the level of data quality maturity in the New Zealand health sector when benchmarked against other organisations. The aim of a review of data quality ‘maturity’ within the New Zealand health sector was to elicit the potential and readiness for change, providing further guidance to develop a national strategy that was both applicable and feasible to implement.

The researcher utilised a tool that was originally developed by Crosby (Crosby, 1980), the Quality Management Maturity Grid. English (English, 1999a) modified the grid for use in TDQM practices to provide a ‘gap analysis’ for data quality practitioners planning improvement programmes, detailed in Table 8. The tool provides a grid that can be used to assess the maturity of organisations in relation to:

- management understanding and attitude;
- data quality organisational status;
- data quality problem handling;
- cost of data quality as percent of revenue;
- data quality improvement actions;
- summation of organisation data quality posture.

Each section is rated from one to five:

1. Uncertainty (ad hoc)
2. Awakening (repeatable)
3. Enlightenment (defined)
4. Wisdom (managed)
5. Certainty (optimising)
<table>
<thead>
<tr>
<th>Measurement Categories</th>
<th>Stage 1: Uncertainty (Ad hoc)</th>
<th>Stage 2: Awakening (Repeatable)</th>
<th>Stage 3: Enlightenment (Defined)</th>
<th>Stage 4: Wisdom (Managed)</th>
<th>Stage 5: Certainty (Optimising)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Management understanding and attitude</td>
<td>No comprehension of data quality as a management tool. Tend to blame data admin or IS for data quality problems or vice versa.</td>
<td>Recognising that data quality management may be of value but not willing to provide money or time to make it all happen.</td>
<td>While going through data quality improvement programme learn more about quality management; becoming supportive and helpful.</td>
<td>Participating. Understand absolutes of data quality management. Recognise their personal role in continuing emphasis.</td>
<td>Consider data quality management an essential part of the company.</td>
</tr>
<tr>
<td>Data quality and organisation status</td>
<td>Data quality is hidden in application development departments. Data audits probably not part of organisation. Emphasis on correcting bad data.</td>
<td>A stronger data quality role is appointed but main emphasis is on correcting bad data.</td>
<td>Data quality organisation exists, all assessment is incorporated and manager has a role in development of applications.</td>
<td>Data quality manager reports to CIO; effective status reporting and preventative action. Involved with business areas.</td>
<td>Data quality manager is part of management team. Prevention is major focus. Data quality is a thought leader.</td>
</tr>
<tr>
<td>Data quality problem handling</td>
<td>Problems are fought as they occur; no resolution; inadequate definition; lots of yelling and accusations.</td>
<td>Teams are set up to attack major problems. Long-range solutions are not solicited.</td>
<td>Corrective action communication established. Problems are faced openly and resolved in an orderly way.</td>
<td>Problems are identified early in their development. All functions are open to suggestion and improvement.</td>
<td>Except in the most unusual cases, data quality problems are prevented.</td>
</tr>
<tr>
<td>Cost of information quality as % of revenue</td>
<td>Reported: unknown Actual: 20%</td>
<td>Reported: 5% Actual: 18%</td>
<td>Reported: 10% Actual: 15%</td>
<td>Reported: 8% Actual: 10%</td>
<td>Reported: 5% Actual: 5%</td>
</tr>
<tr>
<td>Data quality improvement actions</td>
<td>No organised activities. No understanding of such activities.</td>
<td>Trying obvious 'motivational' short-range efforts.</td>
<td>Implementation of the 14-point programme⁹ with thorough understanding and establishment of each step.</td>
<td>Continuing the 14-point programme and starting to optimise.</td>
<td>Data quality is a normal and continued activity.</td>
</tr>
<tr>
<td>Summation of company data quality posture</td>
<td>“We don’t know why we have problems with data quality”</td>
<td>“Is it absolutely necessary to always have problems with data quality?”</td>
<td>“Through management commitment and data quality improvement we are identifying and resolving our problems”.</td>
<td>“Data quality problem prevention is a routine part of our operation”.</td>
<td>“We know why we do not have problems with data quality”.</td>
</tr>
</tbody>
</table>

Table 8: Data Quality Management Maturity Grid (English, 1999a)

---

⁹ Deming’s 14 point quality improvement programme, as applied to data quality by English (English, 1999a)
4.9 Data Quality Improvement Strategy Development

Initial semi-structured interviews were conducted to obtain background information relating to the data quality improvement strategy with the Business Intelligence Manager NZHIS, the Chief Advisor, Health Information Strategy & Policy NZHIS, Acting Manager, Health Information Strategy & Policy NZHIS, and the National Health Index (NHI, see section 3.3.1) Upgrade\textsuperscript{10} programme project manager. The NHI is a database that assigns a unique identifier to health care users including demographic details (Ministry of Health, 2005). The researcher then developed a draft strategy, essentially to provide a ‘strawman document’ for a workshop to be held with Ministry of Health staff. The development of the strategy built on the data quality dimensions included in the framework. Strategy development was supported by the following research:

- extensive review of the current literature;
- emails, correspondence and discussions with stakeholders and data quality experts both locally and internationally;
- review of other national health data quality strategies and discussions with those involved about lessons learned;
- Ministry of Health internal workshop (outlined below) on improving data quality on the National Health Index unique identifier, used to further confirm data quality dimensions, framework and initial strategy development;
- consideration of current policies and strategies that impact on data management and data quality, including the Ministry and DHB Information Systems strategic plans, the Health Information Strategy for New Zealand, the Data Stewardship Policy;
- New Zealand health sector survey from Chief Information Officers (Ministry of Health, 2003a) of current data quality maturity, current IM practices and proposed strategic direction;
- the development of the ethnicity data protocols – a template for the development of data collection and management process standardisation for the major data elements of the national collections, endorsed as a standard by the Health Information Standards Organisation;
- consideration of the Ministry of Health’s readiness or maturity for change;
- a review of international data quality practices in organisations (outlined in section 6.1);
- a review of current practice, a ‘current state analysis’ of data quality practices in the New Zealand health care sector through semi structured interviews and questionnaires with DHB Chief Information Officers, organisation wide data quality teams, Shared Service Agencies, and Primary Health Organisations (discussed below).

As is standard practice at the Ministry of Health, initial strategy development was undertaken within the organisation, with internal management approval required before discussion with external agencies or patient representative groups. The data collected and analysed for the development of the Data Quality Improvement Strategy provide insight into the answers to research questions two to seven.

\textsuperscript{10} The NHI Upgrade was a programme of work to improve the quality of data held on the NHI registry, and in particular to reduce the number of duplicate registrations.
5 Research Method

5.1 The Action Research Cycle

Each change to the DQEF utilised the action research cycle of action planning, action taking, evaluating, specifying learning and diagnosing (Susman & Evered, 1978). The diagnosis phase of the initial action research cycle involved the Current State Analysis of Data Quality within the Ministry, providing the organisation’s reason for wanting to change. The initial action plan was to develop an effective but generic evaluation tool for all national health collections and it was decided by the research participants that the CIHI Framework would provide a good basis for development and learning in the Ministry of Health. A comparison of the CIHI environment with NZHIS showed that, while there are differences that need to be considered, the functions and aims of both organisations were very similar. The action-taking phase involved two focus groups with the newly formed Ministry wide data quality team (MDQT), with subsequent evaluation of the changes suggested. Specified learning from this process provided the researcher and research participants with insight into the applicability of the DQEF and its usefulness within NZHIS and across the Ministry to support a structured data quality programme. The action research cycle was found by the researcher to be similar to that espoused by Argyris and Schön (Argyris & Schön, 1978), where double loop learning takes place; once an issue is found and corrected, the processes used by the organisation are changed in response.

Figure 15 below illustrates the initial action research cycle, adapted from Susman and Evered (Susman & Evered, 1978) and applied to this research.

![Figure 15: Initial Phase of the Action Research Cycle for the Data Quality Evaluation Framework Adapted from Susman and Evered (Susman & Evered, 1978)](image)

Utilising grounded theory techniques to analyse data from the initial action research cycle, with subsequent referral to the applicable literature, ensured the development work remained imbedded in the data provided by focus group participants. The responses provided by the focus groups for each criterion underwent detailed content analysis to elicit the overall agreement of the group as to what changes should be made. The researcher was required to reflect on the outcomes of the initial action research cycle, Argyris and
Schön (Argyris & Schön, 1978) describe this as reflection in action and reflection on action. The researcher needed to build new understandings to inform the new actions of the next action research cycle.

The next action research cycle began with a diagnosis of what was required to improve the DQEF and implement it as a tool used for regular data quality evaluations. The second cycle of action research (Figure 16) involved the development of supporting documents, such as templates and policies, to implement the DQEF. The specific steps in this second action research cycle included:

- the development of a generic template as a guide for completing data quality evaluations;
- piloting of the DQEF on a national health data collection;
- piloting of the DQEF on a clinical database held at a DHB (to ascertain applicability and usability in the wider health sector).

Figure 16: Second Action Research Cycle Concentrating on Implementation of the Framework
Adapted from Susman and Evered (Susman & Evered, 1978).

5.2 Chronology of Research Progression

Table 9 below outlines in chronological progression the data collection processes, numbers of participants and from which type of organisation the participants come from.
Table 9: Research data collection methods, number of participants and where the participants were sourced.

5.3 Grounded Theory Theoretical Sampling

As noted in section 4.1, theoretical sampling was used throughout the research, informed by the use of Grounded Theory methodologies. Where it became clear that particular stakeholders would be likely to be able to provide the appropriate input into the research, interviews and participatory observation targeted participants in particular roles. This is outlined below in table 10.

Table 10: Participant Samples used Throughout the Research Process

<table>
<thead>
<tr>
<th>Research Stage</th>
<th>Participant Sample</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial feasibility</td>
<td>NZHIS Clinical Analysis team staff</td>
<td>Knowledge of NZHIS current data quality issues and solutions</td>
</tr>
</tbody>
</table>
| Development of the DQEF | Information analysts as data users NZHIS staff who work in data NHI Upgrade programme staff | Understanding of data from the user perspective  
Understanding of data quality issues and solutions |
| Development of the DQIS | NZHIS Management team                                                              | Management view of data quality issues and feasible solutions, methods for developing and implementing a national strategy |
|                         | Staff of international organisations outside of health care                         | To understand the data quality issues that face organisations generally and their methods for improvement |
|                         | Health Informatics Conference attendees                                            | The views of the wider health sector on data quality dimensions and their priorities for improvement |
|                         | DBH & PHO CIOs                                                                      | Insight into data quality issues and current practice in the health sector                  |
5.4 Chapter Conclusion

This chapter provides an overview of the methodologies used in this research. The research is interpretive, studying meaning within a social setting. The research then also provides the structure for learning and potential change through the utilisation of action research. Grounded theory provides the structure for the analysis of qualitative data through inductive coding and constant comparison in the analysis phase of the action research iterative cycle. Participatory observation provided considerable rich data, as the researcher was a member of staff within the organisation. Data were also collected at workshops, focus groups, structured meetings and interviews.

The research took place over a two-year period at the Ministry of Health, which undertook to learn about data quality, its current theories and practices, how they might apply to their situation and how to develop these theories into a structured data quality strategy. Although the Ministry of Health ISSP was not a part of this research, its development highlighted to Ministry of Health data users and collectors that data quality improvements were required. Through an analysis of current practice and data quality limitations, Ministry of Health management recognised there was a need to improve data quality. Subsequent learning by the organisation provided the impetus to develop a national strategy to improve data quality across the whole of the health sector.

Research was undertaken throughout the health sector to provide the researcher with a clear understanding of the maturity of the health sector to make change in data quality practices. A comparison of New Zealand health sector data quality practices with those organisations outside of health care and abroad provided a benchmark of data quality maturity. This enabled the researcher to better understand the requirements of the sector for a data quality framework and improvement strategy that would be feasible to implement.

In line with the findings of other studies, the researcher found that action research allowed the researcher to be an active participant in the research process. The researcher and participants developed and implemented a data quality evaluation framework and a draft data quality strategy. The framework and strategy were used as initial discussion documents throughout the Ministry and the health sector, providing guidance at workshops and organisational data quality team meetings. The results of this research are outlined in the subsequent sections. The Ministry of Health Executive Team approved a final ‘Strategy Consultation Document’ for circulation and discussion with health sector stakeholders.
Part C Results and Discussion
6 The Development of a Generic Data Quality Evaluation Framework

This Chapter and the following Chapter describe the results of research on the development of a data quality improvement programme. The action research process enabled the researcher and study participants to work towards the development of an empirical and practical data quality evaluation framework and subsequently, an all-of-sector data quality improvement strategy through learning and reflexive practice. A detailed analysis of the results is provided in Chapter Seven. This Chapter specifically answers the research questions:

- What existing data quality theories assist in the decision making around defining and measuring data quality requirements in national health data collections?
- What steps would be required to initiate a structured data quality improvement programme within the New Zealand Health Information Service?

6.1 Comparison of Canada and New Zealand

The Canadian Institute for Health Information (CIHI) is a single national organisation funded by both federal and provincial Governments. Canada’s strategic advantage in this national health data collection management comes perhaps from the fact that there is a single organisation (CIHI) that is responsible for coordinating the development and maintenance of a comprehensive and integrated approach to health information (Mitchell & Brown, 2002). The CIHI data quality framework was developed through a national consultation process held in 1998 to identify Canada's health information requirements. The resultant roadmap identified a national vision and led to funding for a large number of projects relating to the improvement of health information systems, one of which is the Data Quality Enhancement Project.

The CIHI data quality framework was introduced to provide a common, objective approach to assessing data quality. At an operational level, it also standardises information on data quality for users and is focused on a process of continuous improvement (Canadian Institute for Health Information, 2003a; Mitchell & Brown, 2002). This broad and systemic approach to data quality and its integration across the entire organisation aligns well with the principles of good data quality discussed in the literature (Agmon & Ahituv, 1987; Wang et al., 1996).

The CIHI framework is structured along five general dimensions of quality, defined by Wang & Strong (Wang et al., 1996) and validated in a health informatics setting by Gendron and D’Onofrio (Gendron & D’Onofrio, 2001), and operationalises data quality as a four-level conceptual model (Long & Seko, 2002). At the foundation of the model are 86 basic unit items known as criteria. These criteria can be aggregated using the framework algorithm into the second level of 24 characteristics (e.g. under-coverage, reliability, and interpretability) that, in turn, can be aggregated using the algorithm into the five dimensions of data quality (i.e. accuracy, timeliness, comparability, usability, and relevance). Finally, the five dimensions can be reduced using
the algorithm into one overall database evaluation. Figure 8 below provides a summary of the four-level conceptual model.

<table>
<thead>
<tr>
<th>Overall Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>↑</td>
</tr>
<tr>
<td>5 Data Quality Dimensions</td>
</tr>
<tr>
<td>↑</td>
</tr>
<tr>
<td>24 Data Quality Characteristics</td>
</tr>
<tr>
<td>↑</td>
</tr>
<tr>
<td>86 Data Quality Criteria</td>
</tr>
</tbody>
</table>

Data Quality Analysis and Database Document Compilation

*0. not applicable, 1. unknown, 2. not acceptable, 3. marginal, and 4. appropriate

**0. not applicable, 1. unknown, 2. not met, and 3. met

Figure 8: The CIHI Data Quality Framework (Version 1) Four-Level Model (Long & Seko, 2002)

The inclusion of three levels of hierarchy within the framework is not necessary; the characteristics provided are sometimes not found in other frameworks. In the CIHI framework the characteristics are used to provide structure to logically group the underlying criteria. This logical grouping provides an aggregated score for each criterion and can be useful for mid level analysis of evaluation results when detailed information by each criterion is not required.

The CIHI framework offers a model that demonstrates systemic adoption of a data quality framework. The model recognises that the issue is not simply the quality of the data itself but also the processes surrounding data capture, collation and use, and the education of those involved. The CIHI recognises that improvements are made incrementally and implements a continuous improvement process of assessment, refinement and re-assessment. The model recognises that ultimately people are responsible for data quality and education is as important as system and data improvements. Assignments of data quality across the dimensions are made in conjunction with the staff responsible for each database or registry. It is also clearly articulated that any final determination of the level of quality is to be done once issues of confidentiality, privacy and security have been addressed. It is however noted that the data quality framework itself does not say how this should occur.

New Zealand presents a number of subtle differences from the Canadian health environment, which must be taken into account when developing a data quality framework. While the New Zealand national health data collections are centrally located, due to historical changes to the structure of the New Zealand health service, some systems are managed outside of NZHIS. HealthPAC collections are now managed within the Ministry of Health, as are the Screening Unit collections. This allows for a consistent approach to the management of national data. Further differences are apparent, in particular that any data quality framework must be
cognisant of the importance of data relating to ethnicity for the Government to meet its obligations under the Treaty of Waitangi. Recognition of the importance of data relating to ethnicity, primarily in regard to the Treaty of Waitangi obligations of the Crown, issues relating to ownership of data (whether personal or not) and Maori cultural and spiritual values around health and data are vital. Secondly, there is work that is currently being undertaken to develop all of government metadata standards. Thirdly, privacy is a central and guiding pillar for the use of health information in New Zealand and any data quality framework must consider the significance of this issue. Therefore, the assessment of data quality as being ‘fit for use’ in a New Zealand health context included the assessment of the ability of that data to support the privacy of individual information within the national collections. It is apparent that, while the NHI is flawed, it is an improvement on the diverse regional systems in place in Canada and improvements in the use of NHI present New Zealand with a significant opportunity for establishing a national baseline for data quality.

This initial review found that the CIHI Framework could assist the Ministry to develop a programme for data quality that was more ‘holistic’ and strategic in the way the organisation viewed data quality management. Such a model is easily transferable to New Zealand. However, it could not be transferred wholesale due to the differences in the health service landscape of the two countries. In many ways, New Zealand appears better placed than Canada was prior to its development of the data quality framework.

Having gained initial agreement from management and data custodians that the CIHI framework appeared applicable to local needs, the CIHI framework was assessed for completeness, applicability, and ease of adaptation against current Ministry information strategy documents. These include DHB Information Systems Strategic Plans, the WAVE Report (Working to Add Value through E-information) (WAVE Advisory Board, 2001), and the Health Information Strategy for New Zealand (Ministry of Health, 2005). Compliance with New Zealand legislation, such as the Privacy Act 1993, was also considered. This analysis of the similarities and differences between the CIHI and NZHIS highlighted areas where the CIHI framework was deficient for NZHIS requirements.

6.2 Inclusion of Ministry of Health Information Systems Strategic Plan Recommendations into the DQEF

The New Zealand health sector does not have nationally applied standards that relate to the collection, coding or storage of data, beyond the clinical coding of hospital discharges. The impact of this is described in the Ministry of Health Information Systems Strategic Plan (ISSP) and summarised in Table 11 below.

---

11 http://www.knowledge-basket.co.nz/privacy/legislation/1993028/toc.html
<table>
<thead>
<tr>
<th>Level</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Policy Level</td>
<td>One-off surveys must be commissioned as data are unreliable&lt;br&gt;Policy is implemented without basic data analysis to support it&lt;br&gt;Data collections do not exist to measure effectiveness of policy</td>
</tr>
<tr>
<td>Management Level</td>
<td>Data cannot be compared across collections;&lt;br&gt;Collections cannot be linked;&lt;br&gt;Data management models do not address inter-agency sharing of data;&lt;br&gt;Compliance costs are high;&lt;br&gt;System changes can have major, complex and unforeseen implications through the value chain.</td>
</tr>
<tr>
<td>Operational Level</td>
<td>Data are frequently found to be inaccurate;&lt;br&gt;Data are incomplete;&lt;br&gt;Data are duplicated;&lt;br&gt;Critical information is minimal;&lt;br&gt;There are no clinical coding standards (outside of inpatient data ICD10);&lt;br&gt;Not all activity is captured (such as outpatient care);&lt;br&gt;Recording of common data elements is not standardised.</td>
</tr>
</tbody>
</table>

Table 11: Summary of the Findings of the Ministry of Health Information Systems Strategic Plan (Ministry of Health, 2003c)

The ISSP highlighted the opinions of Ministry of Health data users very well:

- the vast majority of Ministry staff are largely unaware of what data are available to them. They do not know where to get data from or who to ask. The risk of duplicating data collection is high. *The Ministry is data rich but information poor*;

- data analyst skills are a scarce resource in the Ministry. As a result, as highlighted in a session with the senior management of the DHB Funding & Performance Directorate: *Information has been rubbery for so long and it is only as good as your own research capability*;

- there is no training or support provided by the Ministry for the tools used for data analysis (Business Objects), and little support for obtaining a tools license or implementing it on the desktop;

- the integrity of the data is low, with a significant percentage of data received missing key data elements or containing data that cannot be verified;

- there is little or no metadata to describe the content and known limitations of data collections, what information exists is not easily available;

- it is difficult and slow to provide new data sets from already captured data for analysis or reporting.

On a more positive note, the ISSP did plan for a future state where:

*Data quality programmes will be a strategic initiative, with measures being reported as a KPI to the Director General. Data quality issues will be well understood and communicated, such that limitations can be factored into data analyses*.
The review of the ISSP elicited criteria to be added to the draft DQEF, where the CIHI framework did not currently measure issues raised by Ministry staff. These criteria are provided in Table 12 below.

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Characteristics</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>Capture and Collection</td>
<td>Data capture is at source i.e. health encounter</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Data are stored at unit level</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Data quality feedback is given to data suppliers</td>
</tr>
<tr>
<td></td>
<td>Edit and Imputation</td>
<td>Validation checks are done at source</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Data are collected from the health encounter, not for claims/funding purposes</td>
</tr>
<tr>
<td>Usability</td>
<td>Accessibility</td>
<td>Data are available through HADIE (Health and Disability Information Exchange)</td>
</tr>
</tbody>
</table>

Table 12: Criteria Added to Draft Framework Following Review of Ministry of Health ISSP

### 6.3 Privacy and Security Dimension

In the context of the New Zealand health care environment it proved necessary to add two further data quality dimensions – privacy and security. Privacy issues must be specifically addressed in a localised context. The WAVE Project (WAVE Advisory Board, 2001) highlights that New Zealand’s privacy legislation is world leading, as is the application of privacy principles via the Health Information Privacy Code. NZHIS has created guidelines describing the procedure that must be adhered to following a request for information and HealthPAC has also implemented these. Paramount in this procedure is the application of the Official Information Act (1982), which identifies that, where information contains personal details, an objective assessment of the information must be performed to ensure that the protection of personal information is not outweighed by public interest, or if the information must be published under statute, such as through the application of the Health Act (1956). NZHIS also notes that the Health Information Privacy Code (1994) places a limitation on information exchanged between agencies such that rules of privacy governing the original data are transferred with that data.

There is considerable tension within the New Zealand health care environment between the ability to mine data for potentially highly useful information and the expectations of health care users that their information will only be used for the purpose for which it was collected. This was noted to the researcher, particularly by management, whose main concerns were around the political implications of such practices. However, as noted above, legislation does allow for the use of health information where the protection of personal information is outweighed by public interest.

The CIHI states that privacy and security are implicit requirements that are embedded in all their data management processes. Whilst the same could be said of the Ministry of Health, the pervading culture in New Zealand requires that privacy and security of information, and in particular health information, should be made clearer. Therefore, the researcher felt there was a requirement for explicit and transparent consideration of these quality dimensions and added this dimension for discussion with research participants. The characteristics of the Privacy and Security Dimension were developed by the Ministry Senior Advisors for health sector information privacy and security to ensure alignment with the ongoing
development of new privacy and security policies. This dimension specifically addresses legislation, standards, policies and processes.

Security has two components. First is the issue of authorised access to data (and systems using those data) using technology such as passwords, encryption and transaction authentication. Second, disaster recovery planning protects data from potential natural disasters with appropriate backup and alternate processing procedures in place (Levitin, 1998; Miller, 1996). By way of an example of the way security is typically described in the health context, the NZHIS guide to the NHI notes the following:

- access to the NHI is restricted to authorised users, and is permitted by the Health Information Privacy Code (1994) released under the Privacy Act (1993);
- the NHI number is stored in encrypted form on other NZHIS systems.

No reference to user or societal perceptions of the security of data could be found on either the Ministry of Health or the NZHIS websites. It is noted, however, that perceptions of security and trust are critical to the acceptance of uptake of information systems, particularly where systems interact with individuals (Suh & Han, 2003), but that little effort is devoted to security in many organisations (Levitin, 1998). The additional criteria are detailed in Table 13 below.

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Characteristics</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Privacy and Security</td>
<td>Legislation</td>
<td>Data privacy is maintained and complies with regulations</td>
</tr>
<tr>
<td></td>
<td>Disclosure</td>
<td>People accessing a collection are identified and authenticated</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Data are secure from unauthorised access</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Data used for ad hoc and statistical analysis are non identifiable</td>
</tr>
</tbody>
</table>

Table 13: Criteria Added to Draft Framework Following Review of Ministry of Health Privacy and Security Requirements

6.4 Current State Analysis of Data Quality in the Ministry of Health

A ‘Current State Analysis’ (New Zealand Health Information Service, 2003) was undertaken through a survey of data managers and data users from across the Ministry. The survey consisted of open questions requiring free-text answers to elicit information on a set of factors, including historical and contextual information, about each national health data collection, the data collection processes, any changes made to data from within the Ministry, what the data are used for, where they reside, and the nature and perceived effectiveness of existing data quality initiatives. An analysis of the findings of the survey is provided below.

6.4.1 Inconsistent information

The gathering of information on current data quality work proved difficult and this can be the case in many organisations (Levitin, 1998). The information is spread over a range of business units, people and documents so that the Ministry cannot easily assess the scope or effectiveness of its data quality measures. For some data collections, it is possible to locate a person who is knowledgeable about that collection but
there are still some data collections that have not listed a contact person who is clearly associated with or accountable for the collection. Documentation around collections is not consistent or extensive. In some cases, there is inconsistent information on specific collections. Respondents had different understandings, which were contradictory, about data collections. Some web or paper sources were considered to be out of date.

Many of the national collections are known by a short name (e.g. NHI for National Health Index) and some data collections are known by more than one name, leading to confusion as to whether two or more data collection names might be referring to one data collection or not. For example, the Health Practitioner Index and the National Provider Index collections to some respondents were the same collection and to others were two separate collections.

All collections at NZHIS have a data dictionary. Each data dictionary has a consistent layout, but data definitions between collections can vary. Metadata are available for collections, but this also varies and is sometimes difficult to locate. HealthPAC currently has no data dictionaries; the National Screening Unit has dictionaries for both its collections.

6.4.2 Overview of Methods of Data Quality Management

The data quality initiatives found at the Ministry consists of both manual and electronic measures. Manual measures include checks both prior to the data arriving (from the source) and at the time of arrival (at the point of manual data input). Some data quality measures include generating reports that can be checked against original data or forms. This manual approach is discretionary, meaning that people will use ‘common sense’ to decide whether the data are incorrect, and take this into consideration when using the data. This is an accepted ‘metric’ in data quality, where the user of the data makes a subjective decision about the fitness of use of the data for the user’s current purpose (Pipino et al., 2002). The Ministry has a blanket policy of not changing raw data. Where poor quality is suspected, data is sent back to the source for verification. In the case of electronic measures, the data quality from the source is not always known and some respondents noted the vulnerability of national data collections where the original data quality from the source may be poor. Common data quality measurement tools include:

- manual checks from records or reports (reconciliation);
- validation rules/threshold requirements/mandatory;
- business rules/process rules;
- audit trails (who did what, when);
- data matching between systems;
- data quality audits.

A study by Lorence and Jameson (Lorence & Jameson, 2001) on data quality management practices in US health care organisations noted internal audit as the most preferred data quality assessment and problem identification method, followed by examination of errors on individual records and analysis of aggregate data. NZHIS utilises similar methods.
Respondents noted that data quality is easily achievable where data quality can be checked using business rules. However, some respondents pointed out that data quality in terms of ‘content’ could be an issue. The usefulness, relevance and completeness of some data are quality issues where those data are used to make decisions. For example, data quality measures do not necessarily address problems of specificity or content in diagnosis description fields that are of particular interest to researchers and clinicians. Lorenzoni et al (Lorenzoni et al., 1999) found specificity to be an important and common issue in the quality of medical data.

It is considered by some respondents that the electronic transfer of data is likely to minimise the loss of data integrity compared to data that are loaded manually, and this is noted in the literature (Ballou & Tayi, 1989). While there is less likelihood of further inaccuracies when data are electronically transferred, respondents pointed out that data would still only be as good as its source. Compounding this is that data transferred electronically may not be as easy to check. That is, data may be electronically transferred without human intervention and therefore it cannot be located easily or the quantity of data being transferred may mean it would be impossible to check.

Data quality initiatives amongst some data collections are not consistently or regularly applied. Respondents noted that 16 of a possible 33 data collections have either no regular ‘business as usual’ data quality measures in place or they are unknown. It is not certain how accurate this is, as it is possible that with the scattered nature of information and inconsistent information, some respondents do not know all data quality initiatives used.

### 6.4.3 Perceived Effectiveness of Data Quality Initiatives

The *perception* of data quality effectiveness was inconsistent amongst respondents and it is likely that this related to the level of toleration for quality (or lack of quality), which corresponds to the purpose and perceived importance of the data collection. For example, data collections for the purpose of paying contracts or claims held at HealthPAC were seen as self-managing, as the responsibility is on the claimant or contractor to supply the correct the data or not get paid. Business rules further enable auditing of data before entering the data collection. Where data quality might be poor for collections that have no perceived use beyond the final ‘resting’ point, data quality is more likely to be perceived as effective as it has no negative impact further on.

### 6.4.4 Outcomes of the Analysis

Discussion with respondents found that these points are consistent with the views stated in the Ministry ISSP. The difficulties in finding responsible and knowledgeable people and consistent knowledge about collections are reflected in the results.

The researcher, through the Current State Analysis Report, provided the Ministry with recommendations for the steps required to develop a structured approach to data quality:
1. Develop a Data Quality Framework that encompasses the following tasks:
   a. Identify and delegate a responsible person(s) for each data collection. This person is named as first point of contact for data quality queries and is responsible for the implementation of the framework on the collection. This will be informed by the current work being undertaken throughout the Ministry to clarify and define governance around the national collections.
   b. Development of the assessment tool (the DQEF) through focus groups with information and pilot studies on each data collection.
   c. Develop a data quality documentation template that describes the:
      - purpose of the collection and the uses of the data;
      - roles and responsibilities around the collection;
      - results of a point in time assessment using the data quality framework;
      - location of all metadata;
      - historical information i.e. changes made to fields in the past, changes in the purpose of the original collection.
   d. Refine future use of the assessment tool to improve the value of information regarding data quality effectiveness (with measurable benchmarks).

2. Develop a future data quality action plan that identifies data collections and or specific fields for data quality improvement initiatives based on their priority and considers the critical nature of a data collection so that priorities can be identified.

3. Develop a data quality strategy for the health sector, informed by the development of the ISSP and the results of the framework. It is proposed that the strategy should be based on Total Data Quality Management theory.

The Current State Analysis showed that, due to the disparate nature of data suppliers, no sustainable improvements could be made unless a programme of work was placed in the context of a national strategy. At the national level a team, with the requisite data analysis and data improvement skills to solve problems, was required to provide guidance to data suppliers on what processes need improvement. Short-term priorities needed to focus on areas where benefits could be realised easily with long-term projects concentrating on implementing change.

Supplying the Current State Analysis report to management provided the researcher with the required supporting evidence to continue the project. The exercise of discussing data quality issues with a wide range of Ministry staff raised the profile of the project and allowed for participants to provide initial thoughts around the topic of data quality to the researcher, and for the researcher to provide education on the proposed DQEF. The recent publication of the ISSP included a section on data quality and provided further supporting evidence of a need for a comprehensive programme of improvement. By undertaking the current state analysis soon after the ISSP publication release, research participants noted they felt ‘heartened’ that work from the ISSP was being utilised and further developed.
6.5 Focus Group Results

The two focus groups held with the Ministry Data Quality Team (MDQT) enabled a word by word assessment of the draft DQEF that now included the dimension privacy and security and additional criteria from the review of the Current State Analysis and the Ministry’s ISSP. The changes made by the group began to develop a framework that contained language that was more applicable to New Zealand, removed several criteria that were not applicable, and confirmed or removed the criteria proposed by the researcher following the ISSP and Current State Analysis review. Initial discussions on the development of applicable metrics for the New Zealand environment were begun.

6.5.1 The New Zealand Health Sector Data Quality Dimensions

The six data quality dimensions chosen for the DQEF following analysis of the feedback from the focus groups are as follows, with accuracy being the most extensive:

- **Accuracy** is defined within the framework as how well data reflect the reality they are supposed to represent;

- **Relevancy** reflects the degree to which a database meets the current and potential future needs of users;

- **Timeliness** refers primarily to how current or up-to-date the data are at the time of use;

- **Comparability** is defined as the extent to which databases are consistent over time and use standard conventions (such as data elements or reporting periods), making them similar to other databases;

- **Usability** reflects the ease with which data may be understood and accessed. If data are difficult to use, they can be rendered worthless no matter how accurate, timely, comparable, or relevant they may be;

- **Security and privacy** reflects the degree to which a database meets the current legislation, standards, policies and processes.

No one definition can be found for even the most commonly used data quality dimensions (Wand & Wang, 1996), (Levitin, 1998). For this reason it is important for the development of the DQEF to explicitly define each dimension. Within the research participant group the understanding of the meaning of the dimensions varied. It was decided then, to utilise the definitions provided by the CIHI framework as this aligned with the characteristics and criteria found throughout the framework. Importantly, the dimensions were found to be mutually exclusive and collectively exhaustive. There was agreement from the focus group participants that the structure of the DQEF was logical and practical.
An example of the process the researcher undertook to analyse transcribed audio tapes from the focus groups for content themes is provided in Figure 8. The extract is from a discussion around the addition of the criteria relating to ‘Capture and Collection’ – Data capture is at source i.e. health encounter. The text in square brackets [] is the researcher’s added content theme analysis and identifies the person speaking.
[Researcher] The next question that I put in is mainly when I went through the ISSP [source ISSP] and picked out if we had missed anything that they thought was important and I put this one in, data capture is at source, i.e., at the heath encounter. So the person who actually is talking to patients is collecting information that we will actually end up with, and I wonder is it feasible expectation to put this in, and how would we measure it?

[participant 1] If you look at hospital situation you have a lot of people involved who actually put a filter on, you've got all sorts of people ... doctors writing notes ... but it is, I mean can we can say that it is at a health encounter, but it is not as far as our collections are concerned a unit level record of that particular health encounter. [health encounter too specific?]

[participant 2] Even if we say, no, that its not collected at source or we don’t know, that it just gives us that information. You know, so it seems quite valuable information when we are using it ... to sort of say ok ... it's been through a whole lot of other hands before it’s actually been recorded in some way into the collection. So just keep that in mind, when you’re assessing its quality overall. [positive use of data quality information provided by criterion]

[participant 3] I have an example. It’s coming from your work on ethnicity recording which shows ethnicity for example I think if it’s missing, will actually get ethnicity from NHI so if we have 90% missing in the initial record it means that whatever ethnicity we’re measuring ... we’re actually measuring NHI ethnicity recording ... which is why that would be good to know ... yeah ... [positive example of usefulness of criterion]

[participant 4] Or we can ... probably now. We can. That is why we’re working on it this. This is why I think it’s a good one. What are we actually measuring? Or what we think we are measuring. [positive, already seeing the need to measure this criterion] Right ... yes that’s true [positive support]

[participant 5] It may not exclusively just be to measure a collection and when it comes to designing the collections these features it’s going to, by default they become a guide for directorates when they’re designing the collection these features should be there by default and you have your expectations that the data should be collected at source, that should be an expectation. [positive outcome of measuring criterion]

[participant 3] Like ethnicity, we actually should be one that probably should be validated at source rather than from the NHI. So there are some uses for this. [positive example]

[Researcher] So, what would be the metric would we say, all data is collected at source ... or is that ridiculous?

[participant 3] Perhaps if we say all sources of data are documented, so that we know where it has come from. [reasoning on wording of criterion]

[participant 6] One of the issues that we are trying to resolve is getting the data more accurately at the source and one of the things that the slide you put up was around the, it’s not so much the data but it is the quality of the processes that we are trying to do. Possibly here we can make some mention of the quality processes? The metric is the extent of the process exists to collect unit data from the most appropriate source, to ensure data collected from the most appropriate source. [potential of measuring criterion]
From this discussion, the researcher found an overall positive response to the inclusion of the criteria, altering the words to include ‘appropriate source’ rather than ‘health encounter’ as originally proposed by the researcher. The discussion of an appropriate metric for this criterion noted the need for data quality information on the process of data collection (Levitin & Redman, 1993). This discussion was initiated by the focus group participants and reflects the education provided by the researcher on data quality theory prior to the focus groups. It was evident that the discussion held between participants at each focus group was enhanced considerably by two half-hour presentations provided by the researcher. Participants applied their new knowledge directly into the discussion, ensuring the framework remained aligned to data quality theory. The resulting metric stated: ‘Processes exist and are documented to ensure data are collected from the most appropriate source’. This process of discussion and agreement was followed for the review of each criterion. The changes made to the draft framework by the MDQT focus group, including the addition and removal of criteria from the original CIHI framework with the rationale for change, are found in Tables 14 – 29 below. Verbatim quotes from focus groups’ discussions provide an example of the decision process.
### 6.5.2 Accuracy

<table>
<thead>
<tr>
<th>Coverage Criteria</th>
<th>Changes</th>
<th>Rationale</th>
<th>Resultant Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>The rate of under-coverage/over-coverage falls into one of the predefined categories</td>
<td>Divided into two questions</td>
<td>When you use that data you need to know that. As long as we consider both it doesn’t really matter it’s really two separate questions.</td>
<td>The rate of under-coverage falls into one of the predetermined categories. The rate of over-coverage falls into one of the predetermined categories.</td>
</tr>
</tbody>
</table>

Table 14: Modification of the CIHI Coverage Criteria

<table>
<thead>
<tr>
<th>Capture and Collection Criteria</th>
<th>Changes</th>
<th>Rationale</th>
<th>Resultant Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data capture quality control measures exist</td>
<td>Receiver added by focus group</td>
<td>the criteria is fine …doesn’t say where … point of collection … So add to that “and by the receivers” … that would be a big one.</td>
<td>Data capture and receiver quality control measures exist</td>
</tr>
<tr>
<td>Data capture is at source i.e. health encounter</td>
<td>Changed to most appropriate source</td>
<td>Possibly here we can make some mention of the quality processes, the extent of the process exists to meet unit data at the most appropriate source … from the most appropriate source</td>
<td>Data capture is at the most appropriate source i.e. health encounter</td>
</tr>
<tr>
<td>Data are stored at unit level</td>
<td>Changed to available at unit level</td>
<td>We do tend to collect an awful lot more than we need. The manager of the data warehouse can easily get the information if somebody else is after it</td>
<td>Data are available at unit level</td>
</tr>
<tr>
<td>Data quality feedback is given to data suppliers</td>
<td>Changed to comparative feedback</td>
<td>I think the metric should be to also include that it is individually identifiable so they can see that performance relative to their peers, comparative performance</td>
<td>Comparative data quality feedback is given to data suppliers</td>
</tr>
<tr>
<td>Criteria added by focus group</td>
<td></td>
<td>Not all the systems have the user group to give feedback to all the sources or providers of the data. As part of the data governance and a bigger issue really is roles and responsibilities, and there should be someone who is the database manager with who, it is his responsibility for the data collection’s quality and that could be part of the business owner.</td>
<td>Data quality feedback issues, as noted by users, are feedback to database quality managers</td>
</tr>
</tbody>
</table>

Table 15: Modification of the CIHI Capture and Collection Criteria
<table>
<thead>
<tr>
<th>Unit Non-Response Criteria (Changed to Missing Data)</th>
<th>Changes</th>
<th>Rationale</th>
<th>Resultant Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criteria added by focus group</td>
<td>So we could say as a metric, data are collected at a level specificity that is required for the use of …certain purpose - the information reflects certain level of need.</td>
<td>Data are collected at level that provides sufficient detail for its use</td>
<td></td>
</tr>
</tbody>
</table>

Table 16: Modification of the CIHI Unit Non-Response Criteria

<table>
<thead>
<tr>
<th>Edit and Imputation Criteria</th>
<th>Changes</th>
<th>Rationale</th>
<th>Resultant Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validity checks are done for each variable</td>
<td>Added by focus group: at source and at each transfer of data</td>
<td>It’s risk management. Data held at HealthPAC is not the same as what is recorded in the lab house …it’s come out of somebody’s system, already been validated … dump it into a warehouse. How much of that happens?</td>
<td>Validity checks are done for each variable at source and at each transfer of data</td>
</tr>
<tr>
<td>Data are collected from the health encounter not for claims/funding</td>
<td>Criteria added from ISSP but discarded by focus group</td>
<td>Why would they send it in data they don’t get money for? from pharmacy we don’t get any information on drugs that are under subsidy if the surcharge or whatever covers everything or for medicines that are not covered by the pharmaceutical schedule.</td>
<td></td>
</tr>
</tbody>
</table>

Table 17: Modification of the CIHI Edit and Imputation Criteria
### 6.5.3 Timeliness

<table>
<thead>
<tr>
<th>Data Currency at the time of Release Criteria</th>
<th>Changes</th>
<th>Rationale</th>
<th>Resultant Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>The recommended data quality documentation was available at the time of data or report release</td>
<td>And attached added by focus group</td>
<td>Just available or do you want it attached ... some might argue it’s always available ... yeah, available and attached</td>
<td>The recommended data quality documentation was available and attached at the time of data or report release</td>
</tr>
<tr>
<td>Major database/registry reports were released on schedule</td>
<td>Criteria removed by focus group</td>
<td>Actually fit under Documentation currency? yep…. And then this criteria would be just left out</td>
<td></td>
</tr>
</tbody>
</table>

Table 18: Modification of the CIHI Data Currency at the Time of Release Criteria

### 6.5.4 Comparability

<table>
<thead>
<tr>
<th>Data Dictionary Standards Criteria</th>
<th>Change</th>
<th>Rationale</th>
<th>Resultant Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data elements conform to the Ministry’s Data Dictionary</td>
<td>Added or are mapped to</td>
<td>Some of our database elements don’t exactly match at the moment, which implies a whole lot of work, going through all the programs that use that database.</td>
<td>Data elements conform, or are mapped to the Ministry’s Data Dictionary</td>
</tr>
<tr>
<td>Criteria added by focus group</td>
<td>If you’re comparing dollar signs for one collection and then dollar signs from another collection, we’ll know it’s the same information. You need a level of compliance, it needs to be reviewed and updated.</td>
<td>Common data elements across the Ministry are defined consistently</td>
<td></td>
</tr>
</tbody>
</table>

Table 19: Modification of the CIHI Data Dictionary Standards Criteria
Table 20: Modification of the CIHI Standardisation Criteria

<table>
<thead>
<tr>
<th>Standardisation Criteria</th>
<th>Change</th>
<th>Rationale</th>
<th>Resultant Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data are captured at the finest level of detail as is practical</td>
<td>Criteria removed by focus group</td>
<td>Data are captured at the required level of detail as is practical. And that, sort of, fits in with the specificity thing, which we had in other areas so … keep it within the accuracy.</td>
<td></td>
</tr>
</tbody>
</table>

Table 21: Modification of the CIHI Linkage Criteria

<table>
<thead>
<tr>
<th>Linkage Criteria</th>
<th>Change</th>
<th>Rationale</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Geographical Classifications can be used</td>
<td>Changed by focus group to include all coding tables</td>
<td>My understanding is, that you’ve got different tables floating around that potentially use, to code say domicile data too and no defined code table for some of the fields.</td>
<td>Each coded field in the collection uses accepted health sector coding table standards</td>
</tr>
<tr>
<td>Data are collected using a consistent time frame</td>
<td>Added by focus group ‘and format’</td>
<td>So it might be 2003, 08, 11 and one, 11 08 2003 another one and depending on the source, its not always easy to do, quite a lot of manipulation. So I think the word “format” in there is quite important.</td>
<td>Data are collected using a consistent time frame and format</td>
</tr>
<tr>
<td>Standardised codes are used to uniquely identify institutions</td>
<td>Added by focus group ‘Health sector’.</td>
<td>Acceptable institutions codes are available from the database, and all that would fail if we don’t have one standard</td>
<td>Health sector standardised codes are used to uniquely identify institutions</td>
</tr>
<tr>
<td>Standardised codes are used to uniquely identify persons</td>
<td>Added by focus group ‘Sector’. Metric changed to HPI or NHI</td>
<td>Agree – this is an area the centre needs to improve on – it’s a complex problem as a clinical variance may be resolved</td>
<td>Health sector standardised codes are used to uniquely identify persons. Metric changed to HPI or NHI</td>
</tr>
<tr>
<td>Equivalency Criteria</td>
<td>Changes</td>
<td>Rationale</td>
<td>Resultant Criteria</td>
</tr>
<tr>
<td>----------------------</td>
<td>---------</td>
<td>-----------</td>
<td>-------------------</td>
</tr>
<tr>
<td>The impact of problems related to crosswalks/conversions falls into one of the predetermined categories</td>
<td>Language changed by focus group to ‘relationships (one to one, one to many, many to many)’</td>
<td>Conversions is just a one to one, so we have got, one to one, many to one, and many to many. So should we get him to write it out or should we get him to say, you know, we know what we are doing and that you know what mappings are done, and you know how well it’s done.</td>
<td>The impact of problems related to relationships (one to one, one to many, many to many) falls into one of the predetermined categories</td>
</tr>
</tbody>
</table>

| Methodology and limitations of crosswalks/conversions falls into one of the predetermined categories | Changed by focus group to ‘mappings’ | Methodology and limitations of mappings falls into one of the predetermined categories |

Table 22: Modification of the CIHI Equivalency Criteria
### Historical Comparability Criteria

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Changes</th>
<th>Rationale</th>
<th>Resultant Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trend analysis is used to examine changes in key data elements over time</td>
<td>Metric changed by focus group from yearly to ‘as appropriate’</td>
<td>Something here like more periodically at least once a year or not, even twice a year. Once a year has gone, might not look at it for a year and suddenly, you got a whole year garbage data, I mean start seeing things on a month by month comparison with in six months, I would have thought. So it should be in relation to the context of the collection of the data in some cases it could happen continually, for each collection we define the frequency of what is appropriate for that collection</td>
<td>Trend analysis is used to examine changes in key data elements over time. Metric changed from yearly to ‘as appropriate’</td>
</tr>
<tr>
<td>The extent of problems in comparing data over time falls into one of the predetermined categories</td>
<td>Changed by focus group to 'The impact of historical changes are known'</td>
<td>Major key changes have to be documented, anything like that survey or collection for instances, changes methodology that needs to be documented so users, who use the data are aware of other factors that, you know, that may influence results</td>
<td>The impact of historical changes is known</td>
</tr>
<tr>
<td>Criteria added by focus group</td>
<td>Whatever or documentation for a survey to know that they are actually, this survey is a repeat of what the survey … or this series this data at a certain point. Practically it’s almost like a history</td>
<td>Other previous surveys or collections relating to this collection are known</td>
<td>--------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>

Table 23: Modification of the CIHI Historical Comparability Criteria
### 6.5.5 Usability

<table>
<thead>
<tr>
<th>Accessibility Criteria</th>
<th>Changes</th>
<th>Rationale</th>
<th>Resultant Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reports are published through HADIE</td>
<td>Added by ISSP review and ratified by focus group</td>
<td>I was thinking, it’s not just the actual data itself, but what about the metadata? That as well how we with all these tables and things but what about the metadata or what about the information … how we actually put out the survey, quite transparent. People actually using information about the quality of the information they are using</td>
<td>Reports are published through HADIE</td>
</tr>
</tbody>
</table>

**Table 24: Modification of the CIHI Accessibility Criteria**

<table>
<thead>
<tr>
<th>Documentation Criteria</th>
<th>Changes</th>
<th>Rationale</th>
<th>Resultant Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data quality documentation exists per annual subset release</td>
<td>‘Annual’ removed by focus group</td>
<td>Per sub set release might not be annual, that’s right, six monthly or monthly</td>
<td>Data quality documentation exists per subset release</td>
</tr>
<tr>
<td>Database/registry methods documentation exists for internal purposes per annual subset release</td>
<td>Removed by focus group – ‘for internal purposes per annual’</td>
<td>Do we want to you know, if we get all this information from the framework, how far out do we want to give it, and what do we want to give them, you know it’s quite a big question</td>
<td>Database/registry methods documentation exists per subset release</td>
</tr>
<tr>
<td>A caveat accompanies any preliminary official release</td>
<td>Removed by focus group ‘preliminary’</td>
<td>We do that anyway … really, it’s not really a preliminary, though is it. Any official release, they [CIHI] must be thinking along different lines than us</td>
<td>A caveat accompanies any official release</td>
</tr>
</tbody>
</table>

**Table 25: Modification of the CIHI Documentation Criteria**
Revision guidelines are available and applied per annual subset release

<table>
<thead>
<tr>
<th>Interpretability Criteria</th>
<th>Changes</th>
<th>Rationale</th>
<th>Resultant Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Removed by focus group ‘applied’ and ‘annual’</td>
<td>But not annually necessarily …so that revision guidelines for the first quarter publication it also refers to data services. I suppose anything produced basically has some sort of revision.</td>
<td>Revision guidelines are available and applied per subset release</td>
</tr>
</tbody>
</table>

Table 26: Modification of the CIHI Interpretability Criteria

6.5.6 Relevance

<table>
<thead>
<tr>
<th>Value Criteria</th>
<th>Changes</th>
<th>Rationale</th>
<th>Resultant Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>The level of usage of the data holding is monitored</td>
<td>Added by focus group ‘and types of users’</td>
<td>Where actually we have an idea of how much is being used and who by. So should we put level of usage and type of usage to make sure that the purpose of the collection is actually what it’s being use for?</td>
<td>The level of usage and types of usage of the data holding is monitored</td>
</tr>
<tr>
<td>User satisfaction is periodically solicited</td>
<td>Metric changed by focus group from 4 yearly to ‘on a regular basis or when indicated’</td>
<td>A lot of value in doing that, quite possibly, rather than every 4 years, you could almost do it prior to any planned upgrades. In fact, you have to. If anything affects the providers you have to consult. We did a client satisfaction survey after that and that gave us some quite useful feedback</td>
<td>User satisfaction is periodically solicited. Metric changed by focus group from 4 yearly to ‘on a regular basis or when indicated’</td>
</tr>
<tr>
<td>Criteria added by focus group</td>
<td>To measure improvements in satisfaction you really need to nail down some key questions, we always change the questions around how the survey is pointing at the time</td>
<td>A user satisfaction tool has been developed and is consistently used</td>
<td></td>
</tr>
</tbody>
</table>

Table 27: Modification of the CIHI Value Criteria
### 6.5.7 Privacy and Security

<table>
<thead>
<tr>
<th>Legislation Criteria</th>
<th>Changes</th>
<th>Rationale</th>
<th>Resultant Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>All data has been classified in accordance with SIGS (Security in the Government Sector Guidelines)</td>
<td>Criteria added by ISSP review and ratified by focus group</td>
<td>Well all health information is sensitive … that's what they have decided on, most information will be in confidence therefore private so our classifications in the future are going to be classified in accordance with e-government, to classifications, so put that in. Classification is carried out at the development stage and annually. I said you would have to review that, … time, I am not sure annually is the right time period.</td>
<td>All data has been classified in accordance with SIGS (Security in the Government Sector Guidelines)</td>
</tr>
<tr>
<td>Policies around archiving raw data have been developed</td>
<td>Criteria added by ISSP review and ratified by focus group</td>
<td>Data as in original data - we’ve got that elsewhere haven’t we that raw data should be stored, so there is a storage part … do they mean like, making sure we actually got the raw data from PHI and Stats NZ. Its just that we have tried. You really want to keep your raw data free from being modified, accidentally or intentionally. More accidentally is probably more likely.</td>
<td>Policies around archiving raw data have been developed</td>
</tr>
<tr>
<td>Unauthorised changes to data are prevented while the data are in transit</td>
<td>Criteria added by ISSP review and ratified by focus group</td>
<td>Unauthorised changes to data are prevented while the data are in transit</td>
<td></td>
</tr>
</tbody>
</table>

Table 28: New Legislation Criteria
<table>
<thead>
<tr>
<th>Disclosure Criteria</th>
<th>Changes</th>
<th>Rationale</th>
<th>Resultant Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adequate access control mechanisms have been implemented and access granted on a need-to-know basis.</td>
<td>Criteria added by ISSP review and ratified by focus group</td>
<td>That’s how we do this now, we assess each application for information on the individual needs of a person, on a need-to-know basis … on how much information they get, and there’s quite set criteria for that. Well there’s an access policy. There are sort of two different aspects to this. One is access to live systems for example and the other one is actually releasing data from those systems. So we have a data access policy but we also have a release of information policy from other data services</td>
<td>Adequate access control mechanisms have been implemented and access granted on a need-to-know basis.</td>
</tr>
<tr>
<td>Data privacy is maintained in accordance with legislations and available standards</td>
<td>Criteria added by ISSP review and ratified by focus group</td>
<td>it’s worth saying, so what’s the measure there …evidence … of meeting the requirements of the legislation ..and it might be, might be exception reporting produced on a monthly basis</td>
<td>Data privacy is maintained in accordance to legislations and available standards</td>
</tr>
<tr>
<td>The identity of an entity or user is assured before allowing access to classified data</td>
<td>Criteria added by ISSP review and changed by focus group ‘classified’ removed</td>
<td>Or any data really. We don’t have smart cards, do we? No but we could do, in the future, I don’t know how much we’re going to use that data classification, but we will have it, because we have to, so but it will change the level of access somebody gets.</td>
<td>The identity of an entity or user is assured before allowing access to data</td>
</tr>
<tr>
<td>Data are accessible and useable on demand by authorised entities of the data</td>
<td>Criteria added by ISSP review and ratified by focus group</td>
<td>Could take the levels of outages don’t fall below MOU's in place for users of that data. So there's documented levels? Yes for the major data sets, there are, there’s agreed levels to maintain below MOU requirements</td>
<td>Data are accessible and useable on demand by authorised entities of the data</td>
</tr>
<tr>
<td>Disclosure Criteria</td>
<td>Changes</td>
<td>Rationale</td>
<td>Resultant Criteria</td>
</tr>
<tr>
<td>------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Data used for ad hoc and statistical analysis does not contain patient/employee identifiable information</td>
<td>Criteria added by ISSP review and ratified by focus group</td>
<td>Depending on what the data are about, what sort of levels you are looking at, you can do a quick check list and think about all the pros and impact possibilities from releasing the data at this level.</td>
<td>Data used for ad hoc and statistical analysis does not contain patient/employee identifiable information</td>
</tr>
</tbody>
</table>

Table 29: New Disclosure Criteria
6.6 Review of the Altered Framework using the Eppler and Wittig Criteria

The resulting DQEF, with changes as recommended by focus group participants, is outlined below in Table 30. It was assessed using Eppler and Wittig’s (Eppler & Wittig, 2000) meta criteria. Eppler and Wittig (Eppler & Wittig, 2000) define meta criteria as ‘the criteria used to evaluate criteria sets of information quality frameworks’. The meta criteria were developed from articles and academic journal reviews, common sense, and interviews with practitioners. The authors note there is no general agreement on information quality framework meta criteria, therefore the researcher opted to use the criteria used by Eppler and Wittig. The criteria met the perceived requirements for a data quality framework both academically and practically, and the research participants considered this important.

The recommendations from the review centre on the development of business processes that support the effective use of the DQEF in the Ministry of Health environment, and the content of the manual that instructs users on the implementation methodology. Wu (Wu, 2004) defines a business process as ‘a set of logically related tasks performed to achieve a defined business outcome’. This supports Eppler and Wittig’s (Eppler & Wittig, 2000) theory that a framework requires tools, such as guidelines and manuals, to be effectively implemented. CIHI also found that a detailed step by step guiding process is required to implement a framework (Long & Seko, 2002). The Ministry of Health utilised the CIHI manual, but a considerable number of alterations were made following this review and the pilot study of the DQEF, detailed in section 5.7.
<table>
<thead>
<tr>
<th>Eppler and Wittig Meta Criteria</th>
<th>NZDQEF Response</th>
<th>Findings</th>
<th>Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definitions</td>
<td>Definitions of the dimensions and the characteristics exist and are provided.</td>
<td>References/index are not provided making it difficult to locate information quickly. Some definitions are unclear and simplistic and do not relate to the NZ health sector. For example, the use of nursing homes.</td>
<td>An index should be added to the manual. Content should be trimmed down and made more relevant to the audience (systems staff).</td>
</tr>
<tr>
<td>Positioning</td>
<td>Framework is positioned within the existing literature</td>
<td>The context of the framework is clear. The limits of the framework are not explicitly documented.</td>
<td>The limits of the framework should be explicitly documented</td>
</tr>
<tr>
<td>Consistency</td>
<td>Some confusion was experienced in understanding the differences in some framework criteria</td>
<td>If the assessors have not had training in the use of the framework and are not familiar with the collection, this will have some impact. The criteria in some respects are still subjective and so comparing across data collections may be problematic.</td>
<td>Assessors need training on the use of the framework. Data should be assessed by staff who are familiar with the collection.</td>
</tr>
<tr>
<td>Conciseness</td>
<td>The framework is not overly large. It took over four hours to answer all criteria for the collection respondents were not familiar with.</td>
<td>Once the assessor is familiar with the framework it should be easily remembered.</td>
<td>Training should be provided, and/or a trained assessor should assist the assessor. A pre-assessment checklist should be developed to assist assessors and to ensure conditions for an assessment and all required information are available.</td>
</tr>
<tr>
<td>Examples</td>
<td>The examples are not specific to NZ environment and did not seem relevant in some cases. For example: the use of nursing home, the use of Corporation instead of DHB</td>
<td>Providing NZ specific examples helps to guide users with contextual information</td>
<td>Develop NZ specific and illustrative examples to help explain the various criteria.</td>
</tr>
<tr>
<td>Other</td>
<td>Use of colour coding in the manual is meaningless when printed.</td>
<td>Unless the manual is printed in colour and distributed or read online – colour coding is of no value.</td>
<td>The manual needs to be coded in a way that does not depend on colour or only distributed electronically.</td>
</tr>
</tbody>
</table>

Table 30: Assessment of the New Zealand Data Quality Evaluation Framework (Eppler & Wittig, 2000)
The resulting DQEF, incorporating the feedback from the focus groups and the above review, was posted on a shared workspace website. Focus group participants could access the DQEF with password access and provide feedback on the changes. All participants agreed that the content of the framework reflected the focus group discussions. The User Manual was also revised with changes made to incorporate the feedback provided.

6.7 Pilot of the Framework at NZHIS

The DQEF, with the above changes made by the focus group and the review using Eppler and Wittig’s (Eppler & Wittig, 2000) criteria, was piloted on the Mortality Data Collection. Data quality on this collection is considered high, and the collection is used extensively. Table 31 below provides a summary of the uses of this collection for the financial year 2003/04, illustrating the considerable variation in users of the national data collections and the large number of ad hoc queries requested from information analysts. NZHIS provided 72 customised queries (regular reports) and 105 ad hoc queries.

<table>
<thead>
<tr>
<th>User type</th>
<th>Customised Reports</th>
<th>Ad hoc Reports</th>
</tr>
</thead>
<tbody>
<tr>
<td>Universities and Researchers</td>
<td>23</td>
<td>30</td>
</tr>
<tr>
<td>Government Agencies</td>
<td>7</td>
<td>31</td>
</tr>
<tr>
<td>Health Care Providers</td>
<td>9</td>
<td>19</td>
</tr>
<tr>
<td>Internal NZHIS</td>
<td>30</td>
<td>0</td>
</tr>
<tr>
<td>Societies</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>Members of the Public and others</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td><strong>72</strong></td>
<td><strong>105</strong></td>
</tr>
</tbody>
</table>

Table 31: User Type and Report Type Provided from the Mortality Data Collection

Members of the Clinical Coding Team evaluated the Mortality collection (as custodians of the collection) with the assistance of the Clinical Analysis Team Leader. A review led by the collection custodian and the Clinical Analysis Team Leader was considered to be the process by which all national collections would be evaluated. Table 32 summarises the evaluation results.
<table>
<thead>
<tr>
<th>Dimension</th>
<th>Deficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>Data quality control measures are lacking and highlighted when coding of cause specific survival rates. Missing records can be detected through the matching of hospital deaths information, but this is currently only done on an ad hoc basis. Measurement error is not accurately assessed at this stage, but could be through regular audit check if resources allowed.</td>
</tr>
<tr>
<td>Timeliness</td>
<td>Problems with timeliness are caused by delays with data from Coroners’ cases and with the considerable backlog of data not coded sitting at NZHIS. Work is currently underway to reduce this backlog. The Coroners have also recognised this problem and time delays have been reduced somewhat, although still considerable in some cases, particularly suicide deaths. Given the health and political interest in suicide statistics, it would seem prudent to make this information a priority. Documentation is also not provided in a timely manner, although the 'core data quality documentation' requirements are not yet defined within the Ministry.</td>
</tr>
<tr>
<td>Comparability</td>
<td>Outside of the business unit there is either no data dictionary or these have been developed in isolation. Equivalency is problematic as there is some mapping from old codes to new codes, but the documentation around the history of these changes is poor and where it exists, is hard to find. This is also not documented in the data dictionary for the collection.</td>
</tr>
<tr>
<td>Usability</td>
<td>There is no mention of the data quality documentation in the data dictionary, or reference in publications to the collections data dictionary and where to find it.</td>
</tr>
<tr>
<td>Privacy and Security</td>
<td>There is no written policy around the archiving of raw data, although this is done informally. The process of transferring data from Births, Deaths and Marriages is not known or documented; therefore there may be an unknown risk to the data in transit.</td>
</tr>
</tbody>
</table>

Table 32: Findings of the Assessment of the Mortality Collection using the Draft Ministry of Health Data Quality Framework

The information provided by the DQEF evaluation highlighted data quality problems that were already known to the data custodians, and most often work was already being undertaken to make improvements. This provided some validation to the ability of the framework to identify data quality problems applicable to the Ministry environment. Several other deficiencies were highlighted, indicating where improvements in the current data quality work could be made, or in the case of timeliness, highlighting known issues that perhaps should be made a priority for improvement.

The issue of timeliness of the suicide data, shown in Table 30, highlights the trade-off debate between dimensions. For example, if users of the data are aware that suicide data are not complete and the possible implications of missing data to the overall analysis, this data can be released earlier and therefore data becomes usable earlier. The user may need to decide which dimension is most important and wish to use the data immediately, without the delay of waiting for complete data. Alternatively, the completeness of the data may be considered more important than its timeliness, particularly if the impact of the missing data is likely to alter the outcome of analysis.

Eppler and Wittig (Eppler & Wittig, 2000) note trade-offs are not commonly addressed in frameworks. Research participants did not bring up this issue with the researcher and the DQEF does not explicitly address trade-offs between dimensions, however, an analysis on possible trade-offs and their implications would be appropriate to be included in the User Manual. Different trade-offs would apply to different collections, and again may be different for the specific uses of the data (Ballou & Pazer, 2003).
An issues register (Appendix 4) was kept to supply feedback to the researcher on the usability of the DQEF and User Manual. Overall, the DQEF was found by collection users and custodians to provide useful data quality information and to provide sufficient information to make at least preliminary prioritised lists of essential data quality improvement projects. A detailed analysis of the feedback is provided below.

6.7.1 Training Data Quality Practitioners

Further training has been required to ensure evaluators use the DQEF consistently and that it is a practical and easy tool to use. The CIHI also found that, due in part to already heavy workloads, considerable training and change management was required to implement the CIHI Framework (Long & Seko, 2002). Wang & Strong (Wang et al., 2001) and Redman (Redman, 2001) draw attention to the need to including training on assessment techniques in any TDQM process. During training sessions it was found that many users of the DQEF initially made assumptions about the meaning of criteria and needed to be reminded to refer to the metric and definition contained in the User Manual. Training ensures the data quality team have the knowledge and skills needed to carry out their tasks and ensures alignment of processes (Redman, 2001). The alignment of processes assists with consistency in the understanding of the definition for criteria. This is important, as many different users would complete the evaluation within each evaluation cycle and also from one evaluation to the next. For this reason a section was included under each criterion that asked the evaluator to describe in free text how they came to their decision, including references to any data quality information used to make that decision. Further discussion of the importance of training in data quality management is provided in Chapter Seven.

6.7.2 Time

As noted in the findings of the Current State Analysis of Data Quality in the Ministry of Health, data quality information was held in many different locations and attempting to find this information takes considerable time. Time taken, by already busy staff, to complete an evaluation of a collection was estimated with experience to be a minimum of four hours when all relevant documentation about a collection was available. In practice, the information was held in disparate locations by different staff and the initial evaluations took far longer. Repeated evaluations of the same collections would be completed much more efficiently as much of the information could remain the same or would merely need updating.

6.7.3 Intended Unit of Analysis

The researcher and participants also discussed the granularity of evaluation or intended unit of analysis. Participants asked, ‘was the DQEF able to evaluate a collection as a whole and also a column within a collection?’ Price and Shanks (Price & Shanks, 2005) found similar issues when implementing a data quality framework and noted the limitations of a framework on data that are not uniquely identifiable, such as that found in non-key columns. It was therefore considered appropriate to measure each collection as a whole. However, there are specific data elements that can be assessed individually and are extensively used for analysis and decision making. An example would be Registries, such as the NHI, where each field of
demographic information is the reference data for other collections and provides information to inform the decisions around the pre-existence of an allocated unique identifier for a person or place, thus preventing duplicate allocation. In this case, the DQEF can be used as a two dimensional tool for registries, assessing the registry as a whole and then each element.

### 6.7.4 User Manual

Extensive changes to the CIHI User Manual were required to make it useful to the New Zealand health care environment. The findings of this research found that changes were required to make the documents more easily navigated and the audience needed to be clearly defined so that the language could be targeted. Particular attention to the language used in the accompanying User Manual was required as the CIHI wording was found to be too simplistic for the intended audience. Those using the DQEF are likely to be systems administrators, data quality advisors, and members of the Business Intelligence Team, but the language implied the need for little underlying understanding of data and systems. The User Manual was shortened with less background information on data quality and this was produced separately to show the underlying theory used to develop the Data Quality Improvement Programme for those taking part in education programmes.

### 6.7.5 Data Quality Information

Feedback to the researcher from the Clinical Analysis team at NZHIS initially noted that the DQEF did not provide sufficient information to make decisions on the quality of data. The generic questions at first appeared to elicit insufficient detail to provide the level of data quality information required to manage data quality effectively. However, the feedback provided to the researcher became more positive once the team became more familiar with the DQEF and began to understand the context of its use. It was realised that the DQEF provided guidance on what measures should be regularly undertaken by the Clinical Analysis team to provide them with the answers to the criteria. These ‘answers’ provided a complete view of data quality on a collection but required new measures, not previously utilised by the Clinical Analysis team, to be undertaken. In effect, the DQEF was a suitable guide or checklist to enable appropriate and wide ranging measures to be undertaken consistently across all collections and has the ability to raise the standard of work expected in data quality by bringing about awareness of areas of deficiency in the current work programme.

Interestingly, findings from the Mortality data collection evaluation show that the Business Intelligence Team (as data consumers) requires the most detailed information on how the assessment was made for each criterion, whereas management requires only summary information. This variation reflects the different decisions made by the different groups and hence the distinctive uses and required granularity of data quality information. Clearly, reports on the outcomes of assessments should be tailored to provide this information specific to audience requirements. In addition, Chengalur-Smith et al (Chengalur-Smith, Ballou, & Pazer, 1999), Ballou and Pazer (Ballou & Tayi, 1999) and (Sutcliffe & Weber, 2003) all note the importance of the format of data quality information on decision making. Where simple decision making is required, complex data quality information was utilised. However, where complex decisions are required, complex data quality information was not utilised. They consider this may be due to ‘information overload’
where providing too detailed information might be counterproductive in more complex decision environments. In general, decisions made in the health care environment are likely to be complex; however, what is a complex decision to one user may not be complex to another. This researcher suggests that data users need to have input into decisions on the granularity of data quality information supplied. Chengalur-Smith et al (Chengalur-Smith et al., 1999) suggest that it may be most effective to provide users with data quality information that focuses on one or two key criteria which exhibit the greatest impact on the data being used. A later study by Fisher et al (Fisher et al., 2003) observes the impact of experience on the decision maker, whereby increasing use of data quality information is found as experience levels progress from novice to professional.

Overall, collection users and managers found the DQEF to provide useful data quality information and to offer sufficient information to make preliminary, prioritised lists of essential data quality improvement projects.

In summary, the pilot of the DQEF in the NZHIS environment elicited the following issues:

- training was required before using the DQEF;
- time to complete the evaluations was felt to be considerable by users;
- detail of information provided by the DQEF assessment process must meet the needs of data users, such as the Business Intelligence Team, particularly information around how evaluation decisions are made;
- granularity/units of analysis need to fit the type of collection, Registries such as the NHI require more granular, element level analysis;
- language in the user manual and DQEF needed to fit with the users of the tool.

### 6.8 Pilot of the Framework in a DHB

The purpose of this pilot was to assess the DQEF on the Clinical Integrated System (CIS) Model that is currently being used in clinical practice at Auckland District Health Board (A+), giving information on the applicability of the DQEF, and its chosen data quality dimensions in the wider health sector. The CIS Model is an interdisciplinary computerised model of patient care that replaces the paper notes, requiring all staff members to document care via the computer. The software program was developed by the A+ Network Centre for Best Patient Outcomes. This model has been in clinical practice since 2000 and has been used for 6,000 patients.

The dimension, characteristics, criteria, measurement and explanation were reviewed for:

- the appropriateness of the criteria for a clinical database;
- the documentation of data quality of the CIS Model compared to the framework’s criteria;
- the framework’s documentation clarity.
The team was able to undertake the assessment process with minimal training, through one hour of discussion with the director of the Centre and the researcher. Training emphasised the importance of referring to the User Manual for definitions and understanding of criteria. The group reviewing the criteria comprised of four programmers/analysts, a business manager and clinician with a nursing background. Each section of the framework was individually reviewed. The group then met to discuss and collate their findings.

Feedback provided to the researcher following the assessment of the framework using the hospital clinical data collection noted that:

*a data quality framework is an invaluable tool that helps to guide developers to produce robust and valid clinical databases.*

This is evident through the positive feedback given on the usefulness of the information provided by the assessment process and the applicability of the DQEF. The majority of the criteria could be applied to a clinical database held external to the Ministry of Health again as shown in Table 33 below. This table highlights that 52 of the criteria, out of a possible 69 criteria used in the DQEF, conform to the data quality requirements of the clinical database held at the local hospital level.

<table>
<thead>
<tr>
<th>MOH framework criteria</th>
<th>Hospital collection compliance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conformed</td>
<td>53 items</td>
</tr>
<tr>
<td>Not applicable</td>
<td>8 items</td>
</tr>
<tr>
<td>Did not conform</td>
<td>8 items</td>
</tr>
</tbody>
</table>

Table 33: Applicability of the DQEF with a Hospital Clinical Data Collection

The evaluation of the DQEF provided feedback to the researcher on each dimension and this feedback is provided in Table 34 below.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>We believe that within this section all the predetermined data criteria should be easily obtained.</td>
</tr>
<tr>
<td>Timeliness</td>
<td>We think there is a difference here if you are using concurrent clinical databases, as information must be continually fed back to users to help avoid adverse / or to detect problems quickly.</td>
</tr>
<tr>
<td>Comparability</td>
<td>This was seen as an important section, especially if benchmarking is to occur between hospitals.</td>
</tr>
<tr>
<td>Usability</td>
<td>We had to refer frequently to the examples and explanations and there was some confusion as to what was meant.</td>
</tr>
<tr>
<td>Relevance</td>
<td>The difference for us in this area is that if you have a clinical database it must be continually evolving to meet the changing needs of patients and clinicians, hence different criteria for judging relevance is needed.</td>
</tr>
<tr>
<td>Privacy &amp; Security</td>
<td>Agree with criteria, but first criterion is not applicable to our database.</td>
</tr>
</tbody>
</table>

Table 34: Feedback to Researcher By Data Quality Dimension on the Practical Use of the DQEF in a Clinical Setting
The criterion 'User satisfaction is periodically solicited', with a metric requiring four-yearly evaluation of user satisfaction was not suitable for this environment. The evaluation feedback stated:

*we think four years is too long and we think if you are working in partnership the need to formally solicit feedback is unnecessary – i.e. you can continually evaluate your effectiveness.*

This was also noted by the focus group participants as not being an appropriate metric and changed to yearly. The language used for the Usability dimension was further clarified. The evaluation using the DQEF proved valuable to the Centre and has resulted in the identification of areas where improvements in data quality practices are needed. The hospital data analyst, who managed the assessment process of the DQEF with the Centre, suggested that a formal, health-sector-wide data quality management criteria based on the DQEF, together with a certification process such as accreditation, would help to ensure that clinical databases are valid and reliable.

### 6.9 Implementing the Framework

The implementation of the DQEF required considerable time resources. The work centred on the need for business process development to support the use of the DQEF. Managers were required to make decisions about the roles and responsibilities of staff, how often an evaluation should take place for each collection and the processes around decisions on prioritising improvement projects. Therefore the DQEF is not a stand-alone document. The User Manual and business process documentation are required to support the consistent application of the DQEF and communicate policy around its use for each collection.

#### 6.9.1 Tools Developed to Support the DQEF

Along with the modification of the CIHI User Manual to meet the needs of the New Zealand environment the researcher developed a template to provide a structured evaluation-reporting tool. This template was subsequently used for all data collection evaluations to ensure a consistent report format. The headings used in the template were:

- Executive summary;
- Recommendations;
- Summary of dimensions, characteristics and criteria assessment;
- Introduction;
- Summary of assessment by dimension;
- Detailed assessment of collection.

In response to feedback from DQEF users, the researcher requested the Information Technology Services department to develop an automated tool that could:

- reduce as much as possible the time required to complete collection evaluations;
- enable the sharing of the evaluations amongst different members of the organisation;
• provide an electronic filing system for all data quality information.

The Clinical Analysis team member undertaking an evaluation using the automated tool initiated a template for each collection and was able to assign various people across the organisation to a section or all of the evaluation, reflecting the need to gain input from many members of staff. An automated email system informed staff of the allocation of the task and provided a link to the evaluation template, held in Lotus Notes. Each criterion asked for an evaluation rating, with the metric relating to that criterion visible on the screen in the hope that evaluators would make their decisions based on the metric. The evaluation was accessible at all times to the Clinical Analysis team, as coordinators of the evaluation process. A hyperlink was provided to the User Manual so that definitions for each criterion were easily accessible to assessors. It is hoped that the automated template reduced the time required to complete the evaluations by making them easy to update, as previous data quality information will be available in an electronic format that can be directly inputted during the assessment of a collection.

6.9.2 Objective and Subjective Metrics

Further refinements will improve objective metrics used within the DQEF. The DQEF uses subjective metrics for data criteria that have been developed by the data custodians. While this is a valid form of measurement, the robustness of the DQEF also requires additional objective metrics and these will be derived in part from a structured system based on statistical process control (Carey & Lloyd, 2001). The revised metrics will include measures that assess customer requirements for levels of data quality, trends in historical data within the national health collections, current key performance indicators for contracted data suppliers, and legislative requirements for the provision of data by health care providers and the Ministry of Health to international bodies such as the World Health Organisation.

Like CIHI, NZHIS found the concept of ‘fit for use’ extremely difficult to define for the national health collections when this needs to be defined by the consumer. A stakeholder analysis highlighted the extensive range of consumers, all with potentially varying quality requirements. Further, the number of ad hoc queries provided from each collection is considerably more than regular reports. Given these difficulties, and to ensure the DQEF remains generic across all collections, metrics will remain an approximate requirement, with consideration that the most stringent users determine the level of quality (Eckerson, 2002). Consistent metrics would allow for consistent decision making on priorities for improvement, when looking across all of the national collections. Whilst generic metrics are not ideal, a pragmatic approach is required. Miller (Miller, 1996) and Tayi and Ballou (Tayi & Ballou, 1998) note that data can be too accurate when its degree of precision exceeds the user’s processing capability, potentially leading to increasing costs of information systems, strains on the systems credibility, and misuse or abandonment of the data.

The development of metrics to measure the quality of ethnicity data provides an example of the process required to develop and implement effective metrics. A previous project aimed at defining a standard for the collection of ethnicity data provided the health sector with process guidelines on how ethnicity data should be collected, coded, classified and used. This standard was endorsed by the Health Information Standards Organisation and closely aligned with Statistics New Zealand processes for the collection of ethnicity data for
the four-yearly national census. This census data can be used as a benchmark for expected ethnicity rates within a given geographical population. Historical data were also available on the incidence of an ethnic group receiving health care as reported from health care providers. Therefore, the process of metric development for ethnicity data included:

- analysis of expected ethnic group numbers for each DHB geographical region from Census data;
- analysis of historical ethnicity data held at NZHIS;
- existing contractual requirements.

An expected range of values for each ethnic group was then defined for each DHB. Statistical process control allowed for the assessment of outliers from these expected rates. The Clinical Analysis Team informed, through an established ‘feedback loop’, those DHBs that consistently provide data outside the expected range and assisted with improvement measures. This process can be undertaken for each applicable criterion within the DQEF until the full picture of ethnicity data is provided.

KPIs have been developed for expected quality of ethnicity data for inclusion in data supplier contracts with the Ministry of Health. Bauer (Bauer, 2004) highlights the complexity of developing applicable Key Performance Indicators (KPIs). KPIs are defined by Bauer as ‘quantifiable metrics which reflect the performance of an organisation achieving its goals and objectives (Bauer, 2004)’. KPIs should reflect strategic value drivers; in the case of health care this can be the current health strategies in place and contractual requirements between the Ministry of Health and health care providers. Bauer’s pyramid below (Figure 10) illustrates the questions that should be asked when developing KPIs, from the overall strategic vision to key action initiatives.

![Figure 10: Strategic Alignment of KPIs (Bauer, 2004)](image)

The development of the ethnicity data KPIs reflects the value of high quality ethnicity data in the delivery of health care services. Several current health strategies note the requirement to improve ethnicity data in the New Zealand health sector, including the New Zealand Health Information Strategy (Ministry of Health, 2005), The New Zealand Health Strategy (Ministry of Health, 2000) and He Korowai Oranga, the Māori
Health Strategy (Ministry of Health, 2002). Further, changes to payment criteria for Primary Care Services mean that those providing health services to Māori will be paid higher subsidies. This supports the need to actively measure and evaluate the quality of ethnicity data through KPIs.

5.10 Chapter Conclusion

This Chapter provided a detailed analysis of the research results from the development and implementation of a data quality evaluation framework. NZHIS management recognised the need for structure around the management of data quality within the national health data collections. There was a recognised need for assessment tools that provide information on the level of data quality, to see where the problems really were, and who should be accountable. This would then begin to inform the possible solutions required. Previous work done to develop the Ministry’s ISSP and the Current State Analysis of Data Quality provided further support for the commencement of a programme of work on data quality improvement.

The DQEF was based on the work of the CIHI, but went through considerable iterations to ensure local applicability. A comparison with the CIHI and NZHIS environments showed that New Zealand appeared better placed to make improvements, as there is an existing national unique identifier for health care. An additional dimension of privacy and security was required to ensure this is explicitly managed to a minimum standard on all data collections. Further criteria were added following a review of the findings of the Ministry’s ISSP development.

Two focus groups with data users and custodians from across the Ministry provided rich feedback to the researcher on the applicable content of a local framework. Bringing this disparate group together also provided an opportunity for data quality to be discussed from across the organisation perspective and enabled the researcher to provide education on data quality issues.

The DQEF was then evaluated, using criteria developed by Eppler and Wittig (Eppler & Wittig, 2000), to assess the implications of the changes made by the focus groups. The recommendations from the review centre on the development of business processes that support the effective use of the DQEF in the Ministry of Health environment and the content of the manual that instructs users on the implementation methodology.

The resulting framework was piloted on two collections, the Mortality Data Collection at NZHIS and the Clinical Integrated System (CIS) Model at Auckland DHB. The results showed that the information provided by the DQEF evaluation highlighted data quality problems that were already known to the data custodians, and most often work, including education, was being undertaken to make improvements. However, several other deficiencies were highlighted, indicating where improvements in the current data quality work could be made. Users noted that ‘a data quality framework is an invaluable tool that helps to guide developers to produce robust and valid clinical databases’.
Overall Evaluation

6 Data Quality Dimensions
6. Privacy and Security

21 Data Quality Characteristics*

69 Data Quality Criteria**

Data Quality Analysis and Database Document Compilation

*0. not applicable, 1. unknown, 2. not acceptable, 3. marginal, and 4. appropriate
**0. not applicable, 1. unknown, 2. not met, and 3. met

Figure 11: The New Zealand Ministry of Health Data Quality Evaluation Framework Structure

Figure 11 above provides a summary of the resultant New Zealand Data Quality Evaluation Framework following the inclusion of feedback from the two Ministry focus groups, the pilot evaluation of the framework on two data collections and the review of the framework against the Eppler and Wittig meta criteria.

The ability of the DQEF to highlight already known data quality problems validates its applicability and usability in both environments. The DQEF also has the ability to highlight areas where data quality methodology could be improved on in the current situation. The Clinical Analysis Team is prompted, by the results, to raise the standard and breadth of work undertaken to improve data quality by using methods that data users and custodians agree are important and formalising decisions on data quality improvement priorities into a more structured process.

Consideration of the business processes around the implementation of the DQEF highlighted that it is not a stand-alone tool. The User Manual and business process documentation is required to support the consistent application of the DQEF and communicate policy around its use for each collection.

The development of applicable metrics proved difficult in the NZHIS environment and will continue to be refined through the better understanding of data quality by the Clinical Analysis Team. Metrics for the assessment of ethnicity data have been defined, and the process used may provide a guide for further metric development methodology.
For the development and application of the data quality evaluation framework, it is important to:

- define the underpinning data quality criteria carefully, involving all stakeholders to ensure common understanding and direction;
- consider the critical quality dimensions that reflect how the organisation uses data and how data flow throughout the business processes;
- document business processes identifying data sources and their reliability;
- appreciate that the framework requires practical supporting tools, e.g. documentation, to make it effective;
- customise the language of the data quality user manual with regard to the level and experience of the intended users;
- be aware of the importance of both education and training at all necessary stages and levels – training is essential to effect the culture change that must accompany the realisation of the importance of data quality;
- be aware that application of the framework is an iterative, on-going process – the required outcomes cannot be achieved in a single pass.

The DQEF remains an iterative tool, whereby those who use the tool will improve its usefulness with growing knowledge of data quality theory, the level of data quality of the national health collections, and where the priorities for improvement lie.
7 The New Zealand Health Sector Data Quality Improvement Strategy

This Chapter provides the results of the process to develop a data quality strategy for the New Zealand health sector that seeks to ‘institutionalise’ data quality through repeated and consistent data quality processes. The specific research questions answered within this Chapter are:

- What are the data quality dimensions considered important to measure and improve by data consumers, collectors and custodians in the New Zealand health sector?
- What initiatives assist to raise the profile of data quality improvement in the New Zealand health sector?
- What are the data quality roles found and by which stakeholders, in the New Zealand Health sector?
- Does action research methodology provide appropriate research methodologies for understanding learning and change initiation in a data quality improvement programme?

7.1 Survey of Organisations Outside of the New Zealand Health Care Sector

The researcher undertook a survey of data quality work in organisations both within and outside of New Zealand. The survey tool was designed to provide a comparison with the New Zealand health care organisations’ data quality capability maturity and the results of this comparison analysis are discussed in Section 7.5. Responses were provided from 15 different organisations where participants were asked to complete a one-page questionnaire providing free text answers. Content analysis provided for a summary of the themes elicited from the range of responses provided.

All of the respondents had a role in data quality, some with the title of Data Quality Manager or Data Quality Analyst. Some worked in the area of business intelligence, where data quality was one of their many roles. There were respondents who worked solely on data quality in very large organisations, where they developed central requirements for the organisation’s branches, some of which are in other countries.

Of the three respondents whose organisations have a data quality strategy, all noted it was either implemented only recently or not yet implemented. Extensive data quality strategies, existing in their own right, appear rare and most report having ‘some components of a strategy’. There was consensus amongst all the respondents that the data quality strategy should be linked to the business strategy and ISSP, and many organisations had components of data quality requirements in other strategies, such as ISSPs, information management strategies and business strategies. The programmes and strategies were almost entirely developed through intuition, the bringing together of knowledge from other disciplines and in response to particular problems. TDQM literature and Six Sigma theory was also noted as providing
guidelines for strategy and programme development. Strategies (or the components of) included the definitions of roles and responsibilities, data access policies, the link to the business strategy and management goals, standardised data definitions and management of data quality dimensions. In general, the strategies or components had not been tested in any formal way, although one organisation undertook a pilot of a proposed work programme before rolling out to the rest of the organisation. Few organisation-wide data quality teams were in place. The strategic documents covering data quality were high level and generally did not include the daily practices required to manage data quality.

Programmes of work are in place to manage data quality, but much of the data quality work centres on ‘find and fix’ solutions, along with extensive use of auditing, validation, and edit checks. Most respondents noted limited preventative measures are in place. Those that are used included ensuring it was easy to input accurate data through information system design. Data quality practitioners in large organisations tend to provide advice and standards requirements to departments and branches, rather than perform improvements themselves. Some of the respondents reported their organisations have been working on data quality measures for three to eight years, the longest being 13 years. Software is used by some of the organisations to automate cleansing work. One organisation has developed their own software tool to manage all aspects of the data quality programme, including assessment, with two others using automated scripts for assessments. The other organisations do not have any formal assessment tools that are used across all data, but one used statistical process control measures for some data and two have customer surveys for assessing customer satisfaction rates.

The implementation of new collections or warehouses often highlighted significant issues with the data and organisations did not wish to move bad data into a new collection. Therefore, considerable efforts were made to prevent this through improvement projects. The economic value of good information to the organisation was also an impetus for improvement, however this was surprisingly only mentioned once.

Constraints found by data quality practitioners when attempting to implement improvements were varied but management support was the most common issue. Support from management was found to be lacking due to limited understanding of data quality issues and implications. The skills of general staff were also a problem, particularly where the culture of the organisation did not support high quality data quality management, such as where ‘silo’ departments exist and data quality is not seen as an across-organisation issue. Where roles and responsibilities are not clearly defined, difficulties in resources are found. Dedicated resources are required but often not in place. Competing priorities for money and resources were also found to constrain improvement initiatives.

All respondents felt there were areas where their data quality work could improve. The absence of a data quality strategy aligning with the organisation’s overall strategies was found to impact on the ability of organisations to reach their goals. Data quality decisions needed to be made by those in roles that enabled a view of the entire organisation and its issues and with an understanding of effective data quality management practices, but these roles were not in place. Participants thought a formal methodology for data

12 These organisations are not identified to maintain confidentiality
quality management would be effective for improvement. Improvements in process management also needed more work. Providing KPIs in staff job requirements and advancing the knowledge of all staff could also improve data quality work.

The results of this survey informed the development of a subsequent survey (the Current State Analysis below) used throughout the New Zealand health sector to ascertain current data quality practices. The limited evidence of data quality strategies in most organisations indicated that they would be unlikely to be in place in New Zealand organisations. Therefore, the questions used in the New Zealand health sector were adjusted to elicit any data quality programmes, projects, or work taking place with less emphasis on a data quality strategy.

7.2 Current State Analysis of Data Quality in the New Zealand Health Sector

The researcher undertook an extensive Current State Analysis of data quality in the New Zealand health sector and the results provided the Ministry of Health with the impetus for strategic change. The national health collections required a sector wide, standardised approach to data quality and unless a programme of work was placed in the context of a national strategy, no improvements would work. There were no instructions about how data should be collected (the process of data collection) at the health provider level. A need was identified for a national level team to provide guidance to data suppliers on what data collection processes need improvement. There was no clear accountability for data quality in the health sector and the varying requirements for quality between centrally held and managed collections and those ‘at the coalface’ meant there was considerable uncertainty, particularly around roles and responsibilities.

Management at NZHIS noted:

You need a national strategy that provides improvements for everyone that everyone buys into. The role of the strategy at this stage of the evolution is to create awareness across the system around data quality to the extent where people want to embrace a solution. This solution needs to be broken up into bite size chunks i.e. to create an awareness that data quality is important, provide tools that show how this can actually be achieved and then to pick off some key low hanging fruit that might demonstrate to the business units how the strategy will have a positive impact. You can then concentrate on the harder stuff. Once you create awareness you get people taking the issue and driving their own improvements. This has been evident within NZHIS where the Clinical Analysis team have already begun improvement programmes.

Initial research, discussed in this section, was undertaken to determine sector stakeholders’ views of data quality through a survey to validate the data quality dimensions used in the DQEF. The sector organisation’s data quality programme maturity and readiness for change was determined by face to face interviews with stakeholders, including Chief Information Officers (CIOs) of DHBs and Shared Support Agencies. The capability maturity of the New Zealand health sector was compared with international capabilities.
7.2.1 Data Quality Dimensions Survey

The Current State Analysis began with a survey of Health Informatics Conference attendees, who were asked to rate the importance of the six data quality dimensions found in the DQEF and to suggest any other dimensions that should be measured and improved. These dimensions form the basis of the data quality strategy and overall programme, therefore, the researcher considered it was important that the health sector agreed with this set of dimensions. The results show that the dimensions found in the DQEF meet agreement from a sample of data users, custodians and collectors in the health sector.

Thirty Six percent (72) of conference attendees (n = 259) completed the questionnaire. Table 34 below illustrates how the respondents defined themselves, when asked ‘do you consider yourself to be a data custodian, collector, or consumer?’ (for definitions of these terms see section 1.5.2). Respondents were able to circle more than one response as many in the health sector have several roles. Table 35 also provides the results of the survey by type of respondent when asked to rank, from one (= most important) to six (= least important), the most important data quality dimensions to measure and improve.

<table>
<thead>
<tr>
<th>Type of Respondent</th>
<th>Number</th>
<th>Accuracy</th>
<th>Timeliness</th>
<th>Comparability</th>
<th>Usability</th>
<th>Relevance</th>
<th>Privacy &amp; Security</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumer</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Collector</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Custodian</td>
<td>18</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Consumer and Collector</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Custodian and Collector</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Consumer and Custodian</td>
<td>10</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Collector and Consumer</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>All three</td>
<td>7</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Not defined</td>
<td>13</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td>5</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>72</strong></td>
<td><strong>9</strong></td>
<td><strong>27</strong></td>
<td><strong>43</strong></td>
<td><strong>25</strong></td>
<td><strong>35</strong></td>
<td><strong>43</strong></td>
</tr>
</tbody>
</table>

Table 35: Results of Data Quality Dimension Survey Rankings by Type of Respondent

The results of the survey show that all respondent groups considered the accuracy dimension most important. Five groups considered usability to rate second: consumers, custodians, consumer and collector, custodian and collector, all three (consumer, custodian and collector). Three groups considered timeliness to rate second: collectors, custodians, and collectors and consumers. There appears to be a reasonable consensus amongst respondents that accuracy, usability and timeliness are the most important data quality dimensions. This result may reflect the types of issues that are most prominent at the present time in the New Zealand health care sector. The context of the use of the data influences the dimensions considered important (Lee et al., 2002). There is evidence in the literature from overseas that supports these findings (Carson, 2001).
Table 36: Overall Rankings of Data Quality Dimensions by all Respondents

Table 36 notes the overall rankings from all responses. The total score is the total of all scores provided by all respondents for each dimension. For example, 72 respondents provided a score from 1 to 5 on the importance of the accuracy dimension, providing a total of 213. This data support the findings from data in the previous table. Standard deviation figures reveal consistency in the individual scores from respondents.

Therefore, the overall ranking of importance for data quality dimensions was found to be:

1. Accuracy
2. Usability
3. Timeliness
4. Relevance
5. Privacy and Security
6. Comparability

The survey confirmed that the six data quality dimensions used in the DQEF are those that are considered important to measure and improve by sector stakeholders. During discussions with respondents, the overwhelming response was that it was extremely difficult to rank which dimensions were most important because ‘all of them are important’. Lee et al (Lee et al., 2002) note in their survey of importance rankings that using a weighted average model to analyse their data provided little difference between weighted and non weighted responses. Free text answers to the question, ‘are there any other dimensions we should be assessing and measuring?’ provided no new dimensions, but often noted characteristics or criteria already found in the DQEF. The full list of the DQEF characteristics and criteria can be found in Appendix 5.

7.2.2 Interviews with Sector Stakeholders

The researcher undertook 15 face to face, semi-structured interviews to elicit current data quality practices within the health sector. The interviews produced considerable free text data that were rich in content and provided an extensive understanding of current issues, practices and knowledge of data quality in DHBs, PHOs and Shared Support Agencies.

Grounded theory was used to develop thematic categories from the data. Initially, high-level categories were developed. All codes were clearly defined to ensure the researcher remained consistent in allocating data to codes. A code was allocated and defined when the researcher repeatedly encountered the theme, having become intimately aware of the data through repeated coding iterations. Strauss (Strauss, 1987) points out...
that grounded theory does not require entirely inductive coding, as the researcher may have previous knowledge of the subject and apply this knowledge when developing initial codes. Inductive coding would require that codes only be developed directly from data. Having undertaken much of the research on the development of the DQEF by this phase of the research, the researcher had developed considerable understanding on data quality theory and the possible issues that could be encountered. The researcher, therefore, asked generative questions of the data throughout the coding process, and these targeted questions also assisted with the development of initial codes. Generative questions are defined by Strauss (Strauss, 1987) as ‘questions that stimulate the line of investigation in profitable directions; they lead to hypotheses, useful comparisons, and the collection of certain classes of data’. The nodes and their definitions are provided in Table 37 below.
<table>
<thead>
<tr>
<th>Node</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accountability</td>
<td>Who is or is not accountable for data quality, any issues relating to accountability</td>
</tr>
<tr>
<td>Accreditation</td>
<td>Any reference to accreditation of data suppliers in the past or future</td>
</tr>
<tr>
<td>Business management knowledge</td>
<td>Management knowledge of data quality issues and theory</td>
</tr>
<tr>
<td>Clinical coding</td>
<td>Any reference to the clinical coding of data or of participants from the clinical coding area</td>
</tr>
<tr>
<td>Creating change</td>
<td>Anything related to change management</td>
</tr>
<tr>
<td>Data collection</td>
<td>Any reference to collecting data, the issues associated with it. Divides into systems and processes where possible. Reference to data collection processes, the people side, movement of data, method of collection. Reference to data collection systems, the technology</td>
</tr>
<tr>
<td>DHB</td>
<td>Any reference to District Health Boards and the data related to District Health Boards</td>
</tr>
<tr>
<td>Documentation</td>
<td>Any reference to data quality documentation or data management documentation</td>
</tr>
<tr>
<td>DQ Dimensions</td>
<td>Heading code for any reference to data quality dimensions. Should be divided out into each dimension where mentioned in the text</td>
</tr>
<tr>
<td>DQ Programme</td>
<td>Any reference to data quality work. Overall definition of data quality programme, needs to be further defined in child nodes below. Assessment, audit, edit checks, feedback loops, find and fix, prevention, reports, training</td>
</tr>
<tr>
<td>EHR</td>
<td>Reference to electronic health records</td>
</tr>
<tr>
<td>Impact of data quality</td>
<td>Evidence of the impact of data quality in how data can be used, clean up projects required, costs and resources required to work on data quality</td>
</tr>
<tr>
<td>Level of data quality required</td>
<td>Decisions around what level of data quality you need to make data useful</td>
</tr>
<tr>
<td>Link to ISSP or IM Business strategy</td>
<td>Any reference to data quality programme or strategy linking to the Information Systems Strategic Plan (ISSP) or business strategy, now or in the future</td>
</tr>
<tr>
<td>Management support</td>
<td>Any reference to support from management to undertake data quality work of any kind, make change or provide resources</td>
</tr>
<tr>
<td>Moving data around</td>
<td>Moving data from one collection to another or between organisations or business units</td>
</tr>
<tr>
<td>New data collection</td>
<td>Reference to setting up a new data collection</td>
</tr>
<tr>
<td>Ntl health data collections</td>
<td>Any reference to a national health data collections</td>
</tr>
<tr>
<td>Org wide dq team</td>
<td>Organisation wide data quality team, do they have one or not. Also a high level code for notation of source of data</td>
</tr>
<tr>
<td>Organisational attitudes</td>
<td>Any reference to the way the organisation values data quality, only where the text notes this is a 'whole of organisation (cultural) issue'</td>
</tr>
<tr>
<td>Ownership and stewardship</td>
<td>Any reference to the ownership or stewardship of data</td>
</tr>
<tr>
<td>People problem</td>
<td>Problems with the management of staff who collect or manage data</td>
</tr>
<tr>
<td>Primary Care</td>
<td>Any reference to primary care and the data related to primary care</td>
</tr>
<tr>
<td>Priorities for improvement</td>
<td>Any reference to prioritisation of projects for data quality, or the prioritisation of data quality itself when competing for resources with other, has return on investment been considered or measured</td>
</tr>
<tr>
<td>Process Management</td>
<td>Any reference to managing processes within the organisation</td>
</tr>
<tr>
<td>Reporting requirements to MOH</td>
<td>Any reference to mandatory reporting to any section of the Ministry of Health</td>
</tr>
<tr>
<td>Node</td>
<td>Definition</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Roles</td>
<td>What are the roles and responsibilities and who is allocated them</td>
</tr>
<tr>
<td>Source of text - where are the participant(s) from?</td>
<td>Who was interviewed - DQ team, international DQ survey, Ministry meeting or discussion, DHB CIO, HealthPAC, Ministry document, NZHIS, shared support agency</td>
</tr>
<tr>
<td></td>
<td>Sources - DQ Dimensions survey, email, observation note</td>
</tr>
<tr>
<td>Strategy</td>
<td>Reference to any strategy, data or otherwise</td>
</tr>
<tr>
<td>Uses of data</td>
<td>Any reference to how data are used</td>
</tr>
<tr>
<td>Whole of system</td>
<td>Any reference to a systems perspective</td>
</tr>
</tbody>
</table>

Table 37: High Level Grounded Theory Codes and their Definitions

Data collection was considered complete when no new overall categories were required to code the data, indicating that no new themes were emerging and that ‘data saturation’ was reached. Strauss and Corbin (Strauss & Corbin, 1998) note, however, that through extensive examination of the data new themes can almost always be found, but when collecting new data does not add anything further to the research findings the researcher can consider this data saturation.

The initial overall categories were then examined for uniqueness to ensure there were no overlapping categories and none were found. Further sub-categories were then defined to break down the data into ‘child’ nodes. Child nodes are used where it is possible and useful to divide the code further to enable more detailed understanding of the text at that code. For some child nodes, ‘in vivo’ coding was used where appropriate to ensure the analysis remained embedded in the data. In vivo codes are codes derived directly from the words of study participants (Strauss & Corbin, 1998). Figure 12 below provides an example of the breakdown of the initial high level code ‘Roles’ and ‘Organisation wide data quality team’ into subcategories and subsequent child nodes.
These child nodes were then coded using axial coding techniques to explore nodes in detail, searching for relationships, dependencies, and ultimately concepts. Key concepts emerged from the data and provided the researcher with understanding of data quality management and issues within the New Zealand health sector. For example, the relationship between the data coded at 'organisation wide data quality team – yes – impact' was analysed against the data coded ‘Roles’. This provided details on the impact the presence of a data quality team had on roles within the organisation. The key concepts are examined in sections 6.2.3 to 6.2.14 and summarised in section 6.2.15.

7.2.3 Uses of the Data

Discussions with respondents revealed that the national health collections are used extensively throughout the health sector. They are relied upon for decisions around policy and strategy at all levels of the health sector. The Shared Services Agencies analyse regional and national aggregated data to provide information on service delivery back to the DHBs and PHOs which they support. PHOs and Shared Service Agencies undertake similar analysis work for their members and provide them with comparative analysis as a benchmark against other DHBs or PHOs, service analysis around what services to provide related to disease rates in the population, reports on screening rates, KPI reporting, primary care utilisation and expenditure for referred services management. Funding and planning teams make extensive use of clinical and administrative data to ensure population based funding service delivery is managed in the most effective way. Productivity indicators and the mix of procedures and diagnoses provide important data. Some data are obtained from outside of the health care domain to provide a holistic view of population health needs, such as data pertaining to population socio-economic status.
Data are also used to improve clinical practice. Local researchers such as nurses use data for improving nursing outcomes, and decision support departments analyse patient outcomes to advise on clinical best practice. One PHO undertook a project to improve the prescribing habits of GPs and a Shared Service Agency also plans to undertake a similar project within a DHB:

\[
\text{we used the data to improve prescribing habits of GPs, so we saved a lot of money that we can then use to provide preventative services like smoking cessation. Lab tests ordering was also improved. So having the right information is very effective.}
\]

Data quality issues were noted with some data, which caused ‘frustration’ with data users, with two national collections holding data ‘that are not of sufficient quality to use for decision making’. The most common theme to emerge from data coded ‘Uses of data – Management’ was the insufficient data quality of reports provided to DHB managers, so that on occasion managers stated they ‘could not trust the data’.

Managing the quality of these data is a multidimensional issue with the extensive and wide ranging uses of the data by many stakeholders. The value of the data to patient care and management of health care services was evident, with increasing expectations that data will be merged and mined to provide rich information on clinical care and outcomes. A survey of DHB Chief Information Officer’s activity undertaken by the Ministry of Health in 2003, (Ministry of Health, 2003a) noted that the focus of the CIOs work in DHBs is shifting to providing more clinical functionality for the integration and provision of information into and from community and primary care settings, and for sharing information between the DHBs.

### 7.2.4 Accountability

Many of the respondents of this survey noted there was no one person ultimately accountable for data quality in the organisation; the role is the responsibility of the CIO ‘by default’. However the CIOs have varying positions and roles within health care organisations. Sixty percent of DHB CIOs report directly to their executive team, either the Chief Executive Officer or the Chief Financial Officer. Others report to General Managers from service departments (Ministry of Health, 2003a). CIOs felt that management saw data quality as an IS problem. One CIO did note that unclear responsibility for data management in general causes problems with accessing information.

\[
\text{That’s a good question, it’s a bit of an issue for us and there is no one service that is responsible for data. We don’t have an information management department, mine is the IS department. Where the people who want information for reporting and management, where they get the data from tends to be a bit disjointed at the moment. No one has the responsibility in his or her employment contract.}
\]

It was often noted that data quality initiatives were undertaken in ‘silos’, in department areas as and when needed, with no improvements made to the organisation’s systems as a result of these initiatives. The data are fixed for a specific need in that area, with resulting reports that conflict with reports from the organisational systems provided to management. This was often found to occur in the decision support departments where information analysts provide separate reports on clinical and administrative data back to the organisation.
On two occasions respondents reported accountability lay with the information systems team, a general rather than specific accountability. Several respondents noted data quality was the responsibility of the departments or department managers, although they were unaware if data quality was a component of department manager’s performance measurements. Responsibility for data quality across the organisation’s departments is preferred by most CIOs.

7.2.5 Roles and Responsibilities

There are a considerable number of roles that are in some way involved in data quality. The roles are not always clearly defined, and have developed over time in response to data quality problems in different areas. Smaller DHBs noted there were often roles where data quality was one of many tasks for one position. Sometimes this was beneficial, as application managers had extensive knowledge of the data in that application and the specific data quality issues, and were able to identify incorrect data using this knowledge. Some larger DHBs noted that there were disparate areas working on data quality, with no central point of contact, so each area had no knowledge of the work being undertaken in other areas. One CIO noted:

_We have two departments which have data quality functions, one of which is business information unit which is responsible for the Ministry of Health extracts and reporting activity based and financial data to the organisation. Also we in IS have an interest through our systems administration functions of the major systems. Other departments will have a systems administrator and part of their role is data quality._

Considerable data quality work was in the information system (IS) department, either by default or because management and departments see it as their role, along with the delivery of information. The work centres on supporting specific applications and feeding back problems to departments. The IS department sometimes supports data quality projects managed by other departments and is sometimes expected to pay for improvements, supply staff and provide training. The information systems department does often appear to be the best place to have an overview of data across the organisation, as they support and understand the organisation wide information systems.

Decision support departments play a major role, and in one DHB the CIO noted that data quality was ultimately the decision support department’s responsibility. They often worked independently however, without any liaison with the IS department on data quality issues. Finance areas are noted to have analysts who work closely with data, make improvements to data, and one worked closely with the IS team by feeding back potential issues when unusual data values were found.

The Quality and Risk team were mentioned by several respondents, some leading projects with the assistance of the IS Department, others note there are staff with data quality responsibilities in these teams. It is noted that these teams often have a different focus to the IS Department, specifically focusing on patient safety issues, whereas the IS department would have an administrative and clinical information focus.

There was considerable variation in the role of the department or service delivery managers. Some CIOs saw the responsibility for data quality, at least implicitly, to lie with these managers. It appears that no
department managers take an active role in leading data quality improvement, but some do support and work with the IS departments, provide feedback on issues to the IS department, and ensure their staff attend training. Their role appears to be extensive and their support and understanding of data quality important. Respondents often identified department managers as the person they would go to, to initiate improvements in that department.

### 7.2.6 Organisation Wide Data Quality Teams

Three DHBs have established organisation wide data quality teams. The impetus to form these teams is usually an awareness of problems that span the organisation, identified initially by the IS Department. One data quality team established and led by the CIO had extensive membership from across the organisation and defined terms of reference, and was active in identifying issues. The CIO commented on the need for such a group and its role.

> We had a group that I went to when I first arrived. But the issues that were coming up were management performance issues not data integrity issues to in the end I stopped going and the group ceased to exist. Towards the end of last year it was obvious there was a need for it as issues were raising their head all over the organisation so I said I'm prepared to kick off the group again but here are the rules. A I want to chair it and B because it's cross department I need the authority to go and finger people in other departments if needed and it has to have visibility and they agreed.

Membership included:
- IS team members;
- clinical coding manager;
- mental health staff (with an information focus to their role);
- medical records manager;
- administration manager and ‘super user’;
- finance department staff;
- booking systems coordinator (elective surgery booking system);
- training coordinator;
- quality assurance team member;
- rehabilitation team member;
- clinical super user (position not yet filled).

This team met every two months and discussed issues around the capture and processing of data in an effort to standardise these processes across the organisation, to exchange knowledge, support department managers to identify issues, prioritise projects, update on new reporting requirements, provide feedback to department managers on data quality issues, and training and performance management. This was the only data quality team with membership that included clinical staff, although this position was still vacant.

The second team was also well established and comprised mainly funding and planning staff and two IS team members with job titles relating to data quality. Such job titles were not found in other health care organisations. The team was established by the IS staff and did not involve the CIO or senior managers. This team does not yet have terms of reference but meets monthly to discuss data quality issues across the organisation, with an emphasis on funding and planning issues.
The third team was only recently formed by the IS department staff, who felt there was a need for improved data quality, but no one person or department was responsible. They felt that, by default, the data quality team is responsible for data quality, but this had not been formalised in any way. They reported considerable support from management, who stated they were ‘just glad someone was dealing with it’. Membership included three members of the IS department, one of whom is responsible for IS training, and the team leader from the clinical coding department. The team stated their intention was to develop the membership further to include staff from around the organisation, along with the mental health department and clinical staff.

There are a number of committees or teams, e.g. ‘the Health Information Committee’ throughout DHBs that discuss data quality issues as they arise. Although this is not the sole purpose of these teams, CIOs noted that this is where data quality issues were likely to be addressed, as there was no established data quality team. Data quality was not always a standing agenda item, and was discussed only when issues arose. One CIO notes the disparate management of data quality, reactive to problems when they become apparent:

> We have a group called the IS forum and one called the health information committee and those are the groups in which data quality issues could arise but they are not a standing agenda item when they are perceived to be a problem.

Discussion of the impact of data quality teams can be found in Chapter Seven.

### 7.2.7 Priorities for Improvement

Data quality is given a low priority (Tayi & Ballou, 1998). Whilst everyone agrees that data quality is a valid and important activity, in practice few people list it as a top priority (Tayi & Ballou, 1998). Data quality practitioners were required to work hard at ‘convincing’ management that improvement costs or human resources were warranted. Many voiced ‘frustration’ in general around data quality work, often specifying the lack of management support due to competing priorities as a significant factor in this frustration. Further frustration was voiced at the lack of preventative work able to be completed, the lack of staff available to deliver training, the time required to find and fix existing errors and the significant difficulties due to the lack of role definition and accountability for data quality.

Due to demand, the funding and planning requirements of DHBs have previously received a high degree of focus by the CIOs. Therefore the provision of information to support population-based health care is less developed and now a focus of improvements. The 2003 DHB CIO Survey (Ministry of Health, 2003a) found the three main focus areas for the upgrade or implementation of new systems were:

- the provision of clinical information through clinical systems;
- implementation of integration with primary care;
- discharge summaries and electronic referrals.

CIOs report extensive project work activity, with two DHBs managing more than 40 projects at once, although most (nine) reported less than 10 (Ministry of Health, 2003a). CIOs noted the increased work required to implement the New Zealand Health Information Strategy (Ministry of Health, 2005), with ‘12
Action Zones’ expected to be implemented within five years. These factors provide considerable competing priorities against data quality work.

This survey found that competing priorities were often stated as the reason for a perceived lack of support from management on data quality improvement, and data quality was only addressed when a problem arose. Clinical work was noted to take priority over data entry for clinical staff and respondents felt that data quality improvement should pinpoint where clinical decisions were made using that data.

When asked what the priority for data quality improvement should be for the whole of the health sector, most respondents highlighted the national data collections. The NHI was thought to be a significant register and that reducing duplicates on the NHI was very important, as was improving the quality of NHI demographic data. Indeed, there are considerable resources in the DHBs, of both time and money, concentrating exclusively on NHI duplication merges. Many DHBs have a full-time staff member dedicated to this role. Further, respondents often noted the need for data that are not currently collected at a national level including outpatient and data primary care data. Both of these collections are action zones in the Health Information Strategy (Ministry of Health, 2005).

7.2.8 Return on Investment

There is no formula for calculating return on investment (ROI) for data quality projects used in the New Zealand health sector. Many providers do not spend extra funding on data quality improvement, with the exception of major clean up projects, prior to moving data from a legacy system to a new system. In-house staff currently do much of the work and this is often not calculated into the cost of projects. CIOs noted the difficulties justifying ROI for data quality work. In a PHO, ROI was measured on improved patient care and this would need to be demonstrated to obtain funding. A DHB CIO recognised that the DHB was running at a significant deficit, which needed to be reduced, and therefore could not afford expensive rework to fix data. This argument was used to encourage the executive team to support and fund data quality projects.

7.2.9 Management Support

Data quality practitioners were found to elicit the support of managers where possible to achieve their goals. High level management support, such as the executive team, proved effective in solving issues with department managers where the problems were persistent and the department managers resisted change. Sending staff for training was highlighted as one of the issues that required enforcement from department managers, as trainers have no authority to enforce training on staff. CIO’s found the support of department managers difficult to obtain:

there is a lack of management understanding and buy in to data quality issues at all levels. Interesting we see this most when managers who refuse to have their staff collect requested data then challenge why it is then not possible for the organisation to produce the related statistics.
Executive managers were thought to often be unaware of data quality issues, as once the data are in a report for management it has been cleansed as much as possible. Executive managers would be informed of major issues where capital funding was required to make improvements but management themselves otherwise did not see it as an issue to be raised at the executive level. Data quality problems are recognised and discussed by managers when information supplied in reports does not meet their needs i.e. ‘this report is out of date’. Issues with management support often centred around their understanding of roles, as one CIO discussed:

the perception is that it’s an IT call so they are probably not aware of some of the issues that we have unless I have to raise it to that level. I think the expectation is that we will just deal with it.

At one DHB, department managers recognised their responsibility towards data quality following a review of the organisation’s data quality from an external consultancy company. The department managers agreed to work closely with the IS department to actively improve data quality in their departments. Data quality practitioners were very aware of the need for management support and actively sought buy-in from executive management. Several reported significant support for data quality initiatives, however funding for projects or extra staff was not always provided.

7.2.10 The Impact of Data Quality

The key concept to emerge from this category was that the complex and disparate nature of data in the health sector, and within its organisations, impacts on the ability to manage data quality. An example provided noted that one warehouse takes feeds from 21 different data collections, and the custodian of the data warehouse has no control or knowledge of the data quality practices at the supplier end. Suppliers can be small entities, such as an individual pharmacy or GP practice, but the data must be integrated to obtain funding from HealthPAC. Cleansing and interpreting such data is difficult. The data supply chain is long and the aggregate report produced from the source information could look different from the report produced by the data warehouse. The differing views of various stakeholders mean the understanding gained from the data is not always consistent. These different stakeholders may also have different understandings of the importance of the data collection process, e.g. data entry staff who did not have an overview of the organisation were noted to have little understanding of the subsequent uses of the data. Difficulties in interpretation were found at the data field level, with different locations interpreting field definitions differently. Conflicting reports on the same data are also found due to the ‘silo’ nature of data quality work in some organisations. This has caused issues where funding for services is involved, with both parties arguing their data are correct.

Data quality problems were often identified only when the data were used to provide information in reports. Those analysing the data for information are often the first to identify problems that mean that the data are not ‘fit for use’ for their analytical purpose. By this stage data could have passed through several different information systems. This highlights the importance of managing the final data product where the data are used in context. Trust was lost throughout the organisation when data in reports appeared to be inaccurate meaning reports were not utilised. Late data are noted as a problem as staff have little understanding of the
impact of not supplying the data on time, again impacting on the reporting requirements and management of
the services.

Poor data quality impacts on Ministry of Health reporting requirements. A respondent noted a new system
that would not meet reporting requirements in its current format would impact on the organisation’s
performance measures required by the Ministry of Health:

I’m aware that we’re not capturing discharge information that we really need for benchmarking
purposes because people are not being admitted and don’t do the same level of diagnostic coding in
an ED. The data is changing and it will impact on our average length of stay.

Poor data quality, differing data definitions, business rules that are not appropriate to the situation, and lack
of access to data has meant that some DHBs have implemented duplicate data collections alongside those
used for reporting to the Ministry of Health national collections. This imposes a considerable extra cost on
DHBs. In primary care, data in the national collections were compared to the data held in one PHO and the
data did not reconcile. The national data were not trusted and local data were used. Some collections used
for service delivery decisions were found to have data that did not match the source data, indicating the data
collection processes and storage changed the meaning of the data.

Several respondents noted the clinical impact of poor data quality. This was particularly noted where data
move around the health sector, such as in an electronic health record and where patient safety could be
compromised due to inappropriate clinical decisions.

Not having data or having incomplete data available or accessible meant that decision making was affected.
One respondent stated, ‘we’re sitting there blind, trying to run a service’. The devolution of services
previously provided by the Ministry of Health out to the DHBs was not supported with access to historical
information on service provision. Equally difficult was making decisions using out of date data, where no new
data were available. It would be expected that access to data held in the national data collections would be
shared with all stakeholders. At present, users of health data in New Zealand have voiced concerns around
the access to health information held at NZHIS. Whilst NZHIS is constrained by the legislation protecting the
privacy of health data, particularly identifiable clinical data, there appears to be unreasonable constraints on
the access to non-identifiable data.

7.2.11 Data Collection

The data collection category was divided into the sub categories: information systems and processes.
Information systems design is utilised to assist in the collection of good data. In one case, where data were
missing from previous visits, data collectors were prompted by the system to request the information from the
patient. Systems were also noted to need improvement, where fields allowed for the input of inappropriate
data, or where users were in some way ‘still managing to put incorrect data in despite being limited to drop
down lists’. PHOs noted that many GPs use the same practice management system; therefore issues related
to the systems were fed back to the software vendor through the PHO, although there was sometimes limited support provided by the vendor.

Respondents reported that some Ministry of Health information systems feeding into the national collections actively prevent the collection of good data through inaccurate business rules. One example given was the Maternity system held at HealthPAC. This system doesn’t allow for a mother having multiple births, which is a common occurrence.

The ‘data collection processes’ code was further subdivided into the child categories of knowledge (of how the data were collected), training, understanding, and incentives. The disparate nature of data collection and the movement of data around the sector meant many custodians had no knowledge of how data were collected and the processes the data have been through before reaching them. This means any assessment of data quality is not possible and data quality improvement difficult, as custodians have no control over the data collection processes.

Considerable training on data collection is provided to ward and admission clerks. This was evident even where no other preventative data quality work was done. Trainers noted there was little understanding by data collectors on the impact of data collection processes on the ‘big picture’ as many had never been told what the data would be used for. Incentives for data collectors to collect better data included auditing performance and feeding back the number and nature of errors. At present minimum levels of quality expected are not found in position contracts, but this was soon to be introduced in one DHB.

GP\s are paid for data by their PHOs, providing incentive to collect and code more clinical data. Providing GPs with reports from their data on such issues as population disease rates also acted as an incentive, as well as providing service utilisation rates to enable better management of the General Practice.

7.2.12 Creating Change

Under the category, ‘creating change’ the sub categories highlighted what measures data quality practitioners undertook to improve collection practices. These included auditing of performance, feedback loops and paying for data. Action was prompted by poor data quality, including service managers realising they had a role to play in improving the data from their areas. Due to the all-pervading data quality issues that were found, respondents noted that projects for new data collections have now begun to consider data quality from the outset, with more consideration being given to information delivery requirements and process mapping. Organisation wide data quality teams have been formed in three DHBs, and there is increased training of data collection staff and education of managers. In some cases feedback loops have been established where persistent problems have been found. However, much of the work is ad hoc and undertaken in individual departments.

Once data collectors were aware their practices were adversely affecting data quality, through auditing and feedback, improvement was noted. Feedback loops are particularly effective where comparisons with other data collectors or organisations’ performance are provided. Both Fine et al (Fine et al., 2003) and De
Lusignan et al (De Lusignan et al., 2002) found feedback proved to be a highly effective mechanism to improve data quality in medical databases. Respondents also considered data collectors ‘attitudes’ important:

*its about changing the data collectors attitudes so they take ownership, even the managers don’t take ownership.*

Data quality practitioners noted some resistance to changes in practice, with the perception that GPs were particularly unwilling to change. The generally slow rate of change in the health sector was also noted; problems with some of the national collections that had been highlighted several years previous were still being discussed and not yet improved. Practitioners noted that data quality improvement that is linked to supporting the business needs would be more likely to succeed, rather than making data quality a compliance requirement from the Ministry of Health.

### 7.2.13 Staff Skills

Many CIOs (70%), in the 2003 CIO Survey, felt their existing staff did not have the skills and/or they did not have the number of staff needed to meet their agreed organisational service levels (Ministry of Health, 2003a). Budget constraints mean that there are considerable expectations on existing staff and difficulties attracting highly skilled staff. Skilled staff for projects is sometimes provided from other DHBs on secondment. Improving data quality is another skill required of the IS team.

Clinicians with domain knowledge were noted to be valuable in finding data quality issues and understanding the context and implications of these issues. People with a wide understanding of reporting requirements, standards and the way systems work were also noted to be valuable. In general, computer literacy among hospital staff was considered low and training on hospital systems not always sufficient to ensure good understanding by staff, due to limited resources available to deliver training. Staff skill, particularly of data collectors, was noted to be one of the most significant barriers to improving data quality:

*the biggest issue is the quality of the staff rather than the quality of the data*

Training was the most often used method of skill improvement, sometimes targeted to specific people or problems found through auditing. Targeted recruitment of people with the required skills and competency models for performance assessment was beginning to be implemented in one DHB. However, data collectors are often in clerical positions that are low paid, therefore it is difficult to attract highly skilled staff. Staff shortages are noted to be a problem in the current environment of very low unemployment in New Zealand.

NZHIS were considered by study respondents to require leadership skills together with extensive data quality knowledge, making it easier for the sector to implement improvements. Further, considerable analytical skills with domain knowledge are required by NZHIS staff to solve complex problems at a national level.
7.2.14 Data Quality Programmes

This category was divided into several sub categories – training, prevention, find and fix, assessment, feedback loops, audit, edit checks, reports. This provided an overview of the maturity of data quality improvement programmes in the DHBs. No DHB reported having a data quality strategy or comprehensive improvement programme. One DHB is developing a data quality process model through an improvement project in one area that would subsequently be used across the organisation for data quality projects. In this organisation data quality is now included as a component of all IS projects and involves staff from across the organisation, including clinical staff. A minimum data set with standard definitions is being developed; with an assessment of existing data collections to ensure the same data are not being collected twice.

All DHBs stated they provided some form of training to data collectors. For one this was limited to a one-day orientation for new staff. The most comprehensive training programme had recognised the importance of training all levels of the organisation and had spent time educating the executive management team and department managers. The support and understanding of department managers is paramount to making improvements within departments, but this group was often not included in training programmes.

Some difficulties were found with training staff who were already busy, and some training was found to not affect habits in the long term. English (English, 1998) notes the difficulty in erasing already entrenched bad habits to change behaviour, and advocates staff training that is early and continuous. The Ethnicity Data Protocols, developed by the Ministry of Health, provided a training manual for trainers to teach data collectors on the collection of ethnicity data and these protocols were being utilised by trainers. One DHB had developed a manual for all data collection processes. Several DHBs noted that more training was required, but that increased resources were not provided when requested from management.

Preventative measures within data quality programmes mainly included the design of information systems to ensure only correctly formatted data, within an acceptable range, could be entered. Many organisations relied heavily on auditing as a tool for identifying data quality problems. Some organisations reported running regular automated audits of data every day, but most ran weekly audits. On one occasion it was noted that ‘audits are done when I have time’. The audits find ‘irregularities’ in the data, such as in an incorrect format and changes are made to unit level data. The root cause analysis of persistent data quality problems was undertaken only by a few organisations and was a recently implemented initiative where it was found. Audits provide information on the types of data quality problems so that targeted education and improvements can be made. Auditing also provides a trace of staff who entered the data, further targeting education at the individual staff level. Respondents from smaller DHBs, where staff numbers are more limited, found the auditing undertaken by NZHIS staff in the past had provided them with good data quality information, but this no longer takes place.

Considerable work still goes into ‘finding and fixing’ problems in the data. A small DHB had found that data quality was good enough to find problems easily through manual searching. Larger DHBs have begun to assess the persistent problems and target improvements, mainly through education. Much of the reconciliation of duplicate NHI numbers was still done manually due to the difficult decisions required to ensure duplicate NHIs were correctly merged. Manual work concentrates on the quality of the information
systems so that data quality is not reduced by the system. DHBs are prompted by NZHIS and HealthPAC, through stringent edit and validation checks, to improve data supplied to the national collections, particularly inpatient discharge data.

Organisations find formal assessment of data quality difficult, with limited understanding of what to measure. Assessment focused on a limited view of data quality, mainly through auditing. None of the organisations had an assessment methodology that included several dimensions of data quality. Most assessments were to ascertain levels of completeness and formatting accuracy, those features of data that could be assessed by automated auditing of the data or through visual checks. Benchmarking against other collections holding the same data was sometimes used, particularly by the Shared Services Agencies and PHOs, both having access to aggregate regional and national data. One organisation reported valuing anecdotal feedback from data users; however this was provided on an informal, ad hoc basis. Peer review is also used to check that reports contain at least what appear to be good data.

Feedback loops were often utilised and noted to be very important. Some remain informal but in general there are processes in place to report errors back to data suppliers within the organisation and from NZHIS back to organisations. Feedback can be used to reinforce the view that the responsibility for data quality is with the data supplier. The feedback loop from NZHIS, although a formal process, is a one way feedback. Data suppliers are sometimes not followed up after being sent a feedback report, so NZHIS has no knowledge of any action taken on the report. For example, information on who is receiving the report, if it is being used to improve the quality of the collection and whether these changes are provided back to the national collection. Respondents in several DHBs noted that the people who were receiving the NZHIS reports had left the organisation and new or existing staff did not pick up the task of acting on the reports. One data quality team did not have access to these reports and did not know where they went in the organisation. They have directly requested feedback from NZHIS.

7.2.15 Key Concepts to Emerge from the Current State Analysis

The interviews with sector stakeholders provide an outline of data quality practices in the New Zealand health care sector. These key concepts are used to provide insight into the data quality capability maturity of New Zealand health care organisations. The many disparate sources of data in health care lead to considerable complexity in their management. Currently, there is a reactive nature to data quality management, contributing to a lack of trust in data, as users are often made aware of data quality problems for the first time when they try to use data.

There is almost no formal accountability for data quality at management level, some data collectors and department managers have defined responsibilities in their job descriptions, however this is rare. There are many roles within organisations that undertake data quality initiatives and this has led to some disparate silos of work. This issue is being addressed by the formation of organisation-wide data quality teams in at least three DHBs. The data quality programmes are limited to minimal formal assessment of completeness and formatting of data through automated audits and looking at the data. Defining data quality requirements takes place informally where reports are found not to meet the needs of the users.
While the concepts that emerge may appear disparate and somewhat unrelated, they do indicate that data quality work pervades many functions of the organisation and impacts across the organisation. The results highlight the need for an awareness of data quality that thus far has not been consistent within the New Zealand health sector. There is a need for a sector wide management programme that provides clarity on a level of data quality that data consumers and managers can expect of data held in the national collections. These issues are discussed in detail in Chapter 7.

7.3 Capability Maturity Assessment of New Zealand Health Care Organisations

English (English, 1999a) has adapted Philip Crosby’s (Crosby, 1980) quality management maturity grid for assessing the capability of an organisation to make change in the management of data quality. The researcher applied the maturity grid to assess the maturity of the health sector as a whole, using the information provided from the Current State Analysis. The variation of capability between health care organisations is considerable; therefore the researcher made the assessment on two representative DHBs – one with little apparent maturity and one with considerable apparent maturity in data quality management to provide a range of capability maturity for the sector. The measurement categories and stage of maturity of the organisations studied are defined in Table 38 below.

<table>
<thead>
<tr>
<th>Measurement Category</th>
<th>Stage – Immature Organisations</th>
<th>Stage – Mature Organisations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Management understanding and attitude</td>
<td>Stage one – no comprehension of data quality as a management tool. Tend to blame data admin or IS dept for data quality problems</td>
<td>Stage three – while going through data quality improvement programme learn more about quality management; becoming supportive and helpful</td>
</tr>
<tr>
<td>Data quality organisation status</td>
<td>Stage one – data quality hidden in application development dept. Data audits not routine. Emphasis on correcting bad data</td>
<td>Stage two – data quality role is appointed but main emphasis still on correcting bad data</td>
</tr>
<tr>
<td>Data quality problem handling</td>
<td>Stage one – problems are fought as they occur; no resolution; inadequate definition</td>
<td>Stage two – teams are set up to attack major problems. Long range solutions are not solicited</td>
</tr>
<tr>
<td>Cost of data quality as a percent of operating costs</td>
<td>Stage one – Reported: unknown Actual: 20%</td>
<td>Stage one – no organisation has measured the actual cost of data quality and there is no methodology developed</td>
</tr>
<tr>
<td>Data quality improvement actions</td>
<td>Stage two – trying obvious motivational short range efforts</td>
<td>Stage three – data quality programme in place</td>
</tr>
<tr>
<td>Summation of data quality posture</td>
<td>Stage two – “is it absolutely necessary to always have problems with data quality?”</td>
<td>Stage three – “through management commitment and data quality improvement we are identifying and resolving our problems”</td>
</tr>
</tbody>
</table>

Table 38: The Range of Data Quality Capability of New Zealand Health Care Organisations

The ‘immature’ organisation was generally found to be at stage one, called ‘uncertainty’ by English (English, 1999a) and ‘ad hoc’ by Crosby (Crosby, 1980). This is the least mature stage found in the grid, with the organisation highly reactive to problems as they arise. Little is done to solve long term and/or persistent problems.
The more ‘mature’ organisations were found to be variable in the grid rating scale. In particular, no organisation has formally measured the impact of data quality on the organisation and therefore even the most mature organisations are at stage one on the grid. Stage two is the ‘awakening’ or ‘repeatable’ stage, where, due to some kind of incident, the organisation has become aware that there are data quality problems. Data quality roles are in place and teams are set up to deal with problems. All, including immature organisations, are at a minimum fixing obvious data quality problems.

The most mature organisations are at stage three on the capability maturity grid. Stage three is ‘enlightenment’ or ‘defined’, where attitudes to data quality are noticeably different, with supportive and educated management providing resources for improvements. Data quality programmes are in place. Few organisations in this study were found to be at stage three for any category of the capability maturity grid.

Stage four is ‘wisdom’ or ‘managed’, where the organisation sees significant benefits from its data quality initiatives and is continuing to implement and mature its data quality improvement processes. Stage five is ‘certainty’ or ‘optimising’, where there is increased customer satisfaction and virtually complete data defect prevention. All errors are analysed for cause and preventative measures taken. None of the organisations surveyed in this study had any stage four or five practices in place.

### 7.4 Ministry of Health Data Quality Workshop

The information provided by the Current State Analysis, an extensive understanding of the data quality literature, the development of the DQEF, and the Capability Maturity Analysis of New Zealand health sector organisations enabled the researcher to develop an initial draft Data Quality Improvement Strategy (DQIS). This draft strategy was presented to Ministry of Health staff at a one-day workshop held by the researcher on data quality. The purpose of the workshop was to ascertain the participants’ perspectives of the dimensions and characteristics of data quality as defined by the Ministry in the DQEF, to determine what the relative priorities for improvement should be, and how well they felt the business was operating against these dimensions. The workshop was also used to inform participants of the development of the DQIS, as the workshop provided an opportunity to discuss and educate a wider Ministry of Health audience on data quality theory, the DQEF, and the development of the DQIS. For ease of understanding, participants were asked to consider each question in relation to the NHI. Participants were asked to define themselves as one or more of the following:

- data collector – roles where data are being obtained and entered into the NHI;
- data custodian – roles where data are maintained or protected;
- data customer – roles where data are being utilised for information purposes.

Figure 13 below provides a breakdown of the roles of the participants.
Participants were asked to assess the importance of each characteristic within the data quality dimensions included in the DQEF within their defined roles. The DQEF dimensions and characteristics are defined in Appendix 5. A scale of 1.0 (= very important) to 5.0 (= not important) was used. Table 39 provides a summary of the average importance ratings of each characteristic when broken down into roles provided by participants. The assessment of each characteristic, rather than by overall dimension, provided the researcher with a more detailed understanding of the participants’ perception of data quality as each dimension is broken down into several more defined characteristics. As there are several characteristics in each dimension, the researcher wanted to understand what aspects of each were most important.
### Table 39: Importance of Data Quality Characteristics by Role

This resulted in five groupings of priority, all rated above ‘important’. Table 40 shows the five characteristics by dimension that were rated the most important:

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparability</td>
<td>Linkage</td>
</tr>
<tr>
<td></td>
<td>Historical Comparability</td>
</tr>
<tr>
<td>Relevance</td>
<td>Value</td>
</tr>
<tr>
<td>Privacy and Security</td>
<td>Legislation</td>
</tr>
<tr>
<td></td>
<td>Disclosure</td>
</tr>
</tbody>
</table>

### Table 40: Characteristics by Dimension Rated by Respondents as a Priority for Improvement on the NHI

All characteristics were grouped into an importance rating between one and five, with one being the most important for improvement. The results for current performance are the average of the responses provided to the question, ‘how well do we currently manage this characteristic’? The scale for responses to the question...
was 1.0 (= extremely poorly) to 5 (= extremely well). The difference between the two ratings provides an indication of how much of a difference respondents reported between expected performance within a dimension’s characteristic and the perceived actual current performance, potentially indicating areas where the most benefit can be made from focused attention through strategic or business as usual initiatives. This information is outlined in Table 41 below.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Importance Grouping</th>
<th>Importance Average</th>
<th>Current Performance Average</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linkage</td>
<td>1</td>
<td>4.9</td>
<td>3.4</td>
<td>1.5</td>
</tr>
<tr>
<td>Historical Comparability</td>
<td>1</td>
<td>4.9</td>
<td>2.9</td>
<td>2.0</td>
</tr>
<tr>
<td>Value</td>
<td>1</td>
<td>4.9</td>
<td>2.3</td>
<td>2.6</td>
</tr>
<tr>
<td>Legislation</td>
<td>1</td>
<td>4.9</td>
<td>3.6</td>
<td>1.3</td>
</tr>
<tr>
<td>Disclosure</td>
<td>1</td>
<td>4.9</td>
<td>3.1</td>
<td>1.8</td>
</tr>
<tr>
<td>Capture and Collection</td>
<td>2</td>
<td>4.8</td>
<td>2.5</td>
<td>2.3</td>
</tr>
<tr>
<td>Estimation and Processing</td>
<td>2</td>
<td>4.8</td>
<td>2.3</td>
<td>2.5</td>
</tr>
<tr>
<td>Standardisation</td>
<td>2</td>
<td>4.8</td>
<td>3.5</td>
<td>1.3</td>
</tr>
<tr>
<td>Data dictionary</td>
<td>2</td>
<td>4.8</td>
<td>2.9</td>
<td>1.9</td>
</tr>
<tr>
<td>Equivalency</td>
<td>3</td>
<td>4.7</td>
<td>3.1</td>
<td>1.6</td>
</tr>
<tr>
<td>Documentation</td>
<td>3</td>
<td>4.7</td>
<td>2.6</td>
<td>2.1</td>
</tr>
<tr>
<td>Interpretability</td>
<td>3</td>
<td>4.7</td>
<td>2.6</td>
<td>2.1</td>
</tr>
<tr>
<td>Coverage</td>
<td>3</td>
<td>4.6</td>
<td>2.7</td>
<td>1.9</td>
</tr>
<tr>
<td>Data Currency at Time of Release</td>
<td>4</td>
<td>4.5</td>
<td>2.6</td>
<td>1.9</td>
</tr>
<tr>
<td>Accessibility</td>
<td>4</td>
<td>4.5</td>
<td>2.8</td>
<td>1.7</td>
</tr>
<tr>
<td>Missing Records</td>
<td>4</td>
<td>4.4</td>
<td>2.2</td>
<td>2.2</td>
</tr>
<tr>
<td>Edit and Imputation</td>
<td>4</td>
<td>4.4</td>
<td>2.2</td>
<td>2.2</td>
</tr>
<tr>
<td>Documentation Currency</td>
<td>4</td>
<td>4.4</td>
<td>2.5</td>
<td>1.9</td>
</tr>
<tr>
<td>Adaptability</td>
<td>4</td>
<td>4.4</td>
<td>2.1</td>
<td>2.3</td>
</tr>
<tr>
<td>Measurement Error</td>
<td>5</td>
<td>4.3</td>
<td>2.3</td>
<td>2.0</td>
</tr>
<tr>
<td>Item (partial) Missing Values</td>
<td>5</td>
<td>4.2</td>
<td>2.8</td>
<td>1.4</td>
</tr>
</tbody>
</table>

Table 41: Perceived Current Performance of Data Quality Work for Each Characteristic of the DQEF

Participant perception was that many dimension characteristics are not managed well within the Ministry of Health. As a result there are a high number of areas that indicate reasonable gains can be made on current practice. Legislation compliance and standardisation are considered to be the best managed characteristics, each with a difference score of 1.3. The following were considered by the group to be managed poorly in relation to the NHI as they have a difference score of greater than 2.0:

- value;
- estimation and processing;
• missing records;
• edit and imputation;
• adaptability.

The purpose of the next segment of the workshop was to discuss the DQIS and establish a consensus as to which initiatives should take priority for improvement. The strategy proposed a programme of work to introduce regular ‘business as usual’ data quality initiatives to institutionalise the ethos of the DQEF dimensions into the Ministry of Health and the sector. Participants were asked if they considered that the proposed initiatives would bring about effective change by targeting the right aspects of data quality and if they felt any initiatives should be added or taken away. Participants were asked three questions relating to each proposed initiative included in the DQIS and asked to rate them using a 1.0 – 5.0 scale:

1. How much impact will this initiative have on the quality of the NHI? (1.0 = Trivial, 5.0 = significant)
2. What priority should this initiative have? (1.0 = Low, 5.0 = very high)
3. How soon should this occur? (1.0 = within two years, 3.0 = within 6 months, 5.0 = immediately)

Totalling the average results of all three questions, with the highest figure receiving the highest ranking provides composite rankings. The responses provided by participants are found in Table 42.
<table>
<thead>
<tr>
<th>Initiatives</th>
<th>Quality Impact</th>
<th>Priority</th>
<th>Timing</th>
<th>Totals</th>
<th>Composite Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Development of a comprehensive education programme for MOH and sector</td>
<td>4.6</td>
<td>4.7</td>
<td>4.3</td>
<td>13.6</td>
<td>1</td>
</tr>
<tr>
<td>Standard operational initiatives</td>
<td>4.4</td>
<td>4.4</td>
<td>4.1</td>
<td>12.9</td>
<td>2</td>
</tr>
<tr>
<td>Development of best practice guidelines for data collection processes</td>
<td>4.2</td>
<td>3.9</td>
<td>3.4</td>
<td>11.5</td>
<td>3</td>
</tr>
<tr>
<td>Development of KPIs for DHB contracts</td>
<td>4.2</td>
<td>3.7</td>
<td>3.3</td>
<td>11.2</td>
<td>4</td>
</tr>
<tr>
<td>Micro view assessment of data elements for NHI</td>
<td>4.2</td>
<td>3.9</td>
<td>2.9</td>
<td>11.0</td>
<td>5=</td>
</tr>
<tr>
<td>Review DQ team roles (new position for trainer)</td>
<td>3.9</td>
<td>3.6</td>
<td>3.5</td>
<td>11.0</td>
<td>5=</td>
</tr>
<tr>
<td>Database interconnection (sharing data between systems rather than duplicating)</td>
<td>4.3</td>
<td>4.1</td>
<td>2.7</td>
<td>11.0</td>
<td>5=</td>
</tr>
<tr>
<td>Statistical process control</td>
<td>3.9</td>
<td>3.6</td>
<td>3.3</td>
<td>10.8</td>
<td>6</td>
</tr>
<tr>
<td>Review business rules and data dictionary across the Ministry</td>
<td>3.6</td>
<td>3.4</td>
<td>3.0</td>
<td>10.0</td>
<td>7</td>
</tr>
<tr>
<td>Metadata improvement</td>
<td>3.5</td>
<td>3.2</td>
<td>2.6</td>
<td>9.6</td>
<td>8</td>
</tr>
<tr>
<td>Data quality framework implementation</td>
<td>3.2</td>
<td>3.4</td>
<td>2.7</td>
<td>9.3</td>
<td>9</td>
</tr>
<tr>
<td>DHB/PHO accreditation scheme</td>
<td>3.6</td>
<td>3.1</td>
<td>2.4</td>
<td>9.1</td>
<td>10</td>
</tr>
<tr>
<td>Roles and responsibilities for NHI throughout the sector - increased sector participation</td>
<td>3.4</td>
<td>2.9</td>
<td>2.2</td>
<td>8.5</td>
<td>11</td>
</tr>
<tr>
<td>New collections - minimum requirements</td>
<td>3.1</td>
<td>2.9</td>
<td>2.3</td>
<td>8.3</td>
<td>12</td>
</tr>
<tr>
<td>Data production maps</td>
<td>3.3</td>
<td>2.7</td>
<td>2.1</td>
<td>8.1</td>
<td>13</td>
</tr>
<tr>
<td>Prioritise customers and their needs</td>
<td>3.0</td>
<td>2.5</td>
<td>2.3</td>
<td>7.8</td>
<td>14</td>
</tr>
<tr>
<td>Data governance</td>
<td>3.1</td>
<td>2.8</td>
<td>1.8</td>
<td>7.7</td>
<td>15</td>
</tr>
</tbody>
</table>

Table 42: Impact, Priority and Timing of Proposed Data Quality Improvement Initiatives

Of the initiatives discussed, the top four in terms of a composite ranking relating to importance, timing and priority were:

- development of a comprehensive education programme for MOH and sector;
- standard operational initiatives;
- development of best practice guidelines for data collection processes;
- development of KPIs for DHB contracts.

These top four projects had a good correlation to the four streams of the breakout session (see section 6.4.1) relating to the principles of the DQIS. Ultimate priorities may vary, due to other interdependencies at the Ministry of Health and in the sector that were not considered by the participants. The timeframes the group wanted for completion of projects were generally longer than those proposed in planning thus far, with participants preferring to defer many projects and reporting concerns relating to:
• funding for initiatives – participants felt that current sector funding was insufficient and additional significant funding and commitment would be required from the Ministry to ensure initiatives were successful;
• resourcing for initiatives – participants felt that current human resources were inadequate to meet the requirements and intent of the initiatives;
• sector buy-in and enthusiasm for the whole process – participants agreed that there would need to be significant education and cultural change within the sector to ensure the success of the initiatives.

7.4.1 Breakout sessions

The workshop participants separated into smaller groups to answer a series of questions related to the implementation of the strategy. The groups formed under the following streams for this section:

• culture change;
• incentives for change;
• performance measurement and monitoring.

The researcher chose the above topics in order to ascertain the work that would be required to bring about strategic change in the sector. Groups were asked to discuss and report back on each of these areas with a range of questions provided to guide the discussion. The questions were designed to elicit potential barriers to the implementation of the strategy, by assimilating the knowledge of workshop participants, who, by the nature of the organisation, have had considerable experience in strategy and policy implementation. Table 43 summarises the responses from each group.
## Table 43: Feedback from Breakout Groups on Strategy Implementation Requirements

<table>
<thead>
<tr>
<th>Question</th>
<th>Culture Change</th>
<th>Incentives for change</th>
<th>Performance Measurement and Monitoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>What are the data quality improvement initiatives that relate to this area?</td>
<td>Education&lt;br&gt;Best practice definition&lt;br&gt;KPI definition</td>
<td>Definition of best practice guidelines – 'going to the sector and learning from the best'&lt;br&gt;Development of KPIs&lt;br&gt;Education</td>
<td>KPI definition&lt;br&gt;Education&lt;br&gt;Data Quality Framework&lt;br&gt;Practice guidelines&lt;br&gt;Customer Satisfaction survey&lt;br&gt;Statistical process control&lt;br&gt;Accreditation scheme</td>
</tr>
<tr>
<td>What are the priorities for improvement?</td>
<td>A trainer was felt to be essential. This trainer’s role to be to spread the word through training the trainer initiatives providing educators/champions in individual units</td>
<td>Set the standards and behaviours that should be communicated and socialise them to the broader group of users.</td>
<td>In particular the following questions need to be answered to make this achievable:&lt;br&gt;What are the measurements?&lt;br&gt;How will they be assessed?&lt;br&gt;What is acceptable quality?</td>
</tr>
<tr>
<td>What resources would be required?</td>
<td>Costs were not seen to be a big issue in terms of additional expenses. However, some cost would be incurred disseminating the message to the sector through training, road shows ads, progress monitoring, etc.</td>
<td>The group felt that the initiative will cost a lot and up front commitment is essential. It was suggested five long term positions would be required to develop and operationalise the programme.</td>
<td>Analysts to establish parameters/measure performance&lt;br&gt;Trainers to communicate expectations and changes to processes; and&lt;br&gt;Programmers to create the tools.</td>
</tr>
<tr>
<td>Who should own/take the lead in this area?</td>
<td>Requires strong leadership from the Ministry who should act as role models for all other stakeholders helping to drive initiatives out into the sector</td>
<td>Responsibility to operationalise the initiatives clearly fell upon the Clinical Analysis team at NZHIS however the overall ownership lay with the Ministry</td>
<td>The Ministry was seen to be the appropriate leader and owner. The Ministry would initiate the process monitoring DHB performance with the DHBs moving to self assessment as the initiatives mature.</td>
</tr>
</tbody>
</table>
Feedback from across the three groups highlighted similar themes on implementation requirements:

- set clear understandable standards;
- establish good communications plan (education, get buy in from stakeholders, etc.);
- have clear ownership and real leadership.

Recommendations from the workshop suggested the draft DQIS be revised with the inclusion of the areas raised in the workshop, these being:

- database interconnection (sharing data between systems rather than duplicating within each system);
- micro (as opposed to macro) view assessment of data elements for NHI;
- roles and responsibilities for NHI throughout the sector (in particular looking at increased sector participation in the NHI);
- data governance (establish clear roles and responsibilities as well as ownership);
- interoperability of sector systems with NHI.

The Ministry should drive the funding and leadership for these initiatives as part of a project. However, as time progresses more of the funding and resourcing onus will have to fall upon the health sector. This shift must occur with due consideration to the long term funding and resourcing requirements.

### 7.5 Stakeholder Analysis

A stakeholder analysis undertaken by the researcher provided insight into the varying information requirements of the health sector, all with their own perspectives and uses of the national health data collections. Table 44 provides an analysis of Ministry of Health stakeholders and their information requirements and Table 45 provides an analysis of health sector stakeholders and their information requirements.
### 7.5.1 Ministry of Health Stakeholders

<table>
<thead>
<tr>
<th>Stakeholder</th>
<th>Information Interest/Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minister and Associate Ministers of Health</td>
<td>Accurate information and advice from Ministry officials</td>
</tr>
<tr>
<td>Ministry of Health Directorates</td>
<td>Decision-making</td>
</tr>
<tr>
<td>Management Teams</td>
<td>Policy development and analysis</td>
</tr>
<tr>
<td></td>
<td>Assessment of DHB performance</td>
</tr>
<tr>
<td>Policy Analysts</td>
<td>Policy development and analysis</td>
</tr>
<tr>
<td>Policy Advisors</td>
<td>Advice to the Minister, Ministry and sector</td>
</tr>
<tr>
<td>NZHIS</td>
<td>Assessment of health outcomes from national collections</td>
</tr>
<tr>
<td>Clinical Analysis Team</td>
<td>Reporting requirements to WHO</td>
</tr>
<tr>
<td>Data Quality Team (a subgroup within the Clinical Analysis Team)</td>
<td>Yearly publications of health indicators, including mortality analysis</td>
</tr>
<tr>
<td></td>
<td>Information to the Minister, researchers, DHBs, the Ministry directorates</td>
</tr>
<tr>
<td></td>
<td>Data quality caveats to understand limitations of data</td>
</tr>
<tr>
<td>Business Intelligence Unit</td>
<td>Information from data suppliers on collection processes</td>
</tr>
<tr>
<td>Clinical Coding Team</td>
<td>Ensure business processes support data suppliers to provide good data</td>
</tr>
<tr>
<td></td>
<td>Metadata of all national collections</td>
</tr>
<tr>
<td></td>
<td>Data standards and data dictionaries</td>
</tr>
<tr>
<td></td>
<td>Data quality requirements of customers</td>
</tr>
<tr>
<td></td>
<td>Data quality caveats to customers</td>
</tr>
<tr>
<td>HealthPac</td>
<td>Persistent issues with the supply of claims data</td>
</tr>
<tr>
<td>Business Analysts</td>
<td>Provision of data to NZHIS</td>
</tr>
<tr>
<td></td>
<td>Sector used standards, data dictionaries and metadata</td>
</tr>
<tr>
<td></td>
<td>Fraudulent claims</td>
</tr>
<tr>
<td></td>
<td>Accidental over/underpayment of claims</td>
</tr>
<tr>
<td>National Screening Unit</td>
<td>Information from data suppliers about collection processes</td>
</tr>
<tr>
<td>Data Quality Manager</td>
<td>Data standards</td>
</tr>
<tr>
<td></td>
<td>Data quality requirements of customers</td>
</tr>
<tr>
<td></td>
<td>Data quality caveats to customers</td>
</tr>
<tr>
<td></td>
<td>Ensure business processes support data suppliers to provide good data</td>
</tr>
<tr>
<td>Health Information Standards Organisation</td>
<td>Current standards in use</td>
</tr>
<tr>
<td></td>
<td>Priorities for standards endorsement to support data quality</td>
</tr>
<tr>
<td></td>
<td>Provide information to standards working groups</td>
</tr>
<tr>
<td></td>
<td>Dissemination and support of endorsed standards</td>
</tr>
</tbody>
</table>

Table 44: Ministry of Health DQIS Stakeholders and their Information Requirements
### 7.5.2 Sector Stakeholders

<table>
<thead>
<tr>
<th>Stakeholder</th>
<th>Information Interest/Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinicians</td>
<td>Access to clinical data to support decisions&lt;br&gt;Quality caveats to understand the limitations of the data&lt;br&gt;Education as data collectors</td>
</tr>
<tr>
<td>DHBs</td>
<td></td>
</tr>
<tr>
<td>CEOs</td>
<td>Strategic decision-making on funding and planning requirements</td>
</tr>
<tr>
<td>CIOs</td>
<td>Management of data supplied from business units&lt;br&gt;Management of information systems (new and existing) and clinical and administration data&lt;br&gt;Support for business units that collect and store data&lt;br&gt;Strategic analysis of future information requirements</td>
</tr>
<tr>
<td>Quality Managers</td>
<td>Assessment of all aspects of quality across the organisation requiring accurate clinical and administrative data&lt;br&gt;Provision of quality improvement initiatives&lt;br&gt;Meeting hospital accreditation requirements</td>
</tr>
<tr>
<td>Business Units</td>
<td>Collection of data for internal and external reporting requirements</td>
</tr>
<tr>
<td>Information Analysts</td>
<td>Assessment of clinical and administrative data for health outcomes and service utilisation</td>
</tr>
<tr>
<td>Clinical Coders</td>
<td>Contextual information for effective and specific coding of primary and secondary diagnoses&lt;br&gt;Current international data standards</td>
</tr>
<tr>
<td>Shared Services</td>
<td>Funding and performance measurement for regions&lt;br&gt;Health outcomes analysis for regions&lt;br&gt;Strategic directions for service requirements</td>
</tr>
<tr>
<td>Agencies Information Analysts</td>
<td></td>
</tr>
<tr>
<td>PHOs/IPAs</td>
<td>Funding and performance measurements for catchment area&lt;br&gt;Health outcomes analysis for catchment area&lt;br&gt;Reporting requirements&lt;br&gt;Claims data requirements</td>
</tr>
<tr>
<td>CEOs/Managers</td>
<td>Administrative data&lt;br&gt;Clinical data management – possibly used for decision support&lt;br&gt;Claims data requirements&lt;br&gt;Reporting data requirements</td>
</tr>
<tr>
<td>Practice Managers</td>
<td></td>
</tr>
<tr>
<td>Privacy Commissioner</td>
<td>Appropriate management of the privacy of health data throughout the sector including access restrictions</td>
</tr>
<tr>
<td>Health and Disability Commissioner</td>
<td>Accurate and complete documentation of clinical practice&lt;br&gt;Tracking of alterations to data (non-repudiation)&lt;br&gt;Protection of the patient’s information for appropriate users only</td>
</tr>
<tr>
<td>ACC</td>
<td>Ability to merge data to provide health outcomes data</td>
</tr>
<tr>
<td>State Services</td>
<td>Alignment with e-government standards and guidelines</td>
</tr>
<tr>
<td>Commission</td>
<td></td>
</tr>
<tr>
<td>Statistics New Zealand</td>
<td>Alignment with standards used in demographic data to enable benchmarking</td>
</tr>
<tr>
<td>Researchers</td>
<td>Data quality caveats to understand the limitations of the data&lt;br&gt;Contextual information on data collection metadata</td>
</tr>
</tbody>
</table>

Table 45: Health Sector DQIS Stakeholders and their Information Requirements
7.6 Data Quality Improvement Strategy

The researcher undertook the development of a data quality improvement strategy in response to the information provided by the research thus far. The research highlights the need for improved data quality management across the whole of the health sector, given the movement of data around the sector and the absence of clearly defined roles and responsibilities. The aim is to provide high level strategic direction to the health sector to ensure consistency and minimum standards in data quality management.

Change management is the main driver of the DQIS that, together with the data quality dimensions and DQEF, comprise the overall Data Quality Improvement Programme at the New Zealand Ministry of Health. The Ministry of Health Data Quality Workshop (see Section 6.4) informed the subsequent further development of the draft Data Quality Improvement Strategy (DQIS). The DQIS builds on the learning and development of the DQEF and includes the dimensions found in the framework, with the sector survey confirming the dimensions are applicable and comprehensive. Strategy development considered a wide range of areas, reflecting the impact of data quality across the entire health sector. Areas of consideration include:

- cost/benefit and return on investment implications;
- risks of omission and commission (the risk of doing nothing and the risk involved in strategic options);
- long-term goals for data quality in the New Zealand health care environment;
- the current and future use of data and information, including health care strategy requirements;
- the outcome of the pilot study of the framework – providing a gap analysis to identify priorities for improvement initiatives, where consistent, high-level problems appear across many data collections;
- the data governance model developed for the health and disability sector (Ministry of Health, 2004);
- the Information Systems Strategic Plans for the Ministry of Health and the DHBs (Ministry of Health, 2005);
- current ‘good data quality practices’ utilised at the Ministry of Health and by the sector;
- feasibility of proposed strategic change in the New Zealand health care environment, given the capability maturity of health care organisations and conflicting priorities.

Several stakeholders, including DHB organisation wide data quality teams and CIOs, had input into the development of the strategy. Ongoing communication with health sector groups, including the Ministry of Health staff, was essential to ensure sector buy-in and to maintain input and interest in the strategy development and implementation. Full consultation with a wide range of data suppliers and users was also necessary. The resulting Data Quality Improvement Strategy is discussed below. The Data Quality Improvement Strategy Consultation Document can be found at Appendix 8.
The Data Quality Improvement Strategy has the following vision.

*High quality health information will be accessible to all appropriate users through the application of total data quality management principles across the health sector. Data suppliers and custodians will consistently assess data quality as defined by data users, with continual improvements made in accordance with the assessment results.*

Whilst a data quality framework models the data environment and identifies quality characteristics, it must be underpinned and supported by a data quality programme that is broader in scope. Current information processes must be carefully documented to make the meaning of data transparent to all users. For example, data sources should always be identified to provide the data user with context around the data collection process. The health sector is currently uncoordinated across health care providers, funders, and departments. Maintaining consistency with data definitions, business rules and even systems architecture can increase the utilisation of valuable data across the health sector and assist in data quality improvement. Standardisation and consistency allow for the integration of data, further increasing its usability. This is where a ‘whole of health sector’ strategy begins to have a significant effect on the quality of data and where centralised leadership is required to ensure the needs of those who manage and use data are considered.

The data quality standards established by the DQIS aim to achieve improved health outcomes by:

- better decision-making to target areas of improvement in national health data collections, with a transparent prioritisation decision tool for data quality improvement projects;
- improved relationships with data suppliers, developing a whole-of-sector responsibility for data quality;
- improved awareness of a data quality culture throughout the sector;
- improved understanding of the processes involved in developing an information product;
- minimum data quality requirements for existing and new national health data collections;
- minimum requirements for regular operational data quality initiatives across all national health collections, with TDQM principles applied to provide feedback and support to data suppliers on persistent data quality errors;
- best practice guidelines leading to accreditation of data suppliers;
- education, support, and leadership to data suppliers, provided by the Ministry of Health.

### 7.6.1 The Scope of the Strategy

The DQIS includes the development of processes to prevent systemic data quality errors in all existing and proposed new national health collections managed by the Ministry of Health and includes processes along the continuum of data collection from source through to the publication of information products.
Therefore, as suppliers of the data, health care providers and business units within the Ministry are included. The strategy identifies the importance of educating data suppliers about the Ministry’s expectations on collection processes as well as clearly defining the roles and responsibilities across the health sector. The strategy does not include the cost of improving data quality at the supplier end. Data suppliers are expected to manage their own data quality initiatives, aligning them with the DQIS. However, the Ministry of Health will provide support, leadership and guidance. The scope of the strategy is illustrated in Figure 13.

Collections included in the DQIS scope are listed in Table 46 below. The range of collections included in the scope of the strategy highlight the need for a data quality framework and improvement strategy that is sufficiently generic to be applicable to many different types of data that are used in a wide range of contexts.
<table>
<thead>
<tr>
<th>HealthPAC Dunedin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Care Manager Database</td>
</tr>
<tr>
<td>Client Claims &amp; Payment System</td>
</tr>
<tr>
<td>Contract Management System</td>
</tr>
<tr>
<td>Output Collection Programme</td>
</tr>
<tr>
<td>PERORG</td>
</tr>
<tr>
<td><strong>HealthPAC Wellington</strong></td>
</tr>
<tr>
<td>Capitation Base Funding/Capitation Information Cleansing</td>
</tr>
<tr>
<td>Dental System</td>
</tr>
<tr>
<td>Maternity Data Warehouse</td>
</tr>
<tr>
<td>Operational Reporting Database</td>
</tr>
<tr>
<td>PROCLAIM</td>
</tr>
<tr>
<td>Oracle</td>
</tr>
<tr>
<td><strong>HealthPAC Wanganui</strong></td>
</tr>
<tr>
<td>Special Authorities</td>
</tr>
<tr>
<td>WINCAPP</td>
</tr>
<tr>
<td><strong>NZHIS</strong></td>
</tr>
<tr>
<td>General Medical Services Data Warehouse</td>
</tr>
<tr>
<td>Health Practitioner Index or National Provider Index</td>
</tr>
<tr>
<td>Hepatitis B Screening Register</td>
</tr>
<tr>
<td>Immunisation Data Warehouse</td>
</tr>
<tr>
<td>Laboratory Claims Data Warehouse</td>
</tr>
<tr>
<td>Maternity and New-born Information System</td>
</tr>
<tr>
<td>Mental Health Information National Collection</td>
</tr>
<tr>
<td>Medical Warnings System</td>
</tr>
<tr>
<td>Mortality Collection</td>
</tr>
<tr>
<td>National Booking Reporting System</td>
</tr>
<tr>
<td>National Booking Reporting System Data Warehouse</td>
</tr>
<tr>
<td>National Cancer Registry</td>
</tr>
<tr>
<td>National Health Index</td>
</tr>
<tr>
<td>National Minimum Dataset</td>
</tr>
<tr>
<td>Pharmaceutical Information Database</td>
</tr>
<tr>
<td>Private Hospital Data Collection</td>
</tr>
<tr>
<td>Workforce Statistics Data Warehouse</td>
</tr>
<tr>
<td><strong>National Screening Unit</strong></td>
</tr>
<tr>
<td>Breast Screening Register</td>
</tr>
<tr>
<td>Cervical Screening Register</td>
</tr>
</tbody>
</table>

Table 46: National Health Data Collections Included in the Scope of Data Quality Improvement Strategy
7.6.2 Aims of the Strategy

Data move with the patients they refer to creating reciprocal dependence between health care organisations. This means poor data management in one organisation can adversely and incrementally affect other organisations and the quality of care a patient receives. A national ‘systems’ framework is therefore needed to certify that data used for decision making meet the same quality criteria and standards both within and between organisations. Further, organisations need to know of the data quality practices in the organisations where the data came from to make judgements on the quality and usefulness of the data.

The DQIS aims to:

• ensure the flow of information through the sector does not change the information’s meaning;
• prevent poor quality data entering the system through appropriate business rules, validation checks, entity checks, etc, thereby reducing the costs resulting from rework due to poor quality data within systems;
• educate and assist data suppliers to send in only good quality data by applying the TDQM principles that see an information product is managed at every point along its transition, from data collection at source to its final amalgamation into an information product;
• educate Ministry of Health staff on TDQM principles and apply those principles in all processes related to the custodianship of national data.

The DQEF will be used on a regular basis to assess the quality of data in each national health collection. The DQEF will also be used to assess the progress of data quality after improvement programmes have been initiated, thus confirming their effectiveness. Initiatives are to be based on a measure of priority derived from the critical impacts of the health collection on the Ministry’s work. The assessment and improvement of business process, rules and data definitions will ensure consistency and accurate data mapping across the Ministry and the health sector. New proposed collections will undergo an assessment of the technical specifications and collection and management processes to ensure good data quality processes are instilled in the development stages.

Improved relationships with data suppliers through feedback loops and education programmes within the Ministry and out in the sector may further improve the quality of the data supplied from the sector (De Lusignan et al., 2002). Data quality will be the responsibility of the whole sector. Responsibilities within each organisation type (that is, the Ministry, DHBs and PHOs) will be clearly assigned to ensure appropriate accountability.

The DQIS seeks to instil Total Data Quality Management (TDQM) throughout the health sector. This is a continuous cycle of proactive management of data to ensure that only high quality health information is available to all. Data are viewed as a product and this concept is fundamental to TDQM. The formulation of the information product can be managed at every point along its transition, much like a manufacturing process, from collection at source to its inclusion in an information product. An information product within the Ministry of Health could be any one of the national health collections, such as the National Minimum Data
Set (NMDS) or the National Cancer Registry. In DHBs, information products could take the form of collections used for reporting purposes (that is, to NZHIS) or collections used for claiming purposes (that is, to HealthPAC).

7.6.3 Roles in Data Quality in the New Zealand Health Sector

As data quality affects everyone in the health sector, the whole sector is responsible for maintaining and improving data quality. There are three defined roles relating to data quality – the data custodian (for example, Ministry of Health, shared services agencies, DHBs, PHOs); the data collector (for example, DHBs, PHOs, GPs, NGOs) and the data consumer (for example, researchers, policy developers and funding and performance managers). The role of the Ministry of Health is one of leadership and support, while data collectors need to employ all possible processes to ensure only high quality data are collected, using agreed national and international standards, where available.

The Ministry of Health owns the accountability framework for health care provision in New Zealand. While the NZHIS has accepted accountability for the lead on data quality, this does not mean it is responsible for solving data quality problems. The NZHIS is not accountable for health care organisations’ data quality but ensures that there is adequate quality when data move between organisations.

As noted previously, identifying data ‘ownership’ is considered paramount in data quality, as this ownership helps to define the roles and responsibilities throughout the data flow (Loshin, 2001). The health sector has a ‘consortium of ownership’ of the national health data collections. Where there are multiple data suppliers, the suppliers are the owners of the data and a key requirement is that data quality is the responsibility of the data supplier (Loshin, 2001). In the health care environment, this is particularly important. Once the data have left the supplier they have moved from the source, the only place where the ‘truth’ can be verified. The accountability for data quality is at the source through the responsibility of data owners. Data quality needs to be the responsibility of high-level managers in an organisation to ensure the entire organisation makes the required changes for improvement. The entities identified in the Stakeholder Analysis define roles in the management of data quality. The strategy outlines clear roles for the health sector and these are summarised in Table 47 below.
<table>
<thead>
<tr>
<th>Entity</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ministry of Health</td>
<td>As the custodian of the data collected throughout the health and disability sector for the national health collections, the Ministry of Health holds the mandate to manage that data. The Ministry is responsible for maintaining the integrity of the data it receives.</td>
</tr>
<tr>
<td>NZHIS</td>
<td>NZHIS is to be the sector and Ministry leader in data quality, providing expert advice and leading through example. NZHIS will:</td>
</tr>
<tr>
<td></td>
<td>• undertake independent research and consult with experts in areas such as data collection, data analysis and data management;</td>
</tr>
<tr>
<td></td>
<td>• maintain ongoing contact with users and participants in conferences and workshops in order to objectively assess the current and future information needs of the sector;</td>
</tr>
<tr>
<td></td>
<td>• review the quality of information products before they are disseminated;</td>
</tr>
<tr>
<td></td>
<td>• treat data quality as integral to every step of the development of information, including its creation, collection, maintenance and dissemination (TDQM);</td>
</tr>
<tr>
<td></td>
<td>• make its statistical and analytic information products widely available and broadly accessible;</td>
</tr>
<tr>
<td></td>
<td>• identify critical information with known quality problems that have significant negative impacts on the health and disability sector.</td>
</tr>
<tr>
<td>The Clinical Analysis Team</td>
<td>The Clinical Analysis Team will supply the Ministry of Health and the sector with expert advice by employing a subject matter expert. The subject matter expert will, with the assistance of the Clinical Analysis Team, apply consistent TDQM practices to ensure:</td>
</tr>
<tr>
<td>NZHIS</td>
<td>• data suppliers are informed of major data quality issues on a monthly basis;</td>
</tr>
<tr>
<td></td>
<td>• data suppliers are provided with expert advice on possible solutions to data quality problems;</td>
</tr>
<tr>
<td></td>
<td>• regular ‘business as usual’ programmes are in place that assess the quality of data on all national health collections, with effective processes to make prioritised improvements where necessary and possible and in a timely manner;</td>
</tr>
<tr>
<td></td>
<td>• prevention mechanisms are in place, where possible, to confirm that data quality is not degraded once it is received within the Ministry. This includes providing advice to projects developing new collections;</td>
</tr>
<tr>
<td></td>
<td>• the provision and monitoring of KPIs to suppliers. This allows the sector to understand the whole of sector minimum requirements for data quality outcomes;</td>
</tr>
<tr>
<td></td>
<td>• provision of a ‘train the trainers’ education programme for the whole of the sector;</td>
</tr>
<tr>
<td></td>
<td>• data quality caveats are supplied to anyone using data to help them understand the limitations of the data. The Business Intelligence Unit will ensure appropriate caveats are included when providing information to customers.</td>
</tr>
<tr>
<td>The role of the sector -</td>
<td>A national governance group, aligned with the governance structure of the Health Information Strategy (Ministry of Health, 2005) will ensure ongoing sector input into the iterative changes to the DQIS as increasing data quality information becomes available.</td>
</tr>
<tr>
<td>National governance group</td>
<td>The role of the governance group is to oversee the continuing implementation of the DQIS by:</td>
</tr>
<tr>
<td></td>
<td>• championing the programme and educating others within their organisations about the programme;</td>
</tr>
<tr>
<td></td>
<td>• communicating to the rest of the governance board issues that arise in the sector with regard to implementing the DQIS;</td>
</tr>
<tr>
<td></td>
<td>• providing feedback to the rest of the governance board on local effective data quality improvement practices.</td>
</tr>
</tbody>
</table>
Clinical coders
Hospitals employ coders who work in coding units that are closely associated with medical record departments. Most units are lead by a dedicated coding manager. Team leaders will be responsible for coding education and data quality as the need arises and when time is available.

Health Information Standards Organisation (HISO)
HISO plays an important role in data quality by endorsing national e-health standards and provides a forum where key sector interests can be represented to plan, develop and promote the health information standards needed to implement e-health. E-health involves increasing the quantity and type of transactions that can be undertaken electronically.

The development of agreed health information standards will provide for better quality information. Standards allow for the introduction and improvement of health care delivery projects that have an information management component and reduce compliance costs for data suppliers. The ability to transfer data throughout the sector is greatly increased, reducing the requirement for subsequent mapping of data to different coding or classification standards when reaching its destination. Data suppliers and custodians are required to use HISO endorsed health information standards (including the endorsed version of the standard), where available.

Table 47: Roles and Responsibilities for Data Quality in the New Zealand Health Sector

<table>
<thead>
<tr>
<th>Entity</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical coders</td>
<td>Hospitals employ coders who work in coding units that are closely associated with medical record departments. Most units are lead by a dedicated coding manager. Team leaders will be responsible for coding education and data quality as the need arises and when time is available.</td>
</tr>
<tr>
<td>Health Information Standards Organisation (HISO)</td>
<td>HISO plays an important role in data quality by endorsing national e-health standards and provides a forum where key sector interests can be represented to plan, develop and promote the health information standards needed to implement e-health. E-health involves increasing the quantity and type of transactions that can be undertaken electronically. The development of agreed health information standards will provide for better quality information. Standards allow for the introduction and improvement of health care delivery projects that have an information management component and reduce compliance costs for data suppliers. The ability to transfer data throughout the sector is greatly increased, reducing the requirement for subsequent mapping of data to different coding or classification standards when reaching its destination. Data suppliers and custodians are required to use HISO endorsed health information standards (including the endorsed version of the standard), where available.</td>
</tr>
</tbody>
</table>

7.6.4 TDQM at the Ministry of Health – Institutionalising Data Quality

Total Data Quality Management (TDQM) seeks to deliver high quality information products to information consumers and to facilitate the implementation of an organisation’s overall data quality policy. The foundation of the DQIS is the institutionalisation of TDQM principles. Institutionalisation means encouraging data quality management to be accepted as ‘business as usual’; a regular process that is undertaken at all levels in the sector. It is anticipated, through the implementation of the DQIS, that health care organisations will begin to implement similar steps that are appropriate to their organisation. Through the projects outlined below it is expected that expensive, one-off projects can be avoided and money for regular operational requirements will enable an ongoing prevention programme. Projects may only improve processes and/or data in one collection, whereby regular prevention mechanisms help to ensure data across the national collections are of high quality.

TDQM can prevent systematic and repeated data quality problems. Mere human error leading to problems with accuracy may be more difficult to prevent but can be addressed through targeted education. TDQM looks at the processes that data flow through before ending in an information product, making sure that none of those processes change the initial meaning of the data. Some approaches to data quality management target specific errors within a collection or an entire collection but often do not devise solutions to prevent systemic problems. In contrast, TDQM focuses on two key aspects of data management: the data flow processes that constitute the organisation’s business, and the recognition that information is a product, rather than a by-product of these processes (Wang, 1998). Regarding the processes themselves, TDQM seeks to ensure that none of them changes the initial meaning of the data leading to systematic errors and repeated data quality problems. Systematic process errors can be prevented by several means, some of which will depend upon the nature of the business unit and its data. However, we can identify generic prevention mechanisms across business units. Some of the time currently spent daily on fixing errors already in data collections would be better channelled towards the systematic prevention of errors. This can be done
in a variety of ways, and methods may differ from one organisation to the next. Prevention mechanisms common across organisations and business units could include:

- the systematic and ongoing education of data collectors, suppliers and Ministry of Health staff;
- regular ‘business as usual’ processes, that review repeated data quality errors from suppliers and feedback information on issues to suppliers, with support provided for improvement and assistance for suppliers to undertake root cause analysis;
- off-the-shelf or internally made data quality software to reduce time spent on assessing data quality problems (of limited use for complex health data);
- extensive review with the development of a data quality plan for proposed new collections in an organisation outlining data quality risks using the Ministry of Health Principles of Data Quality for New Collections (Appendix 6).

Initial steps have been taken to instil TDQM in data quality processes at the Ministry of Health. How the four steps of the TDQM cycle – defining, measuring, analysing and improving are being managed within the Ministry of Health is discussed below in detail.

### 7.6.5 Defining Data Quality

Initial steps in the strategy were to define what data quality means to the health sector. This was done through the development of the DQEF and the data quality dimensions. The dimensions define what health care data collectors, custodians and consumers considered to be important to measure. Their development was informed by the MDQT focus groups, the Ministry of Health Data Quality Workshop and the Data Quality Dimension Survey, discussed previously. The development of the framework provided for input of internal data customers’ requirements through the organisation wide Data Quality Team (MDQT) and the Ministry of Health Data Quality Workshop. The strategy development then provided for whole of sector input into the data quality dimensions, where data quality improvement should be targeted and how, and where accountability lay.

Continual assessment of the needs of data customers is required and will be managed through a yearly postal survey of customers and discussions at forums and training sessions. The DQEF highlighted the need for a consistent assessment tool of customer satisfaction. The existing situation was ad hoc, using new surveys designed each time customers were surveyed. Therefore, NZHIS was not able to compare previous surveys to ascertain if data quality improvements had resulted in improved customer satisfaction.

Clearly stating the scope of the strategy also provided guidance on defining requirements. Without limiting the strategy to the national health collections, the problems would have appeared overwhelming. However, learning gained to improve data quality on specific collections can later be utilised on all collections found in health care, both administrative and clinical.

Developing data production maps for all major information products will outline the flow of data around the organisation and highlight possible areas of data quality issues by detecting points where data quality may
be impacted (Wang et al., 1998). Data Production maps enable the active management of the production of information, and clarify information requirements from a product perspective, as discussed in Section 1.5.5. When properly developed and supported by senior management they can also be effectively applied to solve complex, cross-functional area problems (Wang et al., 1998). Maps can also be used to summarise the issues for management, to provide a tool for enabling understanding of management of the complexity of improving data quality.

Many of the national health collections have been in existence for more than 10 years. The management of the delivery of health care has changed considerably in this time, but the structure of the national collections has not. Defining the current requirements and uses of existing national collections will provide information on the level of data quality required, indicate where data redundancy is occurring, or highlight collections that are no longer utilised as they do not meet customer needs. Further, the regular information products produced from the national collections, for example, a report on Mortality Data published yearly, may no longer meet customer requirements and their content should be regularly reviewed with customers.

Defining data quality requirements for new collections is also required. The health sector is proposing to implement at least three new national health collections in the next five years. The current situation means there is potential for data quality on each collection to be managed differently, with no structured analysis of data quality requirements prior to the development phase. Therefore, principles for project managers responsible for the development and implementation of new national health collections were developed (Appendix 6). The guidelines require the project team to work with the Clinical Analysis Team to develop a ‘Data Quality Plan’ for each new collection, with clear guidelines for the work required at each step of the project cycle.

### 7.6.6 Measuring Data Quality

The Ministry and the sector needed to improve the data quality information available through objective and subjective measurements. Active and regular measurement of data quality avoids a passive reliance on untested assumptions as to the perceived quality of the data. Further, management are provided with information on the depth of problems, allowing for meaningful comparison with other problems (Redman, 2001).

Reporting on data quality levels to information users is able to be transparent and justifiable. The regular measurement programme involves:

- regular use of the data quality dimensions and DQEF on national collections;
- analysing the availability of data quality assessment software or the potential for in-house development;
- continuing current practice on accuracy measurements where these processes are found to be effective;
- data modelling to assess the usefulness of the data in various contexts when sliced and diced from different perspectives.
The actual levels of data quality found in data collections should be measured using both objective and subjective tools to continually and consistently monitor ‘where we are at’. A consistent measurement process with accurate metrics provides comparable information across time, informing data quality practitioners if improvement was made following interventions. Accurate metrics are the key to effective measurement, but business rules also need to be observed (Wang et al., 2001). The long chain that the data pass through before reaching the national collection means that measurements should be made at various points along the chain to ensure movement of data does not change their meaning. Measurements should consider the maturity of health sector organisations data quality programme as data suppliers (Redman, 2001) and begin with simple and clear data quality requirements from suppliers, progressing to higher expectations as organisations develop more extensive data quality programmes.

Discussions with data quality staff at NZHIS highlighted the current limitations of ‘off the shelf’ data quality software. Vendors have provided demonstrations to the Clinical Analysis Team, who noted the inability of the software and its developers to understand and manage the complexity of health care data. There are many Customer Relationship Management (CRM) software solutions available, but these focus on correct name and address information and are of limited use even on registries such as the NHI.

7.6.7 Analysing Data Quality

Given the early stages of the data quality improvement programme at NZHIS, the analysis phase of TDQM has not yet been implemented. Refinements of measurements are still required, along with the business processes required to support regular measurement of data quality.

The complexity of health care requires an extensive range of decisions, both administrative and health-related, and a single data element can substantiate many decisions and carry impacts across the sector. The importance of data quality analysis is that it allows us to determine the appropriate quality levels and identify areas for improvement where current processes do not support good data quality work. Analysis can draw upon:

- clinical analysis and data modelling – what is not able to be done due to data quality problems;
- analysis of customer survey data;
- consideration of current and future uses of data;
- realistic expectations versus the possible unrealistic expectations of stakeholders;
- corporate knowledge – what the organisation already knows is important and/or relatively easy to fix;
- international standards – does this organisation compare well and are our expectations realistic;
- resources available to make improvements;
- root cause analysis of identified data quality problems.

Analysis could investigate several different aspects of data quality. Using the DQEF to assess collections will provide three levels of data quality information, a score for each criterion, an aggregated score of each characteristic, and an aggregated score of each dimension. Where problems are highlighted across several collections, it may be that further investigation through root cause analysis provides sufficient information to
justify improvements that will impact across several collections. Conducting root cause analysis for identified problems requires methodological enquiry by breaking down all aspects of data collection related to the problem. Developing a cause and effect diagram, also known as a fishbone diagram, can assist in structuring the analysis process (English, 1999a). Often the people working closest to the problem know about the problem and why it occurs (Redman, 2001). An analysis of business processes also helps to define business terminology, metadata and rules for creating metrics, alerts and key performance indicators and provide the logical flow under which data are generated, processed and analysed (Wu, 2004). A decision needs to be made and agreed on around ‘what level of data quality is required for these data to be useful?’

Statistical process control (SPC) measures objectively monitor outliers within a collection or data element by analysing trends, particularly where historical or benchmarking data are available (Carey & Lloyd, 1995). Measuring the data supplied by organisations to the national collections involves performance measurement based on key performance indicators developed through SPC (Carey & Lloyd, 2001), historical information and expectations/uses of data. Statisticians have long been applying SPC to provide quantitative measures to analyse performance measurement. Bowen et al (Bowen et al., 1998) found that SPC procedures have been used to measure and improve manufacturing processes since the 1920’s and these procedures can be used in the production of data. Bowen et al (Bowen et al., 1998) used Pareto charts to track data entry errors and cause and effect diagrams to improve data collection processes. This same methodology also provides considerable information for managing data quality through the consistent measurement of raw data. SPC consists of a set of techniques for measuring process performance, identifying unacceptable variance, and applying corrective actions (English, 1999a). Acceptable parameters can be narrowed as data quality improves, defined and measured in the iterative cycle of TDQM. SPC has been used successfully at the Ministry of Health to define the KPIs for Ethnicity data, as outlined in Section 5.9.2, and this methodology can be used to analyse the different types of data found in the national collections.

7.6.8 Improving Data Quality

Actual improvements may be initiated through one-off projects where current processes do not support good data quality work. Actively finding data quality issues before they cause problems is possible through the regular assessment of collections using the data quality measurements outlined above. Data quality improvement is an iterative cycle, therefore improvement requirements need to be continually assessed by the TDQM cycle as requirements may change over time (Wang et al., 2001). Bowen et al (Bowen et al., 1998) suggest that data quality improvement should be at a level where the expected costs of reconciliation to improve data quality equals the expected costs of decision errors caused by data errors, but this may sometimes be difficult to translate into clinical terms. Any technology investments should be linked with the data quality dimensions which are important to ensure the most effective use of resources (Miller, 1996). Organisations should continue to use existing data quality measures where these are found to be effective. A core set of data quality management activities should be in place for all collections, but for projects where the assessment phase highlights a need, such activities can focus more specifically on developing and improving specific data collection processes.
Prevention is now a large part of the Ministry’s work through a proactive approach that includes:

- regular minimum data quality initiatives for all collections with specific initiatives for single collections only when identified as a specific requirement;
- preventing poor data quality in new collections through a ‘data quality plan’ initiated in the early phases in the development of a new national data collection, with full involvement of the Clinical Analysis team. Data quality is embedded in the system and processes of the new collection prior to ‘go live’.
- continuing what works, as the Clinical Analysis Team have learnt a considerable amount about the operational requirements of maintaining data quality levels;
- the endorsement of standards for the collection, management, storage and use of specific data elements by the Health Information Standards Organisation. This way all stakeholders know and agree to all steps in the data flow process;
- regular ‘business as usual’ processes that review repeated data quality errors from suppliers and feedback information on issues to suppliers, with support provided for improvement.

Data quality improvements are based on an organisation’s priorities and are informed by the previous three steps of the TDQM cycle. The DQEF provides for consistent assessment across collections. Where there are persistent problems found across all or many collections and sufficient resources are available for improvement, these persistent and common problems should take priority. Prioritisation of improvements is also informed through analysis with clear and transparent rules for decision-making. These will be different for each organisation, but the following questions should always be considered.

- What are the uses of the data, now and in the future?
- Do the identified errors impact on clinical decision-making?
- What improvements will provide us with the best return on investment or improved clinical outcomes?
- What improvements can we afford to make in the immediate/medium/long term?
- Do we have the skills available in-house or through outsourcing to make the improvements?

An assessment should be made of the effectiveness of any improvement. This can be done using the DQEF. When applied consistently, the framework provides feedback on the effectiveness of data quality improvement programmes.

### 7.6.9 Data Quality Principles for New Collections

The data quality principles for new collections (Appendix 6) provide a template to plan data quality into the processes and technology of new data collections as early as possible in their development. The principles are designed to assist the development of a data quality plan, with a template and guidelines provided to assist project managers (Appendix 7). A requirement for a plan is included in the documentation requirements of the Ministry of Health Programme Office so that project managers are informed. The
responsibility for the data quality plan lies with the project manager and should be initiated when a new national collection is under development or when significant changes are made to existing collections.

The assessment of a new data collection by applying the principles will highlight the data collection and management points that may impact on the quality of the data. Unfortunately, this research found that whilst the technical aspects of data quality assurance are usually well documented, the documentation of the business processes that support good data quality is often lacking. Information processes must be carefully documented to make the meaning of data transparent to all users. For example, data sources should always be identified to provide the data user with context around the data collection process. Imposing standards and maintaining consistency across business units with data definitions, business rules, and even systems architecture can lead to greater data integration and utility and hence perceived quality. This is where a ‘whole of health sector’ strategy begins to have a significant effect on the quality of data and where centralised leadership is required to ensure that the views of all those who manage and use data are taken into account. Thus, good documentation carries the two-fold importance of establishing standards and managing change. Seven areas of interests are examined within the Principles of Data Quality for New Collections:

- management requirements;
- service level agreements;
- organisational specifications;
- process definitions;
- data collection, processing, data modelling and usage;
- data quality management, measurement, measures, and monitoring;
- technical requirements (data quality related).

Project managers should be aware of the requirements defined by the principles before the project starts to be able to define the related tasks in the project. Project managers and the NZHIS Clinical Analysis Team define milestones in the project plan when the principles are to be applied. The Principles are applied every time a major part of the project changes and requirements as defined by the Principles are impacted. These can be changes in the scope, the technical environment, or in the organisation. The Principles should be applied regularly throughout the project to make data quality risks manageable. For example:

- after defining the processes;
- after defining the organisational needs;
- after the development of the business case;
- after the development of the project plan;
- after defining the technical environment;
- after every step which is related to requirements defined by this checklist;
- in the testing phase of the project;
- at the end of the project, the principles should be a part of the project close-out review.
Not all requirements will be applicable to all data collections. It is up to the assessors, along with the NZHIS Clinical Analysis Team, to select the ‘right’ applicable requirements. Previously, this team had no input into the development of new collections but was expected to manage the quality of data in the new collection once in production. Involving the Clinical Analysis Team in the development phase ensures that expert advice is provided on data quality methodology, the team is aware of potential data quality risks prior to ‘go live’ and can factor this into their regular measurement and analysis of the collection. More appropriate data quality metrics are likely to be developed as the Clinical Analysis Team have a better understanding of the data collection.

The Principles were applied to the development of a data quality plan for the Health Practitioner Index (HPI). The project manager approached the Clinical Analysis Team for guidance on the development of a data quality plan and supplied a draft plan for appraisal. The draft plan highlighted the expertise of the HPI development team in information technology practices, such as detailed data flow diagrams. There was, however, minimal information on the business processes related to data quality management. The project team and the Clinical Analysis Team reported that the template provided clear and applicable guidance to develop a practical data quality plan for the HPI. Management also reported that the template provided the information they required to make decisions on the quality of the collection’s data quality requirements.

7.7 TDQM in the Wider Health Sector - the Proposed Sector Work Programme

Institutionalising TDQM through the DQIS looks to achieve the following outcomes for the New Zealand health sector:

- encourage discussion about data quality through the agreement of a strategic direction;
- raise the profile of data quality within health care organisations through organisation wide data quality groups;
- get the sector moving in the same direction through the development of organisation wide data quality strategies that align with the national strategic direction;
- draw on current practice/knowledge through best practice guidelines developed by the sector and widely disseminated;
- provide clear expectations through accreditation/KPIs/contracts;
- actively reviewing strategic direction regularly.

The initial projects proposed in the DQIS provide for better information on data quality and for prioritising improvements. It is expected that data quality information gathered across the sector will enable a collaborative working party to develop best practice guidelines to further enhance the improvement of data quality.
7.7.1 Accreditation of Data Suppliers

A key issue in a national data quality strategy is how to identify and disseminate best practice and embed it in normal ‘day-to-day’ operation. The approach favoured in New Zealand is accreditation rather than audit. Accreditation suggests devolvement of ownership of data quality with a supporting role from the Ministry. District Health Boards will be responsible for addressing data quality issues within a national framework and getting them approved at the highest level of regional health care funding and provision. Accreditation rather than audit provides a more supporting role from the Ministry. The requirements for accreditation outline an extensive data quality programme that will ensure consistent data quality practices across the health sector.

The Ministry, through the existing Quality Health New Zealand and/or the Royal Colleges accreditation processes, will provide sector organisations with clear guidelines on how to achieve accreditation as good data suppliers. The accreditation process will be developed in consultation with the sector following the development and testing of Best Practice Guidelines. Those organisations that have been able to implement extensive data quality programmes will be accredited as good data suppliers. This may lead to a reduction in the need for time consuming peer review and audit processes.

The proposed sector work programme requires that health care organisations can achieve accreditation if they are:

- taking part in a sector peer review/audit process;
- meeting key performance indicator requirements;
- implementing an in-house data quality education programme;
- developing and implementing a local data quality improvement strategy;
- organising regular meetings with a cross-organisational data quality group;
- aligning with Best Practice Guidelines (when developed).

The components of the accreditation requirements are outlined below.

7.7.2 Taking Part in a Sector Peer Review/Audit Process

It is appropriate that the sector monitor its own performance in relation to data quality practices and outcomes. A peer review and auditing system will provide the sector with the ability to assess data supplier practices and provide advice and learning between organisations. A cyclic peer review process undertaken each year ensures that advice and learning are widely disseminated throughout the sector. The implementation of a peer review process requires staff in the sector to undertake training on how to undertake peer review and audit. The process should be developed and agreed by all stakeholders.

The Ministry will provide and monitor data supplier KPIs. However, KPIs can only review outcomes and not processes. Peer review offers valuable information on data management processes. Those data suppliers that achieve excellent data quality results may provide the sector with valuable learning.
7.7.3 Meeting Key Performance Indicators (KPIs)

KPIs, such as those developed for ethnicity data, (see section 5.9.2) will be included in the DHB District Annual Plans to provide data suppliers with guidelines for minimum levels of quality requirements. Within the DHB District Annual Plans expectations on the quality of data supplied are agreed and signed at the highest level of regional health care funding and provision. The Clinical Analysis Team at NZHIS will assess the performance of data suppliers monthly against the KPIs. Feedback will be given in the form of a report to data suppliers. The report will contain an assessment of any data errors found. Data suppliers will be assisted, where possible, to rectify and prevent persistent, regular data errors. Data quality issues will be expected to be resolved/addressed by suppliers within a reasonable time frame.

7.7.4 Implementing an In-house Data Quality Education Programme

This study found that education of all stakeholders throughout the sector is critical to improving data quality. Accredited data suppliers are expected to provide a formal and regular education programme for management, data collectors and clinical coders within their organisation. NZHIS will assist by offering a 'train-the-trainers' programme, with this education then disseminated throughout the suppliers’ organisations. Data collectors requiring education include clinicians at all levels and in all disciplines, as well as administrative and coding staff. All levels of management should be included in the training programme, given the importance of their support in departmental improvement initiatives.

The training programme should be assessed regularly for its effectiveness and to assess where targeted education should take place. Training tools may be available from the Ministry of Health, such as the Ethnicity Data Protocols training package. Data quality training is further discussed in Chapter Seven.

7.7.5 Developing and Implementing a Local Data Quality Improvement Strategy

Data suppliers are expected to develop and implement their own data quality improvement programme/strategy. The local strategies should work towards preventing errors through TDQM. Such improvement strategies should align, where possible, with the national programme of TDQM, using the six data quality dimensions found in the Ministry’s DQEF as a guideline for assessment and prevention requirements. The data quality strategy should align with the organisation’s business goals and strategies, and ISSPs. The DQIS provides a template for a possible data quality strategy that is applicable to a health care organisation.
### 7.7.6 Organising Regular Meetings with a Cross-Organisational Data Quality Group

A cross-organisational data quality group is needed to facilitate the discussion of data quality issues from an organisation-wide perspective. The group should include representatives of business units that collect, manage or use data, including clinicians and clinical coders. The role of the group is to:

- encourage discussion across business units on data affecting the whole of the organisation;
- create partnerships between IT staff and business areas;
- provide management with the impetus to make changes by providing the same message from all the business units;
- provide feedback to each member’s business unit on data quality improvements or issues;
- provide input into data quality policy and strategy development as well as discussion around proposed projects and the potential impact on the different business units;
- identify critical information on known data quality problems that have significant negative impacts on the organisation.

The organisation’s data quality manager or an equivalent role should facilitate this group.

### 7.7.7 Alignment with Best Practice Guidelines (when developed)

The education and support of data suppliers in their collection processes requires extensive work on behalf of the Ministry, in combination with the sector, to develop applicable and achievable best practice guidelines. The Current State Analysis highlighted that there is already considerable work being undertaken throughout the health sector on data quality. This work will provide extensive knowledge on initiatives and their effectiveness. The information gained from existing and new experience and knowledge of data quality will be used to provide detailed guidelines to data managers on possible ways to improve data quality.

### 7.8 Implementation of the Strategy

The implementation of the DQIS began within the Ministry as the process of developing the DQEF highlighted the need for improvements within the Ministry that could then guide the health sector. Implementation involves the identification of the capabilities, capacities, and competencies that the organisation will need if it is to carry out the results of the formulation component of the strategy (Glaser, 2002). A change in culture around data quality requires the Ministry to provide sector leadership and support for improvement at the data supplier level. The Clinical Analysis Team now has a Senior Advisor, Data Quality, who is a subject matter expert, to assist with the provision of leadership and guidance to the sector.

Much of the proposed work to be undertaken within the Ministry will provide data quality information that will enable the sector to target the solutions that will have the greatest effect on overall data quality. The initial projects aim to develop ongoing, ‘business as usual’, regular assessments of data quality within collections to feed back this information to data suppliers. Table 48 provides details of the projects to be undertaken and led by the Clinical Analysis Team in Phase One of the implementation of the strategy.
### 7.8.1 Phase One

<table>
<thead>
<tr>
<th><strong>Project 1</strong></th>
<th><strong>Data Quality Evaluation Framework</strong></th>
<th><strong>Justification</strong></th>
<th><strong>Key Performance Indicator</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Implement the DQEF and the data quality dimensions</td>
<td>Provides an initial assessment of current data quality status; Provides data quality information to inform the strategy development</td>
<td>All assessments documented</td>
</tr>
<tr>
<td></td>
<td>Develop a priority scale for data quality improvements</td>
<td>The process of deciding what to do first should be logical and transparent</td>
<td>Priority scale for improvements in use on data quality assessments</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Project 2</strong></th>
<th><strong>Customer Satisfaction Survey</strong></th>
<th><strong>Justification</strong></th>
<th><strong>Key Performance Indicator</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Develop a list of priority customers</td>
<td>To ensure we provide the best possible service to our customers</td>
<td>A priority list with justification is documented</td>
</tr>
<tr>
<td></td>
<td>Assist priority customers to ascertain their data quality requirements where necessary</td>
<td>Some customers may not know what they need</td>
<td>Clear requirements for customers are known, and improvement projects target these requirements where possible</td>
</tr>
<tr>
<td></td>
<td>Develop a standardised assessment tool for customer satisfaction to be used regularly in a ‘business as usual’ system to ensure improved customer delivery</td>
<td>We need to consistently monitor whether we are delivering to our customers and whether we are improving over time</td>
<td>Customers are regularly surveyed to assess their satisfaction with service delivery</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Project 3</strong></th>
<th><strong>Principles of Data Quality for New Collections</strong></th>
<th><strong>Justification</strong></th>
<th><strong>Key Performance Indicator</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Develop minimum data quality requirements for new collections</td>
<td>NZHIS currently accepts the management of new collections without input into their data quality requirements at the development stage</td>
<td>Data quality plans are developed for all new collections</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Project 4</strong></th>
<th><strong>Statistical Process Control (SPC) Measures and KPIs</strong></th>
<th><strong>Justification</strong></th>
<th><strong>Key Performance Indicator</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Develop appropriate statistical process control (SPC) measures</td>
<td>To understand our data quality levels; To ensure improvement initiatives have an effect; Analysis of historical data will enable the development of appropriate KPIs for data suppliers</td>
<td>Accurate KPIs are developed and accepted by the sector; SPC is regularly applied to collections and elements to determine outliers</td>
</tr>
<tr>
<td></td>
<td>Develop the business processes required to support the implementation of SPC measures</td>
<td>To ensure TDQM processes are in place</td>
<td>The expected KPIs are assessed monthly against the actual data supplied by DHBs; The NZHIS Clinical Analysis team reports data quality information to the sector on a monthly basis</td>
</tr>
<tr>
<td></td>
<td>Develop the business processes to support the framework implementation as ‘business as usual’</td>
<td>We need to continually monitor our data quality and the impact of improvement projects</td>
<td>The framework is used regularly to assess quality; Action plans for improvement are developed and implemented</td>
</tr>
</tbody>
</table>

Table 48: Phase One, Projects 1 – 4 of DQIS Implementation
7.8.2 Phase Two

The work undertaken in phase one will provide the health sector with extensive information on our current levels of data quality. Business processes developed by the phase one projects will be in place at NZHIS to support the sector in making changes to their own data quality practices and in developing their own improvement strategies. Education of data suppliers has been left until phase two. Whilst education has been highlighted as an urgent requirement likely to have significant benefit, NZHIS required more detailed data quality information to develop an education programme that targeted the most significant data quality issues. The Ministry, with increased information on the quality of the data in the national collections, will begin to make Ministry-wide improvements to enable increased sharing of data, improving the Ministry’s ability to understand health information for policy implications. Table 49 below outlines the projects for phase one of the DQIS implementation. Projects 5 – 11 are to be undertaken and led by the Clinical Analysis team in Phase two of the implementation of the strategy. Project 12 is the first directive for action from health sector organisations to implement an organisation wide data quality team.

<table>
<thead>
<tr>
<th>Project 5</th>
<th>Business Rules and Data Dictionaries</th>
<th>Justification</th>
<th>Key Performance Indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Review business rules and data dictionaries from a Ministry-wide perspective</td>
<td>To ensure a common understanding of data elements no matter where they are used</td>
<td>Whole of Ministry documentation of data dictionaries and business rules, available to all Ministry staff</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Project 6</th>
<th>Review NZHIS team roles</th>
<th>Justification</th>
<th>Key Performance Indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Review the roles and responsibilities of the NZHIS Clinical Analysis Team</td>
<td>To make the best use of our resources and facilitate a more preventative focus on data quality Review of phase one projects – what works, provides good data quality information and prevents data quality errors</td>
<td>The Clinical Analysis Team actively works on strategic prevention of data quality problems, provides sector leadership and education, regularly provides the sector with accurate reports on data quality issues and assists with the prevention of persistent errors from suppliers</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Project 7</th>
<th>Standard Operational Minimum Requirements for All Collections</th>
<th>Justification</th>
<th>Key Performance Indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Develop minimum standard operational data quality initiatives that are feasible and appropriate to apply to all national collections with a review of current requirements for each collection</td>
<td>There is variation between collections on what we do to ensure and improve data quality. Collections now have uses that were not previously planned for</td>
<td>All collections have preventative and improvement measures in place All collections are regularly assessed using the DQEF</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Project 8</th>
<th>Information Product Maps</th>
<th>Justification</th>
<th>Key Performance Indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Develop data production maps for all major information products outlining flow of data around the sector and possible sources of data quality issues</td>
<td>To examine the flow of data that ends in an information product to detect points where data quality may be impacted</td>
<td>All information products have documented data production maps, and quality touch points are identified Where appropriate, improvement projects are in place, targeting areas where data quality could be impacted</td>
<td></td>
</tr>
</tbody>
</table>
| Project 9  
Data Quality Software | Analyse the availability of data quality assessment software or the potential for in-house development | Some processes can be automated with reports generated, saving time on manual work | A report is provided to NZHIS senior management on the available software, with a cost-benefit analysis of the applications considered appropriate |
|------------------------|------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------|----------------------------------------------------------------------------------|
| Project 10  
Metadata for Data Quality | Develop minimum documentation requirements for each collection to ensure that the metadata repository contains adequate information for data quality | To ensure the metadata repository is adequate for maintaining data quality throughout the data life cycle and provides all the data quality information required for users The repository needs to be highly accessible to be of use | Metadata relating to all elements within the national health collections is available to all data suppliers, consumers and custodians |
| Project 11  
Education Programme | Develop a comprehensive education programme for the Ministry and the sector, co-ordinated by one dedicated full-time employee at NZHIS | An education programme that is responsive to data quality information provided by the DQEF and SPC A train-the-trainers approach requires the sector to disseminate the training provided | One dedicated full-time employee at NZHIS to provide a train-the-trainers programme throughout the sector, including Ministry staff Training packages made available to the sector |
| Project 12  
Cross-Organisation Data Quality Team | Organise a Data Quality Team, which includes stakeholders from across the organisation | To bring about discussion among business units and provide information on data quality requirements across the organisation To create partnerships between IT staff and business areas and between business areas Can provide impetus to management to make change when the same message is coming from all the business units | DHBs and the Ministry have data quality teams in place, with clear terms of reference and regular meetings |

Table 49: Phase Two, Projects 5 - 12 of DQIS Implementation

### 7.8.3 Phase Three

The following three projects are large and require considerable input from the whole health sector to ensure applicable guidelines are produced for the sector. The guidelines and the accreditation scheme will be developed through sector working groups, with secretariat support from the Ministry. The guidelines will be developed from the data quality information available through the improved assessment, analysis and management of data quality across the sector following implementation of the first two phases of the DQIS as outlined above. Table 50 outlines the proposed actions for health sector organisations to begin to implement TDQM.
### Project 13
**Best Practice Guidelines**

| Develop best practice guidelines for data quality management processes | Guidelines for data collectors ensure we all have the same understanding of data quality requirements and a common definition/interpretation of data quality processes across the sector | Guidelines are developed from the data quality information available through the improved assessment and analysis of data quality across the sector. A wide range of sector and Ministry stakeholders are involved in the development. Documented best practice guidelines are available and used by the sector |

### Project 14
**Data Quality in ISSPs**

| Include data quality work in the sector ISSP | The ISSPs align with the organisation’s strategic plans, therefore aligning data quality with the organisation’s strategic goals. | All ISSP’s supplied to the Ministry of Health annually include a discussion on how data quality is managed in alignment with organisational strategic goals |

### Project 15
**Accreditation Scheme**

| Develop DHB/PHO accreditation schemes that include: • peer review/audit • best practice guidelines • KPIs • in-house education programme • local data quality improvement programme/strategy • cross-organisational data quality group | To support moving responsibility for data quality to the whole of the sector. Disseminates good practice. Provides clear guidelines for the whole of the sector to follow. Enables data users to trust the data from accredited organisations. | Accreditation scheme developed and agreed by the Ministry and the sector. Peer review system in place providing feedback on best practice and areas for improvement. |

#### Table 50: Phase Three, Projects 13 - 15 of DQIS Implementation

#### 7.8.4 Cost Implications

Increased costs are likely where new staff are required or a change in work practices for existing staff means training is required. Increased costs are most likely in the first year of DQIS implementation, as increased work on prevention may not show immediate benefits through the reduction of rework. Costs may be offset in the future by the ability to prevent errors and therefore reduce rework. Costs will be further reduced through the prevention of systemic errors in new collections if a thorough assessment of potential data quality risks is made and effectively addressed before implementation.
Return on investment is difficult to analyse in data quality, as improvements are likely to impact a wide range of information consumers throughout the sector. It may be more relevant to measure the impact of not making improvements. When considering the cost of not making improvements, organisations should consider that poor data quality:

- compromises decision-making with potentially serious impacts on clinical decision making;
- increases operational costs due to time and resources required to detect and correct errors;
- directly affects the patients’ perception of the organisation;
- reduces employee morale.

Return on investment for data quality improvement is further discussed in Chapter Seven.

### 7.8.5 Monitoring and Evaluation

The Ministry of Health will implement and regularly monitor the effectiveness of the DQIS. Monitoring will take place through the District Annual Plan process, assessment of Information Systems Strategic Plans for the inclusion of data quality strategic plans, and KPI measurements of data submitted to NZHIS. Further, accreditation relies on independent, voluntary programme development that has a focus on training and across organisation assessment of progress towards TDQM practices. Peer review, based on collegiate programmes that focus on the organisation’s data quality processes will provide feedback to organisations within the sector from a peer perspective. A report will be supplied to the Ministry of Health on the outcomes of the review. The governance group within the Health Information Strategy Advisory Committee, as discussed in Chapter Seven, will help monitor the implementation of the programme throughout the sector.

### 7.9 Organisational Change Requirements

The amount of organisational change required depends on current practices. Some DHBs already undertake considerable data quality work but few have staff members who are accountable for data quality. By defining the roles and responsibilities for data quality, all that may be required is a shift in focus to align with the national DQIS, which means the organisation meets KPIs and increased emphasis on training data collectors and department managers, as well as clinical staff. Incremental change is a pragmatic process that allows changes to be assessed for effectiveness on an individual organisation basis. The development of a cross-organisational data quality group will increase awareness of responsibilities across the organisation without impacting significantly on resources.

Data quality software is available, for example, that made by Trillium Software[^13] and Firstlogic[^14], and may assist in automating simple data integrity checks. However, much of the software is expensive and most does not meet all of the complex requirements for health care data. Organisations will need to weigh up the

[^13]: http://www.trilliumsoftware.com/site/content/
[^14]: http://www.firstlogic.com/home.asp?bhcp=1
costs of such software with the benefits of simplified, faster data checks relative to their particular situation and may consider making the change to their existing information only when the system needs replacing.

7.10 Chapter Summary

The DQIS development continues to be an iterative process; priorities should be reviewed once further information about the quality of data within the national health collections and appropriate statistical analysis is available and final consultation has been sought with stakeholders. A survey of 72 health sector stakeholders confirms that the data quality dimensions found in the DQEF reflects the components of data that the sector considers important to measure and improve.

Grounded theory provided the researcher with a structured methodology for the data collection and analysis phases of the action research process. The iterative learning afforded through grounded coding and analysis of the data informed the required changes to the DQEF and the structure of the second phase of the research cycle. The grounded analysis and coding of the data collected from interviews with health sector stakeholders enabled evolved data collection, with changes made to the data collection tools as the researcher began to learn and understand more about the data quality capability maturity of New Zealand health care organisations. The questions moved away from a focus on strategic data quality management to more focus on current operational data quality work, management attitudes and the roles active in data quality management.

A review of data quality within the New Zealand health sector reveals that there are extensive resources being used to ‘find and fix’ data quality problems, such as duplicates on the NHI Register. There are no organisations that identify a management role that has the overall accountability for data quality. It is often found to be the role of the CIO ‘by default’, without any performance expectations in job descriptions. The CIO often has a very good overview of the organisation, but departments such as Decision Support are data users, and were found to play a role in data quality management. Three organisations had in place data quality teams that meet regularly to discuss data quality problems across the organisation. The teams were formed and led by the Information Systems department in response to significant and persistent data quality problems that could not be solved by one department.

Preventative measures centred on education of data collectors, sometimes targeted after auditing of collection practices and data quality errors. Information systems were also designed to prevent the input of erroneous data where possible. There appeared to be considerable understanding of data quality issues and their impacts on decision making, with some organisations displaying a high level of maturity around their understanding of methods to improve data quality. This varied considerably from one organisation to the next. The competing priorities in health care mean that requests for funding for improvement projects or extra staff are not always granted. Whilst some organisations have management who are supportive of data quality improvement, data quality work was entirely initiated by lower level staff.

A Ministry of Health Data Quality Workshop provided the researcher with detailed feedback on how well NZHIS manages the elements of the DQEF and its characteristics, emphasising that there are many areas
where improvements could be made at NZHIS. A draft of the Data Quality Information Strategy was discussed at the workshop and participants felt the strategy provided an applicable work programme to improve data quality in the Ministry of Health and the health sector, and to introduce the institutionalisation of TDQM in the health sector. Participants provided the researcher with knowledge on strategy development and implementation issues likely to be faced.

The resulting Data Quality Improvement Strategy provides the Ministry of Health and the sector with detailed guidelines on how to develop and implement TDQM throughout all levels of the health sector. Roles and responsibilities are clearly defined, along with data ownership. A series of projects provides the required development for business as usual initiatives that institutionalise data quality into every day practice and make use of existing sector knowledge through the development and dissemination of best practice guidelines. For the data quality improvement strategy, it is important to:

- derive and impose standards that facilitate data and information transfer whilst preserving quality;
- re-engineer the business processes to deliver the quality data needed for efficient service planning and the effective practice of integrated patient care;
- identify and disseminate best practice to reduce the development time needed to improve data quality;
- ensure data quality levels are not unnecessarily rigorous to maintain user ownership and workloads at reasonable levels;
- define user accountabilities for data quality and the mechanisms to enforce them;
- seek to embed the search for data quality in normal working practices and recognise its achievement in appropriate ways such as accreditation.

The Ministry’s purpose is to realise the full value and potential of the data that it collects, stores, and manages. Building ‘trust’ in the data throughout the health sector will ensure that data are used frequently and to their greatest possible benefit. With the right framework and strategy, data that are highly utilised for a range of reasons will incrementally improve in quality. Extensive data mining, combining currently disparate collections, will also provide far more granular information and knowledge to improve these collections. A data quality strategy will provide coherent direction towards Total Data Quality Management through a continuous cycle of work. The improved data quality will then ensure that the health sector is more able to make informed and accurate decisions on health care policy and strategy.
8 Detailed Discussion and Further Analysis of Results

Chapter Seven discusses further the implications of the findings related to all of the research questions found in Section 1.6.2 and discusses in detail the research question:

- Does action research methodology provide appropriate research methodologies for understanding learning and change initiation in a data quality improvement programme?

The chapter discusses the implications of the research results on the New Zealand health sector’s ability to use the considerable and valuable data collected under the custodianship of the NZHIS and held in the national health collections.

8.1 The Results in Context

At present, data quality management in the New Zealand health sector is entirely initiated through ‘bottom up’ work, in general through the information services or information technology teams. Management has not yet taken responsibility, and indeed does not understand their role in the data quality process. Data quality work was initiated when staff became frustrated with their inability to use the data for its intended purpose and there was no one else in the organisation responsible for overall data quality management. This bottom up approach has introduced silos of data quality improvement work, introducing more data quality issues through the increasing discrepancy of data from the same source, and producing conflicting reports. This was evident in all types of organisations in the health sector. Porter (Porter, 1991a) states the function of a strategy is to integrate the activities of diverse functional departments to ensure consistency within an organisation with explicit, reinforcing goals and policies when senior management cannot participate or monitor all decisions directly.

The development and implementation process for the DQEF began to initiate a structured data quality programme within NZHIS, almost as a ‘side-effect’. The initial aim of the DQEF was to provide a consistent measurement tool for all national collections. However, the learning provided by the development of the DQEF had an overall effect on the way in which data quality was managed within the organisation. This has been further supported by the development of a sector wide Data Quality Improvement Strategy (DQIS) and there was a noticeable impact on NZHIS management and staff thinking and practices around data quality. Data quality began to be viewed from a more holistic perspective, with the realisation that subjective feedback from data users was an important indicator of data quality, just as much as the objective assessment of the completeness of data, for example.

The DQIS provides the vision and clear direction needed to ensure resources are targeted appropriately, and a context to help stakeholders understand the probable directions of change and the desired outcomes (Lorenzi, Riley, Blyth, Southon, & Dixon, 1997) by analysing the choices available to provide the best possible improvements. This vision provides change agents with an accurate understanding of the change to implement, enabling the communication of realistic expectations for their organisation (Elrod, 2002).
Robson (Robson, 1997) notes that a successful strategy ‘exploits opportunities and fits the circumstances at the time’ with a requirement, therefore, to undertake a systematic, skilful, accurate and realistic assessment of the opportunities and to re-evaluate them. The Current State Analysis and the subsequent Data Quality Capability Maturity Assessment that compared New Zealand health care organisations with international organisations provided the researcher with an understanding of where to base the strategy to exploit current capabilities.

8.2 The Role of the Researcher

The researcher was central to the research process as the project initiator and project manager throughout the DQEF development and evaluation process. The researcher, in the role of Information Strategy and Policy development, was expected to assist the Clinical Analysis Team to develop a comprehensive data quality measurement tool. However, the researcher’s role was for a limited time and work on data quality improvement is the responsibility of the Clinical Analysis Team. The researcher, therefore, had to take care not to ‘take over ownership’ of the DQEF development. Lee et al (Lee et al., 2004), Clark (in (Baskerville & Wood-Harper, 1996)) and Argyris and Schön (Argyris & Schön, 1978), note the essence of action research is the objective of advancing theories as well as facilitating organisational change, with the researcher in a facilitative role. Acceptance of the DQEF by the Clinical Analysis Team was paramount to the success of its implementation and changes to current work practices. Where the research required that the DQEF be tested on collections, the researcher stepped back to allow the Clinical Analysis Team ownership of the DQEF as a tool for their use. The researcher requested feedback on the findings of the use of the DQEF through an issues register, email and face to face discussions.

One of the fundamental principles of action research, according to Baskerville (Baskerville, 1999), is that the research process enhances the competencies of the respective actors. This was possible through the researcher’s role of imparting knowledge on data quality theory and international practices reported in the literature and suggesting ways this knowledge could be applied to the local setting. The research participants provided their knowledge on current data quality practices within NZHIS, the uses of the data collections throughout the Ministry and data quality requirements for those uses, the practicalities of undertaking data quality measurements, and technical expertise on database management. Further, the researcher did not have a technical background, and the other participants had not worked at a national strategic level. This produced a team approach to development and improvement to the framework where all participants gained considerable learning and understanding of the discipline of data quality. Theory was closely aligned with the practical reality of the operational running of the organisation and the impact of poor data quality on the whole of the health sector.

Susman (Susman & Evered, 1978) sees the ‘specifying learning’ phase of the action research cycle as an ongoing process. Considerable learning took place within the organisation throughout both action research cycles. This is highlighted in areas where research participants had little knowledge on how to solve problems, such as the development of applicable metrics for the framework. This required the researcher to return to the literature, study the learning, apply this to the local environment, and impart new knowledge to
the study participants. The study participants became familiar with the current theory and began to apply this in their daily work and to the development of the DQEF.

8.3 Bringing about Change Through a Paradigm Shift

Throughout the research, as learning took place, the needs of the organisation and the researcher evolved. The researcher found there was a need to investigate possible theories to assist in managing the changes required for the implementation of the DQEF and the DQIS and to apply these practically to the research environment and to the DQIS development. The second cycle of action research highlighted issues around change management. Elrod (Elrod, 2002) summarises current change models and finds there are considerable similarities, with most highlighting a transition from normality through some form of disruption and then to a re-defined normality. As organisations pass through the region of disruption, performance is usually diminished. The role of change agents is to minimise this period of diminished performance. The implementation of the DQEF and DQIS required changes to daily work practices by the Clinical Analysis team and health sector organisations. Staff who would use the DQEF were already working to full capacity, had been working at NZHIS for many years in the same or similar roles, and were resistant to change in their daily work practices.

Fry (Fry, 2002), explains the three components of transformational change as being novelty/innovation, transition and notably, continuity, where the past is taken into account. Mintzberg (Mintzberg, 1987) notes that ‘strategies are both plans for the future and patterns from the past’. Appreciative enquiry with a problem based approach was utilised by the researcher as Fry (Fry, 2002) points out that change using appreciative enquiry focuses on the positive rather than the negative, and this theory is supported by ‘reinforcement theorists’ such as Skinner (Skinner, 1963). Long and Seko (Long & Seko, 2002) in their evaluation of the CIHI Framework implementation, stress that staff need to know that the evaluation results are in no way a reflection of their own personal abilities. The researcher needed to take care that the current work of data quality practitioners was still perceived to be of value. The DQEF provided structure to an existing data quality programme; the goal was to continue with practices that proved effective in measuring and improving data quality and to add to this a structure that guided decision-making on priorities of work. Process management seeks to supplement functional management (which is still needed) by linking up the various functional organisations or business units (Redman, 1994). This is reinforced in the DQIS through the development process for the Best Practice Guidelines, with the drawing together and documentation of ‘what works’ in the health sector for dissemination and adoption where applicable.

Lewin (Lewin, 1951) discusses Field Theory in understanding individuals and groups within organisations. Field Theory establishes the need for active involvement of individuals in the change process as those not involved may resist because they might see changes as detrimental to them personally. The issue is one of ownership, if group members perceive that they own the problem and the solution, they will work with researchers to make the change work (Lorenzi et al., 1997).

Pettigrew, Ferlie and McKee (Pettigrew, Ferlie, & McKee, 1992) also note the change in focus in health care management from policy (content) to strategy (content plus process), supporting the requirement of process
management in data quality. Policy development by itself is insufficient to bring about change. Mintzberg (Mintzberg, 1987) notes that managing strategy is mostly about managing stability. Formal planning processes that repeat themselves too often desensitise the organisation to real change. The DQIS should be reviewed for effectiveness and altered where further improvements are required, but the basic ‘rules’ of the strategy, such as the data quality dimensions and the use of TDQM to institutionalise data quality should remain in place.

Coherence in the management of change requires a purpose and belief among the senior management team (Pettigrew et al., 1992). Elrod (Elrod, 2002) found that the CEO’s level of individual development directly influences an organisation’s ability to transform and develop, supporting the notion found in much of the data quality literature that executive level management support is paramount to a successful data quality programme. This research found that senior management were more supportive in organisations where the executive team received education and information on data quality issues.

8.4 The Complex Adaptive System in Health Care

Plsek and Greenhalgh (Plsek & Greenhalgh, 2001) introduce the science of complex adaptive systems (CAS) to understand and bring about change in the health care environment. A complex system is defined as ‘a system with many independent agents, each of which can interact with others’ (Penchas, 2003), that can behave very sensitively and be influenced by small initial differences (Champagne, 2002). Health care systems are complex systems with many independent agents each interacting with the others, creating complex adaptive systems containing emergent learning and change potential (Penchas, 2003). Within each complex system can be found complex sub-systems, such as the Ministry of Health or DHBs within the complex system of the New Zealand health sector.

CAS moves away from the traditional reductionist thinking of troubleshooting then fixing, with a shift in attention towards those actions that seem to be working well. Plsek and Greenhalgh (Plsek & Greenhalgh, 2001) relate CAS to the Total Quality Management (TQM) cycle of plan, do, check, act, with TQM and CAS supporting the exploration of new possibilities through experimentation, autonomy and ‘working at the edge of knowledge and experience’. The concept provides wide boundaries for beginning the work of self organisation (Committee on Quality of Health Care in America, 2000), (Champagne, 2002). Sidthorpe, Glasgow and Longstaff (Sidthorpe, Glasgow, & Longstaff, 2004), of the Australian Primary Health Care Research Institute, further explore the notion of CAS and how people relate to their environment:

Agents respond to their environment using internalised ‘short lists of simple rules’ that drive action and generate behaviour (Eoyang and Berkas 1998; Plsek 2003). The rules need not be shared, explicit, or even logical when viewed by others, but they nonetheless contribute to patterns and bring coherence to behaviours in complex systems. Deliberately exposing and changing underlying simple rules leads directly to innovative ideas (Institute of Medicine 2001). In addition, short lists can be used proactively. When a new system is being instituted, a short list of simple rules (or minimum specifications) may be the most effective way to bring about change. They set the parameters and provide both focus and freedom for system activities. Over-prescription is counter-productive because it stifles creativity and innovation.
Champagne (Champagne, 2002) also notes that complex processes can emerge from the actions of staff who follow some relatively simple rules. The researcher utilised CAS to assist with the change in work practices required to implement the DQEF and the DQIS and to encourage innovation through the development of simple rules that still allow for innovation, whilst supporting the current work practices that are effective.

Involving a team of data quality practitioners and data users in the development of the DQEF and the DQIS meant the strategy had a ‘grass roots’ approach. Those that work with the data every day currently initiate data quality work in the health sector at the ‘grass roots’. This grass roots work may give rise to an emergent strategy, whereby a series of actions converge into patterns that become deliberate if the pattern is recognised and legitimised by senior management (Pettigrew et al., 1992). Strategic management aims to create the future as well as reacting to changing times and is about ‘reading the signs and portents for the future and interpreting them in order to choose an appropriate direction for future development’ (Robson, 1997). The DQEF provides the data quality information required for strategic decision making on effective data quality management, thereby deriving and describing the organisation’s emergent data quality strategy.

The umbrella (Mintzberg, 1987), high level component of the DQIS provides simple rules for health care organisations to follow when developing a data quality strategy for their own organisation, with the detailed components left to each organisation to develop. Highlighting the positive ‘attractors for change’ (Plsek & Wilson, 2001), such as improved data quality leading to less re-work and reduced conflict through clarification of roles, found to be issues in this research, provides for a more constructive approach to change.

8.5 Data Quality Capability Maturity of the New Zealand Health Care Sector When Compared to Organisations Outside Health Care

Organisations surveyed by the researcher from outside of health and some of which were outside of New Zealand were compared with New Zealand health care organisations, with the aim of understanding the context of data quality in New Zealand organisations, and this is shown in Table 51. Hamel and Prahalad (Hamel & Prahalad, 1994) discuss the need for external benchmarks to understand industry best practice and identify key capability challenges. Benchmarking helps to highlight specific areas where an organisation is not performing to best practice standards, and may provide evidence to management that more could and should be done. The comparison of New Zealand health care organisations with overseas organisations provides this benchmark.

The assessment of data quality maturity provides information to management that is a quantifiable score of how well data operations manage the quality of their information. Dravis (Dravis, 2004b) utilises a data quality maturity grid and notes other reasons to understand the data quality maturity capability in an organisation include:

- identifying where in the data quality maturity progression an organisation currently resides provides managers with a benchmark, and a direction to improve;
- scores will reflect that some organisations are further ahead or behind than others;
- managers will understand their organisation's behaviour - good, bad, or indifferent - towards data quality;
- managers will have a framework to begin changing an organisation's attitudes towards management of information;
- managers will know what progressive actions to pursue as the maturity level increases, which will vary depending on level.

The researcher found that both types of organisations are mostly similar in data quality capability maturity. Most New Zealand health care organisations do not have defined roles for data quality, however, and many of the international organisations did define at least one role for data quality. The sample completing the overseas questionnaire may not be a representative sample as many were attending a data quality course and may have more maturity around data quality than typical organisations.

<table>
<thead>
<tr>
<th>Measurement Category</th>
<th>Stage – Mature New Zealand Health care Organisations</th>
<th>Stage – Mature International Organisations Outside Health Care</th>
</tr>
</thead>
<tbody>
<tr>
<td>Management understanding and attitude</td>
<td>Stage three – while going through data quality improvement programme learn more about quality management; becoming supportive and helpful</td>
<td>Stage three</td>
</tr>
<tr>
<td>Data quality organisation status</td>
<td>Stage two – data quality role is appointed but main emphasis still on correcting bad data</td>
<td>Stage two</td>
</tr>
<tr>
<td>Data quality problem handling</td>
<td>Stage two – teams are set up to attack major problems. Long range solutions are not solicited</td>
<td>Stage three – Corrective action communication established. Problems are faced openly and resolved in an orderly way</td>
</tr>
<tr>
<td>Cost of data quality as a percent of operating costs</td>
<td>Stage one – no organisation has measured the actual cost of data quality and there is no methodology developed</td>
<td>Stage three – Reported 10% Actual 15%</td>
</tr>
<tr>
<td>Data quality improvement actions</td>
<td>Stage three – data quality programme in place</td>
<td>Stage three</td>
</tr>
<tr>
<td>Summation of data quality posture</td>
<td>Stage three – “through management commitment and data quality improvement we are identifying and resolving our problems”</td>
<td>Stage three</td>
</tr>
</tbody>
</table>

Table 51: Comparison of Data Quality Maturity in New Zealand Health Care Organisations with International Organisations Outside of Health Care

The most mature of the international organisations studied had implemented Six Sigma methodologies several years ago, and the data quality team used this methodology where applicable. This provided for more maturity than the New Zealand organisations possess, particularly around the measurement of the cost of data quality. Surprisingly however, the data quality work done in this more mature organisation was still somewhat limited due to an emphasis on improvement and data cleansing rather than the prevention of errors through process management. The management of data quality problems was, in general, handled in a more structured way by the international organisations, with clearer reporting lines to departments.
responsible for their own data quality. This may be due to the length of time that data quality initiatives had been in place. In New Zealand the DHBs are relatively new organisations having been set up in 2001 and are still developing institutional knowledge.

The analysis of capability maturity confirms that TDQM and the strategic management of data quality is a relatively new phenomenon, still rarely found across organisations in New Zealand and overseas. New Zealand health care is certainly not ‘lagging behind’, although NZHIS does state as one of its mission goals that it aims to be world leading in its knowledge and leadership of health information management. Leading a strategic data quality programme utilising TDQM principles, for the whole of the New Zealand health sector, would constitute world-leading practice. NZHIS currently has the capability to do this, with strong management support, a structured assessment tool with defined data quality dimensions, and the methodology to develop appropriate data quality metrics.

The New Zealand health sector organisations studied are beginning to move towards preventative initiatives for data quality. Loshin (Loshin, 2004) finds a mature data quality program determines where the risks are, the objective metrics for determining levels and impact of data quality compliance, and approaches to ensure high levels of quality. Lack of accountability and role definition in the New Zealand health care organisation appears to be the most significant hindrance to effective data quality management, along with the need for a formal, regularly utilised assessment tool.

The level of support for data quality from management in the New Zealand health sector varied, and this did not have any correlation with the size of the organisation. Black et al (Black et al., 2004), when studying data quality in clinical databases, found that levels of data quality were not impacted by the size of the organisation. A study of accounting organisations in Australia (Nord, Nord, & Xu, 2005) also found that the size of the organisation did not have an impact on the perceived importance of data quality. Whilst the organisations in this study vary considerably in size and data quality maturity, the level of awareness of data quality issues was reasonably consistent. However, the size of the organisation may impact on the successful implementation of a data quality strategy. Robson (Robson, 1997) states that the size of the organisation affects the nature of its strategic problems and the resources available to deal with them. The need for a generally agreed strategy is more apparent in large organisations, because of the complexity and diversity of their actions. Smaller DHBs and PHOs may not feel that a data quality strategy is required at their level, the national strategy may be considered sufficient to guide the management of data quality issues for the national health collections.

A 2005 study on information management trends in the US and Canada found that improving data accuracy and integrity is the most important issue in business intelligence systems for 75% of respondents (TechRepublic, 2005). The New Zealand health care organisations appeared to have less awareness of data quality at the management level and it is unclear why. It is possible that the lack of formal assessments of data quality means that managers can chose to deny, or may not be aware of, data quality problems in their departments. Those organisations in this study with organisation wide data quality teams appear to be considerably more mature in their overall management of data quality, and this is discussed below.
8.6 The Impact of the Data Quality Evaluation Framework and Dimensions

The process used to develop the DQEF provided the Ministry of Health with a platform on which to begin to improve the maturity of their data quality programme, through the introduction of a structured methodology for the assessment of data quality and the education of data quality practitioners and managers. The Current State Analysis of Data Quality at the Ministry of Health highlighted the need for such a programme, but also that the Ministry of Health is no less mature in data quality management than most organisations. However, the core business of NZHIS is the management of health data, with an expectation to provide leadership to the health sector. It was therefore timely to begin an education and change programme around data quality.

Price and Shanks (Price & Shanks, 2005) reviewed the defining of appropriate data quality dimensions in the literature and found that the process should not be limited to a purely theoretical approach. A complete approach to defining quality must also take into account the suitability for a specific task from a consumer’s perspective. The process of developing and verifying the chosen dimensions in this research utilised a theoretical approach, alongside the practical application of the dimensions on the national collections. The researcher analysed the data quality literature against the CIHI chosen dimensions and added the privacy and security dimension to take into account stakeholder perceptions of requirements, and confirmed by the Ministry focus groups. The Ministry Data Quality Workshop further verified the dimensions from a Ministry perspective. The Data Quality Dimensions Survey provided verification from health sector stakeholders that these dimensions were considered important to be measured and improved.

Tayi and Ballou (Tayi & Ballou, 1998) highlight the importance of understanding data quality dimensions, in stating that data quality cannot be addressed effectively without an understanding of these dimensions. The dimensions included in the framework provide the data quality practitioner with direction and structure for an extensive data quality programme. The dimensions provide some of the ‘simple rules’ required to bring about change in a complex adaptive system (Plsek & Greenhalgh, 2001). Whilst each health care organisation may not wish to implement an extensive data quality programme or the size of the organisation may not warrant or permit extensive work, the dimensions may prompt data quality practitioners to see data quality as more wide ranging than just accuracy. Where current improvement programmes or practices relating to a data quality dimension are already in place new work should support and develop this further. Organisations are able to develop data quality strategies using the data quality dimensions as a guidelines or ‘simple rules’ to follow. Data quality practitioners in NZHIS and the sector felt they had little understanding of data quality theories and gave positive feedback to the researcher on the dimensions for providing simple but evidence-based structure to data quality management.

This research elicited the data quality dimensions considered by the New Zealand health sector to be important to measure and improve. Whilst accuracy was considered the most important, the respondents to the survey rated all of the dimensions closely. Study participants within the Ministry of Health also found rating the dimensions difficult. All study participants thought this was a ‘context of use’ issue, as found by Lee et al (Lee et al., 2002). It would be appropriate to rate the dimensions in order of priority against particular national collections and particular improvement projects. The importance of the NHI to the
movement and use of all health care data does indicate that improvements on the NHI would initially take priority over other improvement projects, and this is reflected in the DQIS. It would be appropriate to develop a priority rating system that data quality practitioners and NZHIS Management develop and agree on, to provide guidance on the prioritisation process.

In health sector organisations, the inclusion of the dimension of privacy and security appears applicable. NZHIS staff felt that the criteria in this dimension should always be managed appropriately and a not known or not met score required immediate attention. The WAVE Report (WAVE Advisory Board, 2001) noted that the privacy legislation was widely misunderstood and often poorly applied. A privacy and security policy is currently under development at NZHIS and will assist to support the implementation of the HIS NZ action zones that pertain to the movement and management of patient data. The new policy will also provide clarification to data custodians around the operational management of data.

The 2003 CIO Survey (Ministry of Health, 2003a) asked CIOs several questions on privacy and security policies and procedures. All DHBs have an information security policy, however only five, from a possible 18 DHBs, felt comfortable with the level of polices, procedures, and people to adequately manage IS security and many were considering reviewing their polices. Perhaps the biggest concern was that seven DHBs felt there was considerable risk around unauthorised insider access to information and inappropriate disclosure of information. Budget was cited as the greatest barrier to improving IS Security. The inclusion of this dimension in the framework serves as a guide to minimum requirements that can be made applicable to DHBs and PHOs. The current characteristics are applicable within the Ministry of Health, and some will be applicable in the wider health sector. Further, the formal assessment of data quality, including the privacy and security dimension could serve to highlight these risks to management, and may assist in securing funding for improvements.

The researcher found that decisions within the Clinical Analysis team at the Ministry of Health on ‘what to measure’ tended to be based on ‘what was easy to measure’ and appeared to be in absence of any structured decision process. Regular assessments concentrated on significant data quality issues that were already apparent. Little, if any, structured assessment of data quality took place in the health sector organisations. Through the assessment of data quality by all of the data quality dimensions, assisted by using the DQEF, the practitioner is aware of all possible data quality problems that are considered by the Ministry of Health and the health sector to be important to measure and improve. This provides for the development of a prioritised programme of improvement work based in part on the outcome of the assessments. Breaking down each dimension into characteristics and criteria provides for detailed analysis that can also be aggregated to provide the appropriate level of detail used by data quality practitioners in data quality reports to management.

The documentation of data quality information provides a history of data quality issues and therefore a means to assess the effectiveness of improvement projects. Documentation at NZHIS on the national data collections and their management was found to be absent, or where present, not maintained. ‘Data Quality Documentation Folders’ were initially proposed by the researcher in response to the difficulties in finding information on national collections, as it is spread out across the organisation. However, through further
learning the researcher and participants found that this information is better kept in a Metadata Repository, as discussed later in this chapter. It is paramount that the data quality information provided from the evaluation of collections is available to data users. Whilst some of this information may be considered sensitive to the organisation, these considerations are outweighed by the benefit of providing data users with context and understanding of the data to enable users to decide if the data are ‘fit for use’. Better data quality information afforded by the evaluation of each collection provides NZHIS with the specific information required to develop applicable data quality metrics or to validate those currently used in the DQEF.

8.7 The Data Quality Improvement Strategy

The aim of the development of the Data Quality Improvement Strategy is to institutionalise data quality in the New Zealand health care sector. The phase one projects proposed in the DQIS attempt to bring about a change in the way data quality is managed through the development of ‘business as usual’ tools. The use of these tools on a regular basis provides a structured and holistic programme with the aim of preventing data quality problems before they arise and to prevent the need for expensive, one off, clean up projects that do nothing to prevent the same problems happening again. The components in the strategy aimed at health sector organisations do not impose the rules on ‘how to’ but provide guidance on effective means of improving data quality from an organisation wide perspective. The formation of an organisation wide data quality programme in the complex adaptive system of a health care organisation is likely to provide the organisation, and the sector through the development of the best practice guidelines, with innovative methods to solve common issues in data quality management in New Zealand health care.

The management of data quality for each national collection reflects the history of NZHIS and the New Zealand health sector. Each collection was developed at different times, under different health care system structures, with varying technology. Much of the data quality work on the collections is in response and reactive to known problems, with the exception of business rules and edit checks to prevent the input of incorrect data. This can introduce problems for data suppliers and users, who need to understand the details of each collection to provide data and use them in their correct context.

8.7.1 The Impact of an Organisation Wide Data Quality Team

A key factor to emerge from this research is the significant impact of the presence of a data quality team. Data quality teams are embedded into three of the organisations studied, with one team only recently established. The researcher found that where a team was present there were significant differences in the way data quality was managed from where there was no team. Plsek and Wilson (Plsek & Wilson, 2001) note that interactions within the complex adaptive system (CAS) are often more important than the discrete actions of the individual parts. Productive or generative relationships occur when interactions among parts of a complex system produce valuable, new, and unpredictable capabilities that are not inherent in any of the parts acting alone. Senge’s theory of the learning organisation also notes the importance of generative learning, where the learning enhances the capacity to create (Smith, 2001). The CAS requires a whole of system view to encourage generative relationships amongst stakeholders that may provoke more creative ideas, and this could be brought about by the presence of a data quality team. Indeed in a CAS successful
strategies do not result from fixing an organisational intention and mobilising around it, they emerge from complex and continuing interactions between people (Rosenhead, 1998).

Stacey (Stacey, 1993) highlights current management theory that extraordinary management, as opposed to traditional management with ordinary day-to-day problem solving, requires the activation of the tacit knowledge and creativity available within an organisation. This necessitates the encouragement of informal structures such as groups set up in response to paradoxes, anomalies and conflicts, with membership drawn from different business units, functions and levels. These groups should be self-organising and capable of redefining or extending their remit rather than being bound by fixed terms of reference. This provides for group learning and innovation. This was the experience of the data quality teams existing in New Zealand health care organisations. All were initiated by staff rather than management in response to organisational problems that could not be solved within one department, were initially informal, and by necessity drawn from different business units across the organisation.

Traditional management thinking has viewed the parts of the organisation in isolation, in contrast, CAS theory suggests that relationships between the parts are more important than the parts themselves (Plsek & Wilson, 2001). In the present study, the organisation wide data quality teams noted the discussion of data quality issues with staff from various areas across the organisation proved valuable. Much of the literature on CAS notes the potential of designing hospitals into ‘complex systems of adjustment’ (Champagne, 2002; Stacey, 1993) which seeks to increase complexity to bring about change, as the probability of change increases with internal diversity. This diversity can be encouraged through the increased participation in decision making by the members of the organisation and better linkages between departments, such as cross department groups. The organisation wide data quality teams in the present study are a good example of increased participation of staff from across the organisation.

The DQIS recommends to sector organisations that an organisation wide data quality team be the first step in the development of a structured data quality programme. Careful analysis of the data provided from the interviews with sector stakeholders, as outlined in Section 6.2.6, and the observations of the researcher when participating in the Ministry of Health MDQT, showed that where an organisation wide data quality team is in place the organisation manages data quality at a higher level of capability maturity. This relationship was discovered through the analysis of interdependencies found between grounded theory codes and is illustrated in Figure 17 below. Where the researcher had coded ‘organisation wide data quality team, these organisations were also found to be coded with many of the following codes:

- whole of system view;
- improved management support;
- ability to create change;
- improved management knowledge;
- better defined roles and responsibilities;
- better organisational attitude to data quality work;
- clearer ownership and stewardship.
These codes indicate data quality capability maturity is at a higher level than those organisations not found at these codes.

**Figure 17: Grounded Theory Codes Identifying the Relationship between the Presence of an Organisation Wide Data Quality Team and More Mature Capability in Data Quality Management**

There is significant benefit from viewing data quality issues across the organisation. The researcher discovered, by identifying where coding overlapped, that there are specific areas of difference in the organisation with a cross organisation data quality team. Differences found indicated data quality was managed from a ‘whole of system’ view, enabled by the presence of staff from a variety of departments. This has the potential to reduce the internal department management of data quality to suit the requirements of that department only. Management support was gained, either through the presence of the CIO, the education of the management team (thereby also improving management knowledge), or through better reporting on data quality issues directly to management. The management support gained assisted in making change happen, and through the ability to infiltrate the data quality programme into all areas of the organisation. Roles, responsibilities, ownership and stewardship became clearer as the organisation began to understand that there was no one staff member responsible for data quality and its implications. Whilst responsibility is still not included in job performance expectations, the ‘all of system’ view highlights the responsibility of everyone in the organisation for data quality as data moves from one area to another. This had the flow-on effect of improving organisational attitudes to data quality, with less resistance to collaborative problem solving.

The structure of organisation wide management of data quality would differ due to the size of the organisation. In the New Zealand health care environment organisations are relatively small and would require only one data quality team for the whole of the organisation. Larger organisations may require teams within departments (or one for each information product), of which one member is the team leader, who attends the organisation wide data quality team meetings as the department or information product
representative. To Redman (Redman, 2001) and English (English, 1999a) the main functions of a data quality team are to:

- ensure the data quality programme penetrates the whole of the organisation;
- resolve across organisational issues;
- provide support, knowledge and education;
- recommend policies and approve standards.

Data quality work in this study was often initiated by the development and implementation of a new system, where organisations wanted to prevent errors in the new system. The organisation wide data quality team is ideally placed to provide input into the development of the new system by providing expert advice to developers, who are often sourced from outside the organisation and possibly have little understanding of the complexity of health information requirements. The data quality teams in the organisations studied saw this as their role and were providing advice to project teams implementing new systems. Ensuring the quality of new information systems is most effective when threaded through each phase of the project and when its significance is honoured by all parties involved (Rusin & Williams, 2001) and involving data users has been found to have a significant effect on the subsequent data quality (Zeffane, Cheek, & Meredith, 1998). Fuller and Redman (Fuller & Redman, 1994) found the need for a ‘corporate data quality policy’ to enable the management of data in a coordinated way. A common policy prevents the development of more information systems with no common data standards and multiple copies of the same data altered in one collection but not others, that lead to data alignment problems (Alshawi et al., 2003), noted to be a significant problem throughout the New Zealand health sector.

It was evident that staff were unaware of issues affecting other areas of the organisation that were similar to their own, until the issues were discussed within an organisation wide group. Further, roles and accountability could be more clearly defined. Through these discussions it became obvious that each department needed to take responsibility for their own data quality. This change in focus was increasingly encouraged when department managers were either on the data quality team or had representation that fed back information from the meetings to their department managers. The importance of department managers’ support in data quality processes in all areas was highly significant and discussed frequently by respondents. Department managers were included in training sessions by two of the organisations studied. Given the considerable support required from department managers, their understanding of data quality issues is paramount and they should therefore be included in training programmes. One team, of which the CIO was a member, provided training to the executive management team through a presentation, which they found successful in securing high level support.

The membership of the organisation wide team is important. The involvement of the CIO through their link or status on the executive management team appears to improve the likelihood of general management support. Rusin and Williams (Rusin & Williams, 2001) note that a quality strategy is sustained by a well articulated philosophy fully supported by top management and based on measurable objectives. The need for top management support is noted throughout the data quality literature and supported by the findings in this research. One CIO from an organisation that is relatively mature in data quality management found there
was a need to engage management before the data quality team could be effective, and noted in the interview with the researcher:

*we had a group that I went to when I first arrived. But the issues that were coming up were management performance issues not data integrity issues so in the end I stopped going and the group ceased to exist.*

However, the membership of the data quality team does not require management presence beyond the CIO, providing support is gained from senior executives for the work the data quality team undertakes. Study participants thought the structure and function of the data quality team must be carefully considered against the cost in staff time to attend meetings and undertake subsequent actions. Black et al (Black et al., 2004) found that there was often technical representation, epidemiologists, statisticians and IT specialists on data quality teams, but few saw the need for representation from managers.

Where possible, data collectors, custodians and users should be members as each has a different view of data quality (Lee & Strong, 2003). The Ministry of Health organisation wide team included information analysts from several departments, who were able to provide the Clinical Analysis Team with information on the usefulness of the data. The inclusion of clinical staff is paramount, it is difficult to judge the quality of data when users have no responsibility for data quality and when they are removed from the collectors (Tayi & Ballou, 1998). Black et al (Black et al., 2004) found data quality teams in the UK generally did consist of medical staff, but there were very few allied health professionals on such teams. Domain knowledge is also important when assessing the quality of the data and prioritising improvement. Croft and Williams (Croft & Williams, 2005) cite the NHS Audit Commission findings that clinician involvement in the validation and use of centrally held data is a key issue in data quality improvement in health care. Most of the New Zealand health care organisations appeared to have little clinical involvement in data quality work, with only one organisation wide data quality team having a position for a clinician, and this had not been filled. No teams had allied health professional input.

Those involved in training staff on data collection were members of the data quality teams. This enabled targeted training to appropriate areas or systems, providing direct feedback on training issues to the departments, and reviewing the success of training programmes. Significant support was required from the department managers, in particular to ensure staff attended training sessions.

Staff from the funding and planning department have considerable interest and input into data quality in many of the organisations studied. Two data quality teams had members from the funding and planning team, highlighting the significance of return on investment for data quality problem solving. Funding and planning departments and the data quality teams recognised that poor data quality was negatively impacting on funding claims through HealthPAC. Further, this department has significantly different uses for data than clinicians or administrators, with possibly differing data quality requirements. It may be possible, through the increased awareness and analysis of data quality on funding claims, to secure funding by providing management with information on return on investment of data quality work.
In summary, an organisation wide data quality team provides for a higher level of capability maturity in data quality across the whole of the organisation. Data quality is perceived as a ‘whole of system’ problem requiring ‘whole of system’ solutions, with members finding that data quality issues were similar across the organisation. Teams were able to engage management through education and data quality reports bringing increased management support for data quality work. Membership of the team assists in ensuring each department is aware of their role, with the CIO playing a pivotal role in communicating with executive management. Roles and responsibilities became more clearly defined, as the team members from diverse departments understood the impact of poor data quality in each department and the need for each department to have a role in the management of data quality. Data quality team members were often involved in training staff as they had considerable knowledge of persistent data quality issues. Funding and planning staff were also present in data quality teams, with different data needs from those of clinicians. Their presence reflected the importance of accurate data for reporting and claiming to obtain funding for services.

8.7.2 Accountability, Roles and Responsibilities

A significant impairment to managing data quality was the absence of accountability and roles for data quality in any of the organisations studied. Redman (Redman, 2001) feels the head of an organisation should be responsible for data quality and should not delegate this responsibility due to the implications of poor data. Loshin (Loshin, 2001) feels that no data quality programme will be successful without the support of senior managers due to the effect on the operational and strategic sides of the business. None of the New Zealand organisations had their Chief Executive Officer as accountable for data quality. The highest level of ‘self claimed’ responsibility was at the CIO level, however, there was no real accountability as data quality is not a requirement in any employment contracts. This is reflected in the low priority of data quality work, with no data quality projects initiating from the executive management teams.

The impact of having no accountability and no defined roles and responsibilities throughout the sector was evident. Individuals worked on data quality improvements to serve their own or their department’s requirements, with little thought to the impact of the organisation as a whole. This created silos of work, overlapping of work, and little cross organisation communication of data quality issues. Consequently, management were receiving reports with conflicting information.

Appropriate roles and their definitions are likely to differ between organisations due to their size and existing roles. However the DQIS seeks to highlight the absence of accountability in data quality and defines the roles of the Ministry, including the Health Information Standards Organisation, NZHIS and the Clinical Analysis team, collectors and suppliers of data to the national collections, and clinical coders in line with Loshin’s (Loshin, 2001) concept of a consortium ownership model, where the data collectors have ownership of the data and the whole of the sector is seen to have responsibility for the quality of data held in the national collections, with erroneous data to be corrected at source where possible.

The role of the Clinical Analysis team at NZHIS was reactive and not expected to be proactive, and their skills were under utilised. For example, the team were not included in the development phases of new
collections or major changes to existing collections despite their obvious expertise. The development of the Data Quality Principles for New Collections places the Clinical Analysis team in the path of the project manager, and outlines the organisation’s expectation that an extensive data quality plan will be developed with the assistance of the Clinical Analysis Team. Phases one and two of the DQIS provide the Clinical Analysis team with the tools required to implement preventative data quality initiatives, to undertake a leadership role in the health sector, and to further develop data quality knowledge and practices in the sector through the implementation and ongoing review of the DQIS. Ownership of the DQIS sits within the Health Information Strategy and policy team, but the onus of implementation sits with the Clinical Analysis Team as an operational team. NZHIS is responsible for promoting the DQIS and assisting health care organisations to implement phase three. The need for defined roles within health care organisations is highlighted in the DQIS, but organisations should define these specific roles themselves, as each organisation may have differing structure and need.

8.7.3 Best Practice Guidelines

The proposal in the DQIS to develop best practice guidelines is primarily to elicit ‘what works’ from stakeholders currently working with data quality. In keeping with the framework of CAS, current practice should be valued, particularly where good results have been shown (Plsek & Wilson, 2001). Many health care organisations already undertake considerable work in data quality and this knowledge and experience could be harnessed and disseminated throughout the health sector. Study participants expressed interest in the tools developed by the researcher, many noting that data quality had not previously been anyone’s role and therefore knowledge and skills around data quality practices are sometimes lacking in the organisation.

A working group made up of interested stakeholders from throughout the sector should develop the guidelines. This could include members of NZHIS, the Shared Services Agencies, DHBs, and PHOs. Through their collaborative development with appropriate sector wide consultation, it may be possible to have the guidelines endorsed as a national health standard through the Health Information Standards Organisation (HISO). This provides the sector with clear guidelines on data quality management, encouraging consistency throughout the sector. Software vendors, while not required, are very much encouraged to follow HISO standards, which may further embed the data quality standards throughout the health sector.

At present the guidelines cannot be developed through local empirical evidence of the best methods for data quality improvement due to the absence of formal assessments of data quality in health care organisations. However, anecdotal evidence, the experience of local data quality practitioners, and a review of international best practice may provide useful guidelines for the sector. The guidelines would need reviewing when more data quality information is available as data quality work matures within the New Zealand health sector.
8.8 A Metadata Repository for the Health Sector Including the National Data Dictionary

The Health Information Strategy for New Zealand proposes the development of a National Data Dictionary to address the standardisation of the major data elements found in the National Health Data Collections, providing a common language through which information could be shared electronically. This standardisation is likely to provide for considerable improvement to data quality through a nationally consistent understanding of data definitions. McGrath and More (McGrath & More, 2001) found that data standards development and implementation improved data and systems integration, supporting the aims of the Health Information Strategy. All collections at NZHIS have a data dictionary. Each data dictionary has a consistent layout, but data definitions between collections can vary. HealthPAC currently has no data dictionaries; the Screening Unit has dictionaries for both its collections but these do not correlate with the NZHIS data dictionaries.

While data dictionaries are simple data element documentation tools, metadata repositories manage the total information processing environment. Metadata are data that describe the properties of other data, or data about data (McFadden et al., 1999). Metadata repositories generally support the assorted data aspects of IT systems and include both technical and business metadata. Technical metadata generally support IT staff and users, while business metadata support an organisation's business users. Metadata repositories provide information necessary to evaluate and improve data quality, rather than measuring data quality themselves (Rothenburg, 1996).

Work is underway to develop national discovery level metadata frameworks in New Zealand government organisations. This is targeted at providing front-end access to metadata collections (Department of the Prime Minister and Cabinet, 2002). However, this development is still in its formative stage and, while Government does see the New Zealand Government Locator Service (NZGLS) as an emerging standard for government agency metadata collections, it is not apparent at this point how this will relate to data in the health sector (Ministry of Health, 2003). It seems logical that those involved in managing data quality in health care should work closely with the State Services Commission and other government departments in terms of recognising the impact of metadata standards on health data and to influence the development of these standards.

The researcher proposes that an effective way to manage data quality alongside the implementation of the Health Information Strategy for New Zealand is to develop a metadata repository that is available to all data users. Included should be information on what each data element represents, the coding standards or domains used, the source of the data, date of collection, and the quality levels of the data from the results of the assessment of the collection using the DQEF. Included also should be information on whether or not the standards used for the element are endorsed by the Health Information Standards Organisation.

The development of the National Data Dictionary requires that all stakeholders have access to the standards enclosed and a metadata repository would be a tool for providing such access. Olson (Olson, 2003) sees the metadata repository as the primary software tool for managing data quality, with the value of any data quality...
initiatives dependent on the maintenance of an accurate repository. As noted in Chapter One, data quality practitioners including English (English, 1999a), Wang et al (Wang et al., 2001), Olson (Olson, 2003) and Loshin (Loshin, 2001) agree that to make effective improvements to data quality the organisation needs to provide contextual information and clear standards on data definitions through a metadata repository that is maintained and available to all stakeholders. Problems can arise where users and practitioners are not informed of this context and assumptions on the quality of the data can be made without access to sufficient data quality information. It should be noted that data quality of metadata also needs to be managed using similar methodologies as those used in TDQM (Olson, 2003).

A metadata repository is included in the DQIS to be a project requirement for the Ministry of Health. The repository needs to be developed and managed centrally to ensure a national view of the repository contents. The Ministry of Health is required to develop minimum documentation requirements for each collection to ensure that the metadata repository contains adequate information for managing data quality. Metadata relating to all elements within the national health collections need to be available to all data suppliers, consumers and custodians to be effective. Further, the terminology needs to consider all users and to avoid terminology that is too technical.

The benefits of a metadata repository include increasing the longevity of the usefulness of the data. Data users are provided with metadata that enable decision making on data in the data warehouse. The policies and procedures under which these data were collected may have changed, but with sufficient data quality information the user is able to utilise data from historical collections. Further, while the data may be accurate, if users do not understand the meaning of the data or the context of their collection, their interpretation of the data may be inaccurate (Olson, 2003). As many data users become increasingly removed from any personal experience with data (Fisher et al., 2003), the importance of contextual information on their collection increases.

8.9 The National Health Index

The development and implementation of the action zones found in the Health Information Strategy for New Zealand (Ministry of Health, 2005) will lead to the increased ability to move data around the health sector. Data users will, in future, rely heavily on the presence and accuracy of the NHI to ensure correct identification of data for the correct patient, and to enable the linking of these data. Ministry of Health staff considered the characteristic ‘linkage’ to be the highest priority for improvement and this was reinforced by those in the health sector who noted the importance of the NHI as a unique identifier to enable the linking of disparate unit-level data. The 2003 CIO Survey (Ministry of Health, 2003a) highlighted the move towards electronic discharge summaries, electronic referral and integration with Primary Care. All these initiatives require the accurate identification of the patients’ data through the NHI.

Ministry of Health staff felt that the NHI Register should be assessed using the DQEF at an individual element level. Micro (as opposed to macro) assessment of the NHI means assessing all data elements, each used to identify patients when searching for an already existing NHI number. Price and Shanks (Price & Shanks, 2005) note that the level of granularity of assessment is a common discussion held when
implementing data quality assessment tools and should be documented in the business processes. The importance of the quality of the NHI is reflected in the resources allocated to maintaining this quality in all types of organisations. Duplicate NHI numbers are a significant and increasing problem with the register. Improving the data quality of all the elements in the NHI allows for the reduction of duplicate NHI numbers for individuals through better matching of searches. The Health Practitioner Index (HPI) is a new register to uniquely identify health care practitioners and provider organisations. The NZHIS should require the same level of assessment as the NHI to prevent the introduction of duplicates and this should be documented in the HPI Data Quality Plan.

The NHI acts as the ‘primary key’ in the management of relational data. Many national health care sectors do not currently have national unique identifiers, only regional or federal. Their value has been recognised, with many countries working towards the implementation of unique identifiers. Decision makers in health care find the ability to link health care data provides rich information, previously unavailable.

Lee, Conley and Preikschat (Lee, Conely, & Preikschat, 2000) identified fragmented information systems as a major problem world wide in the health care industry. The cooperative use of data where systems are diverse and distributed but work together to share information, is somewhat facilitated through the presence of a unique identifier. This provides for the ‘tracking’ of a patient throughout their health event, between primary and secondary providers, and referrals to specialist consultants; and reflects and supports the changing nature of health care where the care is provided on a continuum of the patient’s life, rather than event based. Data mining is enabled, where data analysis can discover meaningful new correlations, hidden patterns and relationships, providing for knowledge discovery with possible implications on national policy decisions. Care needs to be taken, however, that the use of patient data is within the requirements of legislation around health information, in particular the Privacy Act 1993, with the patient giving informed consent, when the data are collected, for its use by other providers and researchers.

It is not known what level of quality is required for the NHI to remain ‘fit for use’ and this would be difficult to ascertain. Despite increasing duplicates, with DHBs in 2004 creating duplicate NHI’s for 5.6% of the total new NHI’s created and with, on average, 704 per month remaining unresolved, the NHI is highly valued and utilised. It is envisaged the NHI will provide a ‘superset’ of the New Zealand population from which specialised systems could access specific views, or subsets, of the population. A subset could be by disease, age, sex, ethnicity etc, for the analysis of population health programmes. The NHI will provide health care services utilisation predicting for particular populations in particular geographical regions, and targeted health promotion and prevention activities.

The increasing reliance on the NHI and the future reliance on the HPI for unique identification, preventing the creation of duplicates needs to be actively managed by the data custodian if the registers are to remain effective. This would be assisted through the TDQM approach utilising root cause analysis and process management. All these methods are time consuming and resource intensive on the surface, but the evidence of this research is that management of the NHI throughout the sector is already using considerable resources, to correct bad data, that would be better channelled towards the prevention of duplicates.
8.10 Skills of the Data Quality Practitioner

Whilst the staff within NZHIS did have the requisite skills to complete the DQEF, management recognised the need for a senior position within the Clinical Analysis team to provide ongoing support to the strategic development and implementation of a data quality programme. A position titled Senior Advisor, Data Quality was established, with the main task of implementing the DQEF and DQIS. A review of the literature and position descriptions from other organisations informed the researcher and management on the knowledge and experience required to undertake such a role. Chung, Fisher and Wang (Chung, Fisher, & Wang, 2002) and Fletcher (Fletcher, 2004) found that both academic researchers and executives highlighted the interpretive capabilities of a data quality practitioner as essential. Interpretive capabilities enable the practitioner to understand how data quality affects both formal and informal organisational structure and decision making. The practitioner then needs to be able to articulate impacts to the management team. Adaptive capabilities are also important, where the practitioner is able to identify user requirements and measure user satisfaction (Chung et al., 2002).

Pierce (Pierce, 2003) analysed internet posted job descriptions for data quality analyst positions. Knowledge of statistics and data analysis, understanding of relational databases, and conceptual and analytical skills were the most common attributes found in the job advertisements. Most advertisements required people with prior experience in the applicable industry or domain. Thornton, in (Karr et al., 2000) highlights the need for domain knowledge required to detect outliers. Lee and Strong (Lee & Strong, 2003) also discuss domain knowledge, described as knowing why, whereby organisational members are more likely to raise more meaningful enquires if they understand the reasons for collecting and storing specific data. It is possible that a qualification in health information management would be of particular advantage in the management of health care data quality. Lee and Strong (Lee & Strong, 2003) do note, however, that the data quality custodian with domain knowledge does not lead to better data quality; therefore this is not a skill required. The research appears inconclusive in this area.

At present there are few specifically trained data quality practitioners worldwide. The Massachusetts Institute of Technology (MIT) has introduced a competency programme over the last three years, aimed mainly at those already working in the area of data quality, with incremental certification as a data quality practitioner through three week long courses. Pierce (Pierce, 2004b) notes there is as yet no research to understand the impact on data quality outcomes of having data quality practitioners with specialised training.

The skill set for such a role at NZHIS required someone with:

- knowledge of data analysis, ideally in the domain of health care;
- leadership and education skills;
- experience in project management;
- knowledge of applied statistics;
- understanding of information management practices;
- the ability to work in a complex political environment.
NZHIS did not attempt to find someone who also had knowledge and experience in quality management practices, as it was considered highly unlikely that such a person would be found. This is consistent with the findings of Pierce, who noted few job advertisements included such requirements (Pierce, 2003). Pierce (Pierce, 2004b) also notes the importance of the placement of the data quality practitioner in the organisation. The placement of the Data Quality role at NZHIS as equivalent to that of Senior Advisor allowed for a position that had reasonable access to, and influence on, middle and upper management. The role encompasses an expectation of strategic and forward thinking not found in operational data quality roles, but is within the operational team to ensure any proposed improvement aligns with the operational realities of data quality management. Olson (Olson, 2003) sees the need for a dedicated data quality professional to make data quality improvement effective. Assigning data quality to staff who have other priorities is unlikely to be effective.

8.11 Training

English (English, 1999a) provides an analysis of how data quality improvement relates directly to Deming’s 14 points of quality (Deming, 1982). Two of these points relate to training: training is required for both management and staff, and training needs to be continuous. NZHIS, as the sector leader in data quality management, has a responsibility to provide training in data quality methodologies, as well as on the requirements for data suppliers to the national collections. The positive feedback provided to the researcher by those in the sector who undertook the ‘train the trainers’ programme for the collection of ethnicity data, with an accompanying training manual provided to organisations, shows evidence of the willingness of those working with data in the sector to follow national data requirements. The lack of clarity around some requirements for data on the national collections has introduced some data quality issues that could be solved through better training and communication by the collection custodians. Approval was not given for an extra staff member at NZHIS to work solely on the education of health sector trainers, when requested as part of the NHI Upgrade programme and the implementation of the DQIS. However, a review of current positions within the Clinical Analysis Team is included in the DQIS for phase two of the improvement programme, as study participants felt this allowed time for improved data quality capability maturity in the organisation before making resourcing changes.

All of the organisations studied undertook some training towards ensuring good data quality, but most felt that more training would be an effective way to improve data quality. One study participant's response provides an indicative summary of the frustrations found throughout the sector when attempting to deliver training:

So each time an issue is flagged, we then go and arrange the training with the user, if it’s turned down it gets flagged and we say well we tried but it didn’t happen so if it happens again we go to their manager. We have a full time trainer now but in an organisation of this size if you don’t have focus then you are only ever going to deliver a half hearted training approach. I tried to push training up to another level but that hasn’t succeeded.

Competition for limited resources has meant that the amount of training provided is often less than was thought to be required by data quality practitioners. Much of the training appears to focus on teaching data collectors ‘how to collect the data’, providing little information on the importance of the data and possible
impacts of poor data quality. Kmietowicz (Kmietowicz, 2004) and Lee and Strong (Lee & Strong, 2003) found the understanding of data collectors on the impacts of data quality has a significant impact on the quality of data collected, with a need for data collectors to know what, how, and why data are collected. One DHB, with a relatively high level of data quality maturity has recognised that understanding by data collectors increases their interest in the quality of the data:

*The training focuses on where their data goes, as people don’t realise how important their data is. We are really bad at sharing information about this and so they don’t really care about the data. Data collectors are very receptive, they just had no idea the information even went to the Ministry.*

The evidence from this research suggests that the content of data quality training should be targeted to specific audiences, depending on their role in the data management chain, and this is supported by English (English, 1999a). Training on the theory of data quality management or the impact of poor data quality was only evident in one organisation, which provided some information to management. Feedback as a method of training may be appropriate to data entry and collection staff as they are then able to see where their performance needs improving. This has been found to be effective in improving medical data quality (Lorenzoni et al., 1999).

Clinical coders undergo a course to qualify as a clinical coder and are required to undertake refresher courses. In New Zealand this has become difficult, as training is no longer provided and those wanting training have to go to Australia. Internal training is provided within health care organisations, but the current shortage of clinical coders may mean that it is no longer possible for organisations to provide sufficient training to ensure good data quality. Clinical coding is highlighted in the DQIS as requiring a high level of data quality. There is a strategy under development at NZHIS that specifically addresses the management of clinical coding in New Zealand and this strategy aligns with the DQIS.

The assessment of the impact and effectiveness of training relies on the auditing of data related to specific data collectors or data elements in some organisations, therefore no assessments appear to be made on the quality of the process of data collection or management. None of the study participants had undertaken any formal training or education programmes specifically in data quality work and the trainers of data collectors did not have formal educational qualifications. Little empirical evidence is available in the literature to ascertain the most effective methods for data quality training, with feedback loops being the exception. Much of the training given in New Zealand organisations was not based on any proven methodologies, but those providing training appeared to have a very good understanding of the organisation’s data quality issues due to their other roles within the organisation. It is evident from the study that considerable education is required for the effective collection and management of data, but there is currently no evaluation of this training.

Study participants noted staff incentives for improvement were lacking, again related to the lack of responsibility for data quality in job descriptions. Including data quality levels in staff performance measurements puts an onus on the organisation to provide adequate staff training.
8.12 Feedback Loops

The ‘feedback loop’ is extremely important to improving the quality of data from suppliers. Orr’s (Orr, 1998) work on data quality and systems theory has the recurring theme of the need for continual feedback from users to ensure that data quality is maintained. Providing a formal feedback process for all users across the organisation through the data quality team provides an overview of users’ issues, allowing for appropriate prioritisation of data quality work. There is considerable evidence in the literature, particularly in health care, to support the value of feedback loops in improving data quality (De Lusignan et al., 2002; Fine et al., 2003; Gribben et al., 2001; Lorenzoni et al., 1999; Porcheret, Hughes, Evans, Jordan, Whitehurst, Ogden, & Croft, 2004).

All data in the national health collections are supplied from disparate collections, across a wide range of health care providers. An analysis of interviews held with data suppliers highlighted the current problems with information feedback from NZHIS, but also noted the value where they existed and were actively managed. The processes within NZHIS were reviewed to ensure the feedback was reaching the appropriate person or group within the health care organisation and the current data quality reports and their contents were appropriate and useful. A template was developed that ensured information on the content of the data quality report was provided to report users each time a report was sent, in case the usual recipient did not receive the report. Process improvements are required within NZHIS to understand the actions taken by organisations once the reports are received and if any support is required from NZHIS to action the improvements required in the health care organisation, as currently little follow-up on data quality reports occurs.

Feedback was one of the primary methods found for improving data quality between NZHIS and the health sector organisations, and also heavily utilised within organisations to feedback issues to departments. Considerable feedback is supplied to larger organisations that create NHI numbers to assist with duplicate resolution, and these reports were heavily utilised by the DHBs. Many had full time staff allocated to search for NHI duplicates, so the reports usually reached the intended recipients, as there were clearly defined roles. The support from the data quality literature on the benefits of feedback loops, the appreciative enquiry theory of Fry (Fry, 2002), and the study participants responses indicate that feedback is an area where the health sector is ‘doing what works’ and should be encouraged to continue with current practices, with some improvement for more formalised processes.

8.13 Technology and Data Quality

The underfunding of technology in the New Zealand health sector is evident from the CIO Survey undertaken in 2003, with many of the patient management systems reported to be over 10 years old, and two being 16 years old (Ministry of Health, 2003a). This is common in health care internationally, with other industries using information technology far more extensively (Bates, 2002). This is despite the evidence that some IT applications lead to improvements in safety, effectiveness, patient centred care (Ball & Douglas, 2002). The quality and reliability of information systems in health care is still inferior and the presence of internally developed software means that sharing information can be difficult. DHB CIOs reported feeling that their
budget was one of the constraints in developing new systems, along with lack of skilled staff, hindered again by budget constraints (Ministry of Health, 2003a). The rapidly changing environment, requiring the integration of disparate systems and the increasing requirements to support health care providers through better information has placed considerable pressure on the CIO and IS departments, and New Zealand health software vendors. Funding and planning requirements have driven improvements in information technology in health care in New Zealand in recent years to support health sector restructuring. CIOs are now beginning to move investment towards clinical systems (Ministry of Health, 2003a).

Interestingly, technological skills were considered the least necessary in data quality practitioners (Chung et al., 2002). The strategic management of technology is required to ensure optimum data quality. Technology can enable an organisation to materially alter the nature of its processes, such as enabling data collection through a PDA from a community nurse out in the field to improve the timeliness of the data and removing the need to use paper records that are later transcribed into an information system. Analysis of current processes should be made when considering applying technology to process improvement. If existing problems in processes remain, the technology will not significantly improve the process and may even make the process more error prone (Cunningham, 2001). It is also harder to fix flawed processes after the implementation of new technology, since the new technology supported process may now have an additional source of complexity hidden within the technology. Data quality problems can be hidden by technology that collects large amounts of apparently correct data.

The Health Information Strategy for New Zealand provides the strategic direction for New Zealand to increasingly use technology to assist with the management of health care data, and this is discussed below.

8.14 The Implications for the Health Information Strategy for New Zealand

The Health Information Strategy for New Zealand (HIS NZ) (Ministry of Health, 2005) is currently being implemented. The strategy notes the increasing complexity and cost of delivering health care services with 7-10% of the Gross Domestic Product of New Zealand currently being spent on health care. There is a demand for better quality of care from providers, funders and patients. The focus is now on population health, wellness, and illness prevention through increased utilisation of primary care services to prevent the need for hospitalisation and prevent chronic disease. The HIS NZ provides a link between the national health strategies and the data collected or not collected in the health sector. This link of information requirements to the organisation’s strategy needs to be made at all levels of the health sector and to all relevant strategies. The management of data quality needs to align with the business strategy and the ISSPs of an organisation. At present many of the ISSPs in health care organisations do not explicitly discuss the strategic management of data quality and there is little documentation of the operational management of data quality. The ISSP may be the best document to provide the link with the information and therefore data quality requirements of the whole of the organisation and its development should be done in conjunction with all of the organisation’s departments. This improves the capability maturity of New Zealand health care data quality work. The organisation is provided with all the information required to make appropriate strategic
decisions as defined by the strategic direction of the organisation, at a level of management sufficient to make these decisions.

HIS NZ recognises the importance of quality information in the management and delivery of health services and is focused on the connectivity and movement of health data throughout the health sector. National connectivity supports integrated care and the implementation of a distributed electronic health record (EHR). Local systems will hold unit level data; regional systems will consolidate those data, feeding eventually into the national systems, as shown in Figure 18 below. Data are held at varying levels of granularity according to the decisions required on those data. This requires the flow of data through connectivity, shared standards, extending national collections to data not currently collected, and the better use and sharing of information. The development of the EHR, as proposed in HIS NZ, needs to consider data quality throughout the development and implementation phases, building data quality into the technology and processes (Orfanidis et al., 2004).

Figure 18: New Zealand’s Distributed Electronic Health Records (EHR) Model (Ministry of Health, 2005)

One of the first tasks in the Health Information Strategy is to review all of the national health data collections. A review of all collections at one time is likely to highlight the anomalies between the collections in their management, stewardship, access and subsequent data quality. Definitions of data quality clearly stress that data requirements are derived from the intended use of the data. Thus the intended use of the collection data determines the necessary properties of the data (Arts et al., 2002). Some of the national collections were developed more than 20 or 30 years ago when the structure and function of the health service was considerably different. Many subsequent restructures of the health service have meant that national collections were developed under many different health care structures and the use of their data has evolved with time. For example, the national health index (NHI) was initially designed under a completely centralised health service, and was a unique identifier for hospital patients only. The NHI is now used for all patients for all health encounters, including primary care, increasing the complexity of the management of the register.
There may now be duplication and redundant data collection that can be rationalised through an ‘all of systems’ review. The management of the national collections needs to take into account the original use of that collection and consider its usefulness in today’s environment and if changes to the collection will improve its utilisation and subsequent data quality.

The Health Information Strategy specifically discusses the DQIS in relation to the NHI, as a requirement on every health event record. The quality of the data held in other national health collections is not discussed. The 12 action zones found in the HIS NZ are currently undergoing an assessment of the scope of work required to develop and implement each initiative. This would be the stage at which data quality should be considered as an integral area of each action zone, to ensure sufficient funding and human resources are available for data quality work. Increasing the sharing and access of information is likely to improve the quality of data, as the quality of data required will be that of the most stringent user (Orr, 1998). If data are not used over time real world changes will not be reflected in the data collection. The use of the data by many different users will highlight inconsistencies in standards that may prevent the integration of disparate data or require mapping. Much of the work proposed in the HIS NZ will ultimately lead to improved data quality, however, specific consideration of data quality is required to fully utilise the valuable data held by NZHIS on behalf of the sector.

8.14.1 Stewardship and Ownership of Data

While there is consistency in custodianship through NZHIS management of the national collections, various stewardship arrangements exist for the different national collections. The current stewardship arrangements for national collections tend to be characterised by some or all of the following attributes:

- existing groups are set up for individual collections;
- responsibilities and accountabilities of these groups are not always clear;
- no group that looks at overall requirements for collections or issues across collections;
- NZHIS has assumed stewardship responsibility in cases where no structures exist (body corporate function), and where it is the custodian; and
- sector and consumer interests are represented in different ways across collections, or not at all (Ministry of Health, 2004).

The Ministry of Health Draft Stewardship Framework (Ministry of Health, 2004) suggests requirements for audit and assessment of data should include a definition of what is acceptable data quality within a national collection, which organisations are responsible for maintaining the quality levels at different stages in the data flow, and the mechanisms for assessing quality and reporting on it. This framework remains in draft form under consultation. The implementation of the stewardship policy alongside the recommendations made in HIS NZ, would provide the sector with much clearer lines of accountability around data quality and the management of national collections data, as well as input into the decisions made around the collection through the governance framework. The implementation of the Health Information Strategy Action Committee (HISAC) to manage the implementation of the Health Information Strategy for New Zealand (Ministry of Health, 2005) provides a structure for the implementation of a governance group for data
stewardship. The Stewardship Framework will enable consistent and better management of data quality on the national collections through documentation on clear specifications for each collection. The proposed specifications are provided in Table 52 below.

<table>
<thead>
<tr>
<th>Specification Components</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reason Purpose for of the Collection</strong></td>
</tr>
<tr>
<td>This defines the purpose of the collection: what it is to be used for, why it is to be used in that way, what health and/or disability strategies are the drivers for the collection (including the Health Information Strategy for NZ, any legislative and/or contractual requirements).</td>
</tr>
<tr>
<td><strong>Coverage of the National Collection</strong></td>
</tr>
<tr>
<td>Description of the population groups targeted by the collection and confirmation that the geographic coverage is nationwide.</td>
</tr>
<tr>
<td><strong>Stakeholders of the National Collection</strong></td>
</tr>
<tr>
<td>Identification of the stakeholders whose activities will be supported by the collection. Description of the key information needs and reporting requirements of the stakeholders and how that information will benefit the stakeholders’ activities. Identification of stakeholders who will be the source of data for the collection, and the relationship of the data collection to their activities.</td>
</tr>
<tr>
<td><strong>Composition of the National Collection</strong></td>
</tr>
<tr>
<td>Description of the data elements that comprise the collection, with an associated data dictionary.</td>
</tr>
<tr>
<td><strong>Access, Privacy &amp; Security Requirements</strong></td>
</tr>
<tr>
<td>Description of how the stakeholders will access the national collection, and how these access arrangements will protect the privacy of the people who provided the data in the first place. Description of security controls (including authenticating the person seeking access to the national collection) and how the effectiveness of the controls will be monitored.</td>
</tr>
<tr>
<td><strong>Key Process Flows</strong></td>
</tr>
<tr>
<td>Description (including process maps) of the process flows for collection, production and maintenance of data. This includes the timeliness required from this process as well as the process itself.</td>
</tr>
<tr>
<td><strong>Business Rules &amp; Data definitions</strong></td>
</tr>
<tr>
<td>Definition of this defines the business rules applied to the preparation of data which also validate and transform data received.</td>
</tr>
<tr>
<td><strong>Quality &amp; Audit Requirements &amp; Quality Assessment</strong></td>
</tr>
<tr>
<td>Definition of what is acceptable data quality within the national collection (including completeness, accuracy, integrity, consistency and validity). Definition of which organisations are responsible for maintaining the quality levels at different stages in the data flow. Description of the mechanisms for assessing quality and reporting on it.</td>
</tr>
<tr>
<td><strong>Sustainability</strong></td>
</tr>
<tr>
<td>Description of how the national collection will be maintained to enable it to survive 20+ years, the approach to making the collection sustainable long term and identifies responsibility and liability of different elements. This includes a description of the stewardship and custodianship arrangements. It also includes a description of the operational funding for the collection.</td>
</tr>
<tr>
<td><strong>References to Standards</strong></td>
</tr>
<tr>
<td>Identification of the standards relevant to elements of the collection.</td>
</tr>
<tr>
<td><strong>Future Requirements</strong></td>
</tr>
<tr>
<td>Request for additional functionality planned for the collection.</td>
</tr>
<tr>
<td><strong>Collection Awareness/Communication</strong></td>
</tr>
<tr>
<td>Definition of the level of awareness and communication to be maintained across stakeholder groups.</td>
</tr>
</tbody>
</table>

Table 52: Specification Components for New National Health Data Collections (Ministry of Health, 2004)
8.14.2 Access to National Health Collection Data

As custodians of the national collections, NZHIS has the role of managing access to the national health data collections through the management of national access, security and usage policies (Loshin, 2001). Many of the study participants noted the limited access to the national health data collections as being a significant impairment to service planning and delivery. Access to information held in the national health collections remains an issue in the New Zealand health sector at all levels. Detmer (Detmer, 2003) feels that most of health care’s avoidable shortcomings can be linked to data, information or knowledge that are inaccessible or inaccurate. Detmer (Detmer, 2003) also found that some organisations have demonstrated that improved access to knowledge for clinicians improves the quality and/or safety and efficiency of health care services. This is also discussed in the Draft Stewardship Framework (Ministry of Health, 2004), finding that the national collections are valued as a resource to the Health Sector, although under utilised, and support exists for the improvements in the stewardship of national collections to formalise the wider involvement of the sector in national collections. Improving access to the national health collections is one of the 12 action zones in the Health Information Strategy for New Zealand. This will be enabled through improved technical infrastructure and policy changes such as the Stewardship Framework (Ministry of Health, 2005).

Access to more data held throughout the health sector, but not currently available nationally is also addressed through the implementation of new national collections. Participants stated, however, that GPs sometimes consider the data they collect from patients are ‘owned’ by the practice. At present some GPs ‘sell’ these data to their PHO for analytical purposes. With the implementation of a national EHR and the implementation of a National Primary Care Dataset, this perception of ownership is likely to be challenged. At present patients are not explicitly informed that their health data are on-sold within their own PHO, and this needs to be addressed.

8.15 Chapter Summary

This chapter provided a discussion of the research results found to be of significance to the New Zealand health sector. The researcher applied change management and systems theory to increase understanding of strategic data quality management in the New Zealand health sector. Action research provided the researcher with a guide for the research process, applicable to the complex adaptive system of the Ministry of Health. The research methodology could not be highly structured, as the researcher was a member of the organisation’s staff, needing to abide by management requirements for other work priorities. However, the benefits to the research of working within such a complex environment far outweighed the sometimes slow progress of the research. The action research process encouraged a team approach to learning and problem solving, thought to be highly effective in complex adaptive systems when the team is from across the organisation’s departments.

Change management was required to encourage the Clinical Analysis team to embrace the paradigm shift required in their management of data quality to a more strategic and holistic methodology. The researcher utilised the theory of appreciative enquiry (Fry, 2002) to positively encourage change, and utilise existing knowledge within the team. Simple rules, such as the TDQM process and the data quality dimensions guided
the change, leaving room for innovation. The theory of ‘complex systems of adjustment’ (Champagne, 2002; Stacey, 1993) can be instilled in the organisation to encourage change through the constant interaction of people throughout the organisation.

The impact of the presence of an organisation wide data quality team is considerable. The research emphasises that those organisations with an established data quality team were more mature in their management of data quality. Management were more informed and supportive of data quality work, and roles and responsibilities were clearer, with departments beginning to understand their responsibilities that brought about improved attitudes towards data quality throughout the organisation. Membership of a data quality team is important and should include those who use data as well as those who manage data. The DQIS requires first that health care organisations put together an organisation wide data quality team, as this appears to be a relatively simple but highly effective way to improve data quality.

Roles and responsibilities for data quality are unclear throughout the health sector, including the stewardship arrangements for the national collections, the role of the Ministry and roles within health care organisations. This impacted considerably on the quality of data. Within health care organisations staff worked in siloed departments, fixing data for their own use without considering the rest of the organisation. This led to confusion at management level as conflicting reports were provided, with wasted resources due to overlaps in work. There are no senior managers in New Zealand health care organisations who have data quality as a measured job performance requirement. The DQIS clarifies, at a high level, the roles of the Ministry of Health and the sector around data quality with the aim being to highlight to health care organisation managers the requirement to consider introducing data quality into staff performance measurements and to take overall responsibility. The skills of the data quality practitioner should include interpretive capabilities so the practitioner understands how data quality affects both formal and informal organisational structure and decision making.

Training in data quality is extensive throughout the health sector. However, most organisations felt that more training would improve data quality. The research evidence from this study suggests that the content of data quality training should be targeted to specific audiences and this is supported by English (English, 1999a). No formal assessments appear to be made on the quality of the process of data collection or management. Little empirical evidence is available in the literature to ascertain the most effective methods for data quality training, with feedback loops being the exception. Feedback loops are extensively and successfully used between NZHIS and health care organisations as well as within health care organisations. The use of these feedback loops should be encouraged, with some increased formalisation of processes, as this method appears to assist data quality practitioners in their role.

The HIS NZ action zones are likely to have considerable impact on the quality of data held in the national collections with the increasing access and use of data which will improve its quality. The review of the national collections is likely to highlight the disparities in management between collections and the changing uses of historical collections. The development and implementation of a Metadata Repository, based on the proposed National Data Dictionary will provide all stakeholders with the information on each collection, including data quality information, on which to base decisions on data ‘fitness for use’.
The DQEF and the DQIS provide the clear direction for a holistic and ‘whole of health sector’ way of viewing data quality, with the ability for organisations to develop and implement local innovations through locally developed strategies and data quality improvement programmes. Clearer governance of data through the implementation of the HIS NZ governance structure will help to define the roles throughout the sector more clearly. The process of developing Best Practice Guidelines, as proposed in the DQIS, will encourage the dissemination of methods of data quality work that organisations throughout the health sector have found are effective. This further reinforces the notion of encouraging change but keeping ‘what works’.
Part D Conclusions and Future Work
9 Improving Data Quality in the New Zealand Health Sector

The NZHIS is the custodian of data collections that hold information on the health of all New Zealanders, an increasingly important role with ethical and political responsibilities for its management (Ryan, 2004). Indeed, many in the New Zealand health sector refer to the data held at NZHIS as ‘a national treasure’. As noted in Chapter four, for Māori, data are embodied with significant spiritual and cultural significance. Kamira (Kamira, 2003) notes it is possible to apply the concept of **kaitiakitanga**, meaning guardianship, protection, care and vigilance to data. This approach offers the potential to locate Māori as the primary beneficiaries of data relating to themselves, to ensure that data collected are appropriate and relevant and that issues of collective ownership and collective privacy are addressed (Kamira, 2003). The researcher found through this research, that not only Māori in New Zealand have strong opinions about the care and protection of their health care data. This is reflected in part through the inclusion of the privacy and security dimension in the data quality evaluation framework.

Van Gool, Lancaster, Viney, Hall and Haywood (van Gool, Lancsar, Viney, Hall, & Haywood, 2002) note that without adequate information it is difficult to determine the success or failure of health polices. The national health collections at NZHIS are used in the development and evaluation of national health policy. The quality of these policy decisions is reliant to an extent on the quality and nature of the data held in national collections. There is a significant trend towards evidence based medicine in New Zealand health care, and the philosophy is translated into decision making for policy development. Evidenced based policy requires an investment in health information for the ongoing maintenance of existing collections, the development of new collections to address the significant areas of incomplete data, and the development of the skills of information analysts and researchers to glean the best information from that data (van Gool et al., 2002). NZHIS provides data to international organisations and Murray et al (Murray et al., 2004) highlight the need for improved global health monitoring through the availability of better health information, requiring global standards for core health reporting. Enhanced global reporting will increase government commitment to collect high quality data (Murray et al., 2004).

Many stakeholders are not aware of the information that is already available in the national collections. Black et al (Black et al., 2004) found the under utilisation of valuable clinical data to be a common phenomenon. Black et al (Black et al., 2004) also note the considerable effort and resources that are required to collect and maintain such data, with many database custodians working in isolation of support from similar colleagues. Despite this, many respondents in this study noted the need for more national health data collections for particular data, evidencing the value the data held in national collections provides to stakeholders. The implementation of HIS NZ makes strategic data quality management imperative. The overarching aim of the strategy is to collect more data and to move data throughout the sector, in line with international trends to implement electronic health records.

Michael Porter, the renowned MIT strategist, and colleagues, support the notion that more and better information is required to strategically manage health care, along with providing patients access to their own health information to enable them to make decisions about their care (Porter & Olmstead Teisberg, 2004).
Providing direct access to personal health information has been found to improve data quality, particularly with medical histories (Porter & Mandl, 1999). Black et al (Black et al., 2004) found that data collections where the patient is informed of the collection and use of their data were more likely to contain complete data. At present there is little public education on the importance of collecting health care data in New Zealand (Ryan, 2004), perhaps reflecting the political opinion that highlighting the collection and use of health data will bring misunderstanding and apprehension in the general public. Ryan’s (Ryan, 2004) study of the opinions of New Zealanders towards electronic health records found that many participants had questions, that if answered, could dispel many of the current fears. None of the participants in Ryan’s study were aware of the existence of national health collections. It is likely in the future that managing the quality of health care data will be increasingly difficult without the participation of the patient, due to the increasing complexity of health care delivery and the increasing collection and use of data by disparate health care providers and planners. The UK NHS also notes the need for significant public education (NHS Information Authority, 2003).

The sharing of information by disparate users requires a shared and consistent understanding of the data quality conceptual framework used by each organisation providing shared data. Lorence and Jameson (Lorence & Jameson, 2001), in studying data quality in US health care organisations, found that inconsistencies in the regional and national management of data quality could in itself be a quality issue as it precludes comparative assessments or benchmarking. The health sector, without a consistent view of data quality, is unable to say which sources of data are likely to provide data of appropriate quality. The implementation of a national data quality strategy in New Zealand is aimed at preventing this lack of standardisation so all data users can have an appropriate level of confidence in data they are using.

9.1 Action Research and Data Quality

This research, using action research methodology and a combination of interpretive and positivist data collection and analysis methods, provides the health informatics community with an understanding of the issues related to developing and implementing programmes to improve data quality. The applied discipline of health informatics requires research designed for the purpose of improving the effectiveness in actions, more so than natural science truth. Plummer (Plummer, 2001) sees health as a large, complex system that is highly political and culturally diverse and that health informatics research that improves outcomes and performance is required for applied practice development (Plummer, 2001).

This research was undertaken in the highly political environment of a government health department. Limitations were placed on the research due to this; external input into the strategy development was limited until executive managers from across the organisation had approved the draft consultation document contents. The study of data quality from within an organisation using action research requires that the researcher abide by the rules of the organisation. In this study the researcher was a member of staff, and this had an impact on the research process. The researcher found that she had little control over the timing of the research process and was often less able to be methodical, having little control over impacting work priorities. This differed from the previous experience of the researcher when undertaking qualitative research and sometimes proved frustrating. Even with high level support from management there were considerable
competing priorities on the researcher’s time. There were considerable delays in the research which impacted on the researcher’s thesis. However, the benefits of working within the organisation eventually outweighed the researcher’s frustrations as a practical and pragmatic data quality programme could be developed. Working within the complex adaptive system provides considerable insight into the real needs of the organisation. These could be translated into the health sector following extensive discussions with sector stakeholders. The researcher’s background in health care, having trained as a nurse, provided an understanding of the structure of the New Zealand health sector and domain knowledge when discussing health care issues. It is unlikely that an ‘outsider’ could develop such a depth of understanding in such a complex environment in a limited time frame. Further, the researcher had the full support of high level management, providing guidance as well as the resources required to undertake the research.

9.1.1 Action Research in the Complex Adaptive System

The process of action research appears to align well with the needs of the complex adaptive system, and provides a suitable mechanism for bringing about change in such an environment. Champagne (Champagne, 2002) notes the similarities between complexity and learning theories. Both see change in a global, integrated way as forming part of the routine life of organisations, with the change process a collective one. Action research provides the organisation with the means to learn through collaborative groups.

Baskerville and Myers (Baskerville & Myers, 2004) outline four key action research premises that arise from the pragmatist philosophy of asking the right questions to get the right answers. The first is that arising from Charles Pierce (1839-1914), arguing that all human concepts are defined by their consequences. In order to experiment, or to even determine purpose, people must have independence of will to decide what actions they could choose to undertake. This theory aligns with the notion that CAS posits of supporting the exploration of new possibilities through experimentation and autonomy.

The second premise is that of James (1842-1909), who found that truth is embodied in practical outcome. There must be practical action in the problem setting to reveal the relative truth-value of the theoretical concepts underlying the action (Baskerville & Myers, 2004). The concept of CAS provides wide boundaries for beginning the work of self organisation (Committee on Quality of Health Care in America, 2000) and that organisations learn by doing (Champagne, 2002).

The third key notion of pragmatism found in action research is Dewey’s (1859-1952) logic of controlled inquiry, whereby inquiry is the directed or controlled transformation of an indeterminate situation into a determinately unified one. This involves the conceptualisation of possible ways of resolution and practical experiments that change existing conditions with the pattern of inquiry requiring action (Baskerville & Myers, 2004). In the CAS agents respond to their environment using internalised ‘short lists of simple rules’ that drive action and generate behaviour (Eoyang and Berkas 1998; Plsek 2003). The rules need not be shared, explicit, or even logical when viewed by others, but they nonetheless contribute to patterns and bring coherence to behaviours in complex systems.
The fourth assumption is Mead’s (1862-1931) tenet that all human action is socially reflective and explains why collaborative teams are required in action research. This team provides the interaction required for the action to be rationalised and operationalised in the reality of the social world. The participants must be socially situated in the research environment. Action is socially relative, and this means the action researcher must be a participant observer. Champagne (Champagne, 2002) discusses the CAS theory finding that organisations are more able to be innovative and bring about change where there are many cross department groups working together, the change process is then a collective one, with managers as change agents alongside workers.

The effect of the organisation wide data quality team on the maturity capability of the health care organisation in their management of data quality is evidence of the applicability of both CAS theory and action research assumptions to bring about effective change and innovative learning in the discipline of data quality. The development process for the DQEF allowed for researcher and staff participation and learning, drawing on current change management theories of appreciative enquiry highlighted by Fry (Fry, 2002), to transition the change with the minimum period of diminished performance (Elrod, 2002). Innovation was encouraged through providing a voice to the experts who work on a daily basis with health care data within the organisation.

There is considerable theory available in disciplines outside of data quality to support the findings of this research that organisation wide teams enhance the innovation and learning of those who participate and through this improve the organisation as a whole. Encouraging staff to develop solutions themselves assists to institutionalise data quality in the organisation, through the development of emergent strategy. Data quality practitioners are empowered to implement ‘business as usual’ initiatives that they themselves believe work. This emergent strategy is guided by the national strategy that provides the simple rules that encourage the seeds of innovation, whilst maintaining a momentum towards a defined national vision for data quality management.

9.2 Institutionalising Data Quality in the New Zealand Health Sector

The process of developing the DQEF provided NZHIS and the Ministry of Health with considerable learning around data quality. NZHIS is in a good position to implement TDQM, with the knowledge to undertake such work. The implementation of the Senior Advisor Data Quality position within the Clinical Analysis team provides this team with the ability and expectation to view data quality in a strategic way. Further, the health sector requires guidance and leadership from the Ministry of Health if improvements are to be made. Other areas within the Ministry of Health have also improved their understanding of data quality, having voiced frustration through the ISSP Project and the Current State Analysis at the current levels of data quality.

The presentation of the DQEF at the annual national health informatics conference and at a DHB CIO Forum began to raise the profile of data quality in the health sector. Further, the research involved many participants from throughout the health sector who were eager to obtain the tools developed within the Ministry of Health for use in their own organisations.
At present NZHIS does not have a metadata repository that holds data quality information, is continually maintained, and available to all stakeholders. Without a metadata repository it will be difficult to manage data quality effectively in the health sector, as NZHIS staff will be the only stakeholders with an understanding of data quality requirements. If the whole of the sector is to take responsibility for data quality, then the whole of the sector needs access to data quality information about the national health collections.

Many data quality practitioners voiced a feeling of isolation in their work, even where an organisation wide data quality team was present. The establishment of a national forum for data quality in health care will provide the sector with support from colleagues and assist in the development of best practice guidelines. This research has shown that the establishment of organisation wide data quality teams within the DHBs and PHOs will provide considerable improvement in data quality capability maturity. Given the limited resources within health care organisations, the organisation wide team may prove to be the most cost effective way to manage data quality within these resources. The work undertaken by the team is likely to raise the profile of data quality with management and highlight areas of concern for the entire organisation, with subsequent funding for the implementation of the next phases of the DQIS.

### 9.2.1 Data Quality as a Standard

The development process for the DQEF involved collaborative work within the Ministry and confirmation of the data quality dimensions with stakeholders in the sector. This consultative approach to the development may mean that the data quality dimensions and components of the DQIS can be endorsed as a national standard by the Health Information Standards Organisation (HISO). HISO requires that standards follow strict guidelines for sector consultation before they will endorse the standard, to ensure the majority of the health sector agrees to the standard, further encouraging utilisation of the standard.

This endorsement has the benefit of providing established guidelines to the sector and software vendors. Standards endorsed by the HISO do not have to be used; there is currently no legislative requirement. However, health care organisations are likely to follow standards when implementing new systems and request that software vendors meet standards requirements. Encouraging vendors to build data quality into systems at the development phase assists system users to prevent poor quality data entering the system.

### 9.3 The Strengths of the Research

This research provides a methodology for practical and relevant research in the field of data quality by solving problems while expanding scientific knowledge. The research was able to bring about organisational change and learning through the development and implementation of the Data Quality Evaluation Framework. Perhaps the greatest limitation was the time available to the researcher, impacted in part by the role of the researcher as a staff member in the organisation. This meant that validation of the impact of implementing the DQEF on measured data quality improvement was not possible. The effect on data quality is likely to take a considerable amount of time, outside of that available to those completing PhD research.
The researcher was able to access many respondents from many different areas. Purposive sampling was used to ensure information was gained from many different areas outside of health care, across the Ministry of Health, throughout DHBs and PHOs, and from respondents attending a health informatics conference. This purposive sampling may have led to the inclusion of respondents that had a higher level of understanding and interest in data quality than most working in health care. Semi-structured interviews were held, whereby the researcher sought to gain further understanding of particular issues raised by previous interviewees. This may have limited the responses given by participants to fit with the researcher’s questions.

The DQEF has been piloted on one clinical data collection held in a DHB. While the evidence provided from this pilot shows that the DQEF is likely to be applicable throughout the health sector, further research is required in a range of health care organisations to ascertain if the framework is generic to data across the whole of the health sector. However, it was evident that the DQEF did highlight areas of known data quality problems, validating its applicability and usability where tested. Further work on the development of the business processes required to implement the DQEF as ‘business as usual’ work is required, would be specific to each organisation, and need to be agreed by management. These business processes should include decisions around the ongoing training of those using the DQEF, the frequency of its application to each collection, the next steps based on the outcome of evaluations, and ownership of data quality issues. The DQEF remains an iterative tool, whereby those that use the DQEF will have increasing knowledge and experience on which to base improvements to its content.

Further experience in using the DQEF and the outcome of better data quality information will allow for accurate development of data quality metrics. This could not be done within the time frames of the research, but accurate metrics are paramount to the success of the DQEF to provide effective data quality information and to provide data suppliers with accurate key performance indicators for expected levels of data quality. The research provided a beginning methodology for metrics development through the development of ethnicity data metrics, but the methodology will benefit from further refinement and validation through empirical testing.

The research was limited to the management of data quality on the national health data collections and to the public health sector. There are many other types of data collected throughout the health sector. The DQIS noted that health care organisations might wish to use the processes outlined in the DQIS on collections held internal to the organisation for consistency and ease of management. The DQEF and DQIS development processes included Ministry of Health departments outside of NZHIS who manage data not included in the national collections. It was recognised that consistent management of data quality across the Ministry of Health is required and the aim of including all departments was to begin an education process around data quality management and to inform the other departments of the processes to be used at NZHIS, as they are users of national health data. The development of the DQIS was aimed at the development of a high level, national strategy, and therefore does not consider operational aspects of data quality.

The research provided for the development of a national health data quality strategy, and the Ministry of Health Executive Team approved the content of the strategy. The implementation of the strategy in the
health sector was outside of the researcher’s responsibility. Further, the timeframe for implementation needed to coincide with the implementation of the Health Information Strategy for New Zealand. Therefore, the DQIS has not been tested through a post implementation evaluation process.

9.4 Generalisability to Other Countries or Domains Outside Health

Grounded theory is based on the social psychological theory of symbolic interaction which assumes that meanings are constantly evolving and do not remain static over time (Bowers, 1988). This research utilises grounded theory method (Bryant, 2002) to establish the truth for this environment. A parallel issue for grounded theory study would be to identify the conditions under which the theory is expected to work. As this research has not been replicated in a similar environment, it is not possible to ascertain these conditions. The use of grounded theory is supported where data from the participants and the researcher are included in the research results (English, 1999a). Baskerville and Wood-Harper (Baskerville & Wood-Harper, 1998) also note that action research is only relevant to the social setting in which the research took place as the same research in a different social setting may provide different results. This is due to the multivariate nature of human interaction.

The use of grounded theory in the analysis phase of action research provides validity to the methodology, as the researcher is embedded in the research process, encouraging the continual reflexive practice of the researcher. Grounded theory provided the researcher with a structured methodology for the analysis of large amounts of textual data. The results of this research are directly from the data to provide a theory grounded in the data, in particular with the development of the DQIS, where very little research has been published. The extensive interviews undertaken throughout the health sector and the careful analysis of the data provide results that are a true as possible reflection of the reality of data quality management in the New Zealand health sector and their requirements of a data quality strategy.

The types of data quality issues found in other national health care systems appears very similar to the New Zealand health sector and the learning provided from the Canadian Institute for Health Information was applicable and valuable to this research. It may be that this research provides guidance to those institutions that are tasked with the custodianship of national health data.

The data quality capability maturity assessment undertaken in this research that assessed both New Zealand health care organisations and international organisations using the Capability Maturity Grid indicates that health care has similar data quality issues to other organisations. However, the methods required to solve these problems in health care may require a more multidimensional approach due to the complex adaptive systems found in health care organisations and health sectors. Further, Atkinson, Eldabi, Paul and Pouloudi (Atkinson, Eldabi, Paul, & Pouloudi, 2001), in investigating health informatics methods, note the ‘special case’ for health informatics with unique demands. For example, governments provide the legislative and regulatory frameworks as well as the delivery of care. The structure of the health care system drives the shape of IS application developments, not the need to enhance profit or market share, where a predominantly public health service exists as it does in New Zealand. A dominating factor in health informatics is the particular social, professional and cultural context of health care. These drivers mean the
content of the DQEF and DQIS may be applicable in other public health care services, but not outside of this domain.

9.5 What Further Research is Required in Data Quality in Health Care?

Further research into the impact of the presence of an organisation wide data quality team, the ideal membership and function of the team, could provide further evidence for data quality practitioners that such a team is effective in improving the capability maturity of data quality work. A New Zealand wide data quality team, made up of representatives of health sector organisations and the Ministry of Health, may provide for more effective national management of data quality in the New Zealand health sector. This team reflects the movement of data throughout the sector and shared ownership and responsibility of the national health collections and it may be worthwhile to study the impact of such a team on the capability maturity of the New Zealand health sector to manage data quality. Respondents in the study indicated a national group would provide them with support and knowledge sharing capability, many finding work on data quality to be 'isolated and unsupported'. Further, a national group may provide support for the implementation of the DQIS.

Research providing the analysis methodology for return on investment of data quality projects in health care would assist data quality practitioners to gain funding and human resources for data quality projects. Many of the study participants noted the difficulty in ascertaining return on investment and therefore gaining funding. It may be possible to develop a generic methodology for the New Zealand health sector, as priorities throughout organisations may be similar.

Data quality metrics are often cited as requiring further research in the data quality literature and there is unlikely to be one formula that could be applied to all organisations. However, empirical research into the methodology for the development of applicable data quality metrics for the New Zealand national data collections may provide considerable insight for other data custodians of national health collections.

All of the large organisations that took part in this research provided some form of data quality training. The lack of empirical evidence to support the most effective methods for training in data quality means that the content of the training may not be as effective as it could be. It may be that existing educational theories provide sufficient guidance but research in this area is warranted as it has the potential to improve the cost effectiveness of training programmes.

Eppler and Wittig (Eppler & Wittig, 2000) note that tradeoffs between data quality dimensions are not commonly addressed in frameworks. Further research in this area may be warranted as the assessment of data collection potentially provides the data quality practitioner with a considerable list of possible improvements. Prioritising these improvements should be based on the needs of the organisation with associated guidelines included with the development of a framework.
As the DQIS is not yet implemented, it has not been possible to undertake a post implementation evaluation of the strategy. Indeed, this is likely to be most beneficial over a period of several years, given the slow nature of change across a whole health sector and the phased approach to the strategy, changes to data quality work as a result of the strategy may be some years away.

The health sector has some way to go before reaching second and third generation data quality management discussed by Redman (Redman 2001). Continuing to research the possibilities of preventing the introduction of poor quality data through process management could lead to a health sector where all data quality is managed through prevention rather than ‘find and fix’. However, given the complexity of the health sector, considerable research and ongoing resources would be required. Continuing research on preventative data quality for the national health collections would provide a focus but also learning that could be applied to other data collections.

9.6 Implications for Professional Practice

Work on data quality in health care in New Zealand is at present initiated and undertaken at the grass roots level of organisations. This research highlights that strategic direction is required in order to harness and encourage the innovation taking place within organisations. Data quality management should, in the future, concentrate on the prevention of errors and TDQM methods of management. Increased resources should not be required as staff who currently work on ‘find and fix’ work can be redirected to actively assess and improve data.

The data quality team within the health care organisation should have a pivotal role in the development and implementation of new systems, including ongoing training of data collection staff on the impact of errors and the downstream uses of the data. Further, the importance of data quality, and the staff who work in this area, needs to be recognised by management teams. Recruitment of staff should include those with the skills of strategic thinking and analysis. Providing such a team with a structured programme of data quality management, with clear roles and responsibilities for the organisation to manage their data from a whole of system view, will enable data quality practitioners to concentrate on improvement initiatives.

There are an increasing number of training programmes available that include a data quality component and a certification course in data quality is now available through MIT. The data quality practitioner of the future is likely to be able to be qualified in the field, however, it is noted that there is little interest at present from graduates for this field, with many coming to the discipline with a wide range of experience in many fields. A strategic view of data quality will draw on a wide range of skills that may mean these experienced practitioners are the most appropriate for the discipline at the strategic level.
9.7 Chapter Summary

This chapter provides the conclusions elicited from the action research process, the development of the DQEF and the DQIS. Table 53 below highlights how each of the seven research questions have been answered.

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Answered through Research Process</th>
</tr>
</thead>
</table>
| 1. What existing data quality theories assist in the decision making around defining and measuring data quality requirements in national health data collections? | • Literature review  
• Review of current CIHI work |
| 2. What are the data quality dimensions considered important to measure and improve by data quality consumers, collectors and custodians in the New Zealand health sector? | • Current state analysis of data quality in the New Zealand health sector including a Survey of data quality dimensions as proposed by NZHIS  
• Interviews with sector stakeholders  
• Comparison with organisations outside of health |
| 3. What initiatives assist to raise the profile of data quality improvement in the New Zealand health sector? | • Current state analysis of New Zealand health sector highlighting the impact of data quality teams, management support.  
• Ministry of Health data quality workshop |
| 4. What are the data quality roles found and by which stakeholders in the New Zealand health sector? | • Current state analysis of New Zealand health sector  
• Stakeholder analysis |
| 5. Does action research methodology provide appropriate research methodologies for understanding and change initiation in a data quality improvement programme? | • Framework development process – all of chapter 5  
• MOH data quality workshop |
| 6. What steps are required to initiate a structured data quality programme within NZHIS? | • ISSP review  
• Current state analysis of Ministry of health data quality management  
• Focus groups to review CIHI Framework and develop New Zealand dimensions  
• Review of Eppler and Wittig framework requirements against revised New Zealand framework  
• Pilot of the New Zealand framework at NZHIS and a DHB |
| 7. What are the components of a national health care data quality improvement strategy that initiates effective ‘institutionalisation’ of total data quality management across the New Zealand health sector? | • Strategy development  
• Tools developed to support strategy |

Table 53: Methods to Answer the Research Questions

Action research enabled the data quality practitioners to learn the data quality improvement needs of NZHIS, the Ministry of Health, and the health sector. The collaborative environment of action research encouraged the development of a framework that is applicable to the national health data collections and the Clinical Analysis team at NZHIS to have ownership of the DQEF implementation.

Action research also appeared applicable to the complex adaptive system found in health care organisations and across the whole of the health sector, where collaborative groups form to bring about change through shared learning. The four key premises of action research align well with the theory of CAS, in particular
Mead’s assumption that all human action is socially reflective and more able to be innovative in cross department groups. These groups go some way towards reducing the isolation expressed by data quality practitioners in the health sector, however a national group could also be effective.

Further developments around the implementation of the DQEF are required to support its ongoing and applicable use as an evaluation tool, including business processes for the frequency of evaluation of collections and the process following each evaluation. The DQEF is an iterative tool that can be incrementally improved once organisations have better data quality information. The DQIS provides the guidelines to the whole of the health sector on a way forward for data quality improvement. The detail is left to the organisations to develop and disseminate innovative solutions to common and persistent problems. Both the DQEF and the DQIS could be endorsed as a national standard by the Health Information Standards Organisation. The limited time available to the researcher restricted the scope of the research and the DQEF and DQIS would benefit from further research post-implementation.

This research provides results that may be applicable to other health sectors overseas, but the research has only been undertaken in the New Zealand environment. There are many similarities between health care systems, in particular with the need to manage national health data. The capability maturity of the New Zealand health care organisations, when compared to other organisations indicates that they are reasonably consistent with other organisations. Therefore, the results of this research may provide insight into the needs of many organisations.

The strategic management of data quality requires wide ranging skills from practitioners with experiences in many different areas. The challenge for the discipline of data quality is to encourage these experienced practitioners to the discipline and then to undertake further study in the field of data quality to ensure the application of skills towards the development of strategic programmes of data quality management.

This research began a journey towards improved maturity around data quality management in New Zealand health care, where TDQM is ‘business as usual’, institutionalised into the daily practices of all those who work in health care. The increasingly information intensive nature of health care demands a proactive and strategic approach to data quality to ensure the right information is available to the right person at the right time in the right format, all in consideration of the rights of the patient to have their health data protected and used in an ethical way.
Part E Appendices
Appendix 1: Data Quality Improvement Strategy Workshop Workbook

Ministry of Health

Data Quality Improvement Strategy Workshop

Tinakori Room Level 8, Deloitte House
61 Molesworth St
Wednesday 27th October
10 am to 3 pm

PLEASE BRING THIS BOOKLET WITH YOU TO THE WORKSHOP

212
Background

The Ministry is working towards developing and implementing a “Data Quality Improvement Strategy” (DQIS). The Ministry's goals in developing the strategy are to ensure that:

- consistent processes are in place across the Ministry of Health business units to manage data quality;
- continual assessment of the level of quality of the data held in the National health collections managed by the Ministry of Health;
- priorities are identified for data quality improvement initiatives;
- a continual cycle of data quality improvement and assessment is in place.

The NHI has a fundamental role to play in many of our national collections by providing a reference framework. We are therefore concentrating initial data quality improvement efforts on the NHI collection, with subsequent rollout to all other national health data collections. The uses of the NHI will change with the changing needs of the sector, such as the move to population based funding and PHO's. We therefore need to ‘future proof’ the NHI data quality improvement initiatives.

The overall objective of the workshop is to develop a targeted roadmap and portfolio of data quality improvement initiatives as a basis for practically informing the development of the DQIS. The intention is to qualify the approach thus developed on the NHI Data Collection, as the ‘demonstration collection’.

What do I need to do to prepare?

All workshop participants will need to be familiar with the following documents relevant to the topic and the workshop objectives:

- Ministry of Health’s 2003 Information Services Strategic Plan;
- Draft Data Quality Improvement Strategy;

Please take some time to think about the objectives of the workshop in respect to your role, experience and expectations of data quality improvement and be prepared to contribute to the workshop discussions.
**Workshop Content**

**Exercise 1. Current Business as Usual Initiatives?**
Given the future requirements of the NHI, will the current BAU work be applicable in 3 years time?

<table>
<thead>
<tr>
<th>BAU initiative</th>
<th>Is it the right thing to do for the future of the NHI (1-5)?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monthly reporting of duplicates to DHBs – number of duplicates etc.</td>
<td>5 – Agree strongly 1- Disagree strongly</td>
</tr>
<tr>
<td>Weekly statistics are provided on the NZHIS website.</td>
<td></td>
</tr>
<tr>
<td>Nightly scripts to find duplicates with manual matching by the operations team.</td>
<td></td>
</tr>
<tr>
<td>Ad hoc searches for duplicates based on observed issues</td>
<td></td>
</tr>
<tr>
<td>Manual matching of PHO registries</td>
<td></td>
</tr>
<tr>
<td>Birth and death indexes every month are checked against changes in the NHI.</td>
<td></td>
</tr>
</tbody>
</table>

**NHI Upgrade Programme**

<table>
<thead>
<tr>
<th>NHI Upgrade Programme</th>
<th>Is it the right thing to do for the future of the NHI (1-5)?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data tags – e.g. do not merge</td>
<td>5 – Agree strongly 1- Disagree strongly</td>
</tr>
<tr>
<td>NOAH</td>
<td></td>
</tr>
<tr>
<td>NHI Training programme</td>
<td></td>
</tr>
<tr>
<td>Improved geo-coding</td>
<td></td>
</tr>
<tr>
<td>New data elements</td>
<td></td>
</tr>
<tr>
<td>New Application Programming interface</td>
<td></td>
</tr>
<tr>
<td>Public awareness</td>
<td></td>
</tr>
<tr>
<td>Views for different population groups</td>
<td></td>
</tr>
<tr>
<td>Audit of users</td>
<td></td>
</tr>
</tbody>
</table>

**Exercise 2: What are the most important data quality dimensions for the NHI?**
What are the data quality requirements for the NHI in relation to the data quality dimensions outlined the Ministry of Health Data Quality Framework? Please rate your overall response for each characteristic i.e. How important is **Coverage**? Included is the assessment criteria used in the data quality framework to give you an understanding of the meaning of each characteristic.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>How important is this (1-5)</th>
<th>How well do we do this now (1-5)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Coverage</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>o The population of reference is explicitly stated in all releases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o Known sources of under/over-coverage have been documented</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o The frame has been validated by comparison with external sources</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o The rate of under coverage falls into one of the predefined categories</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o The rate of over coverage falls into one of the predefined categories</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Capture and collection</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>o Practices exist that minimise response burden</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o Practices exist that encourage co-operation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o Practices exist that give support to data suppliers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o Standard manual and electronic data submission forms and procedures exist</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o Data capture quality control measures exist</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o Data capture is at source i.e. health encounter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o Data are stored at unit level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o Data quality feedback is given to data suppliers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o Data quality feedback issues as noted by users is feedback to database quality managers</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Missing records</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>o The magnitude of missing records is mentioned in the data quality documentation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o The number of records received are monitored and measured for unusual values</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o The magnitude of missing records falls into one of the predetermined categories</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o Data are collected at level that provides sufficient detail for its use</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Item (partial) Missing values</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>o Item missing records is identified</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o The magnitude of missing records falls into one of the predetermined categories</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Measurement error</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>o The level of overall error falls into one of the predetermined categories</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o The level of bias is not significant</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o The degree of problems with consistency falls into one of the predetermined categories</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Edit and Imputation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>o Validity checks are done for each variable at source and at each transfer of data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o Edits and imputation rules are logical and consistent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o Edit reports for users are easy to use and understand</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o Any imputation that is done is automatically derived from edits</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Timeliness</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>o The difference between the actual date of release and the end of the reference period is reasonably brief</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o The official date of release was announced in advance of the release in the schedule of publications</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o The official date of release was met</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o Database/registry methods are regularly reviewed for efficiency</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Documentation currency</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>o The recommended data quality documentation was available at the time of data or report release</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Comparability

- **Data dictionary standards**
  - Data elements are evaluated in comparison to the appropriate data dictionary
  - Data elements conform to the appropriate data dictionary
  - Common data elements across the Ministry are defined consistently

- **Standardisation**
  - For any derived variable, the original variable is also maintained on the main database

- **Linkage**
  - Each coded field in the collection uses accepted coding table standards
  - Data are collected using a consistent time frame and format
  - Ministry of Health standardised codes are used to uniquely identify institutions
  - Codes are used to uniquely identify persons

- **Equivalency**
  - The impact of problems related to relationships (one to one, one to many, many to many) falls into one of the predetermined categories
  - Methodology and limitations of mappings are documented

- **Historical comparability**
  - Trend analysis is used to examine changes in key data elements over time
  - The impact of historical changes are known
  - Accessible documentation on historical changes to the database exists

Visibility

- **Accessibility**
  - An official subset of micro-data are defined, created, made available, and frozen per release for users where appropriate
  - Standard tables and analysis are produced per release
  - Products are defined, catalogued, and/or publicised
  - Reports are available through HADIE

- **Documentation**
  - Data quality documentation exists per subset release
  - Database/registry methods documentation exists per subset release
  - A caveat accompanies any official release

- **Interpretability**
  - A mechanism is in place whereby key users can provide feedback to, and receive notice from, the product area
  - Revision guidelines are available per subset release

Relevance

- **Adaptability**
  - Client and stakeholder liaison mechanisms are in place to keep informed of developments
  - The database/registry can adapt to change

- **Value**
  - The mandate of the data holding fills a health information gap
  - The level of usage and types of users of the data holding is monitored
  - User satisfaction is periodically solicited

Privacy and Security

- **Legislation**
  - All data has been classified in accordance with SIGS
  - Policies around archiving of raw data have been developed
  - Unauthorised changes to data are prevented while the data are in transit
  - Data privacy is maintained in accordance to legislations and available standard

- **Disclosure**
  - The identity of the user is assured before allowing access to data
  - Data are not made available or disclosed to unauthorised individuals, entities, or processes


Details can be found in **Section 5** of the Draft Data Quality Strategy.
### Exercise 4: Brainstorming

The workshop will be split into 5 separate groupings. Each group is required to complete the elements for the assigned area to look at and discuss their viewpoint with the rest to the workshop participants when the group as a whole reconvenes. Possible areas are:

1. Education (Sector, Consumers, Ministry)
2. Culture Change Sector wide)
3. Incentives for change (DHB’s, PHO’s)
4. Performance Measurement and Monitoring

### Elements to be addressed by groups:

<table>
<thead>
<tr>
<th>Question to ask yourselves</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>What are the projects that relate to this area?</td>
<td></td>
</tr>
<tr>
<td>Which aspects need to be done first?</td>
<td></td>
</tr>
<tr>
<td>What resources would be required?</td>
<td></td>
</tr>
<tr>
<td>Who should own/take the lead in this area?</td>
<td></td>
</tr>
<tr>
<td>Where should the money come from?</td>
<td></td>
</tr>
<tr>
<td>Who are the stakeholders?</td>
<td></td>
</tr>
<tr>
<td>What kind of communication strategy may be appropriate for them?</td>
<td></td>
</tr>
<tr>
<td>How will we know if we are succeeding with this area?</td>
<td></td>
</tr>
</tbody>
</table>

### Strategies

<table>
<thead>
<tr>
<th>Strategies</th>
<th>How much impact will this have on the quality of the NHI (1-5)?</th>
<th>What priority should this have (1-5)?</th>
<th>How soon should this occur?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ministry</td>
<td>5 – Significant 1 – Trivial</td>
<td>5 – Very high 1 - Low</td>
<td>5 – immediately 3 – Within 6 months 5 – Within 2 years</td>
</tr>
<tr>
<td>Data Quality Framework Implementation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prioritise Customers and their needs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard Operational Initiatives</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metadata improvement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Collections – minimum requirements</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Production Maps</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistical Process Control</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Review DQ Team roles (new position for trainer)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Update data access policy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Review business rules and data dictionary across the Ministry</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sector</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Development of KPIs for DHB Contracts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Development of best practice guidelines for data collection processes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Development of a comprehensive education programme for MOH and sector</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DHB/PHO Accreditation scheme</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Peer review/audit</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Best Practice guidelines implementation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• KPIs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• In-house education programme</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• In house data quality improvement strategy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other Ideas</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 2: Current State Analysis of Data Quality in the New Zealand Health Sector Interview Guidelines

Data Quality in the New Zealand Health Sector

Roles and Responsibilities

1. Who is responsible for data quality in your organisation? Is this responsibility reflected in employment contracts?

2. What are the main positions working on data quality?

3. What role do you play as CIO?

4. Do you have an organisation wide data quality team? If yes, who are the members?

5. What is their role? Do they have a Terms of Reference?

Data Quality Programme

6. Do you have a structured programme for data quality improvement? Do you have an organisation wide data quality strategy?

7. Can you outline the programme?

8. Is the programme specifically linked to your business and ISSP?

9. Do you formally measure and or assess data quality levels – if yes how?

10. What are the regular processes to improve data quality?

11. What are the preventative data quality measures regularly taken?

12. What are the constraints you face(d) in implementing data quality improvements?

13. What is the priority for data quality improvement in the health sector?

Management Support

14. What is your management team’s perceived attitude to data quality?

15. Do you consider your organisation is lacking in any particular area of data quality management?
Appendix 3: Data Quality Dimensions Survey

This brief survey is designed to elicit the dimensions of data quality that the health sector considers important for the national health data collections. The concept of data quality is defined by a set of dimensions, usually considered in data quality literature as quality properties or characteristics of data and a set of data quality attributes that most data consumers react to in a fairly consistent way. A rating of 1 is the dimension you consider to be the most important for measuring and improving in the health sector. A rating of 6 is the dimension you consider to be least important for improvement. Assign each number once from 1-6.

Please place the completed survey in the box at the conference reception desk.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Definition</th>
<th>Rating 1 (most important) to 6 (least important) use each number once only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>Accuracy refers to how well data reflects the reality it is supposed to represent.</td>
<td></td>
</tr>
<tr>
<td>Timeliness</td>
<td>Timeliness refers primarily to how current or up-to-date the data are at the time of use.</td>
<td></td>
</tr>
<tr>
<td>Comparability</td>
<td>Comparability is defined as the extent to which databases are consistent over time and use standard conventions (such as data elements or reporting periods), making them similar to other databases.</td>
<td></td>
</tr>
<tr>
<td>Usability</td>
<td>Usability reflects the ease with which data may be understood and accessed. If data are difficult to use, they can be rendered worthless no matter how accurate, timely, comparable, or relevant they may be.</td>
<td></td>
</tr>
<tr>
<td>Relevance</td>
<td>Relevance reflects the degree to which a database meets the current and potential future needs of users.</td>
<td></td>
</tr>
<tr>
<td>Security and Privacy</td>
<td>This dimension reflects the degree to which a database meets the current legislation, standards, policies and processes.</td>
<td></td>
</tr>
</tbody>
</table>

Are you a (please circle all that are applicable)?: Data User  Data Collector  Data Custodian (manage data on behalf of others)

Are there any other dimensions of data quality that should be considered important for measuring and improving in health? Please list:

---

I am a PhD student at The University of Auckland in the Department of Information Systems and Operations Management. I am conducting this research for the purpose of my thesis on the development of a Data Quality Improvement Strategy for the New Zealand Health Sector. You don't need to put any identifying information on your response. Please contact me if you would like further information or to discuss data quality in your organisation at karolynkerr@hotmail.com or on 021 740 162.

Thank you for your assistance. Karolyn Kerr

---
Appendix 4: Issues Register for Pilot of Data Quality Evaluation Framework

1. The questions are not straightforward. For example we have to make assumption in question 14 that the question is about internal process and not external.

2. Question 1 the question is not very clear, therefore we have made assumption it is data release rather system release the respondent need to think about when answering this question.

3. Different and unfamiliar terms being used such as unit response, item response etc. for example confusion about questions 17 and 19. Everyone agreed that it would have been better to use familiar terms such as a record, a field or a variable.

4. The teams are concerned that if it is difficult for them to understand the questions how valuable it will be for others.

5. We believe question 10 should have been two questions rather than just one. One question related to the suppliers and a second questions related to the data receivers. Because of the way the question is currently structured, we found it difficult to score question 10 since one score has to cover both the suppliers and the receivers.

6. Question 21 needs to be split into two questions to cover both internal and external level of errors. We suggest the following:
   - What is the level of error in the data supplied?
   - What is the level of errors introduced internally?

7. Question 25 - we made an assumption that the question is related to internal factors.

8. Question 27 – the question is not clear and the team didn’t support the arguments that were provided in the Data Quality Framework Manual - Criterion 27.

9. Questions 21, 22 and 23 will be answered from NZHIS perspective.

10. Suggestions for questions 21-23:
   - **21. Overall error** assesses the degree to which the reported derived data matches the values on the raw data entered on the database that is the number of errors in a sample of data.
   - **Criterion** The level of overall error falls into one of the predetermined categories (as on page 32 of the DQF Manual).
   - **Measurement** % of error in a sample of data, as identified by an annual internal data audit and/or ad hoc external data audit.

   - **22. Bias** assesses to what degree the difference between the reported values and the raw data values on the database occurs in a systematic way.
   - **Criterion** The level of bias is not significant.
   - **Measurement** Criterion met if there is no evidence of/reason to believe there is a bias significant enough to affect the data to a noticeable degree.

   - **23. Consistency** assesses the amount of variation that would occur if repeated measurements are taken, that is the number of times the same error occurs in a sample of data.
   - **Criterion** the degree of problems with consistency falls into one of the predetermined categories (as on pages 33/34 of the DQF Manual)
• **Measurement** % of inconsistency in a sample of data, as identified by an annual internal data audit and/or ad hoc external data audit.

11. Question 23 tries to assess the data quality at a very high level. Often the data quality issues we deal with are field specific. Overall, the question is too broad and not very helpful for identifying lower level issues with the data quality.

12. Error in the manual starting from criterion 39 or 40 onwards. The issues identified is that the numbering in the data Quality Framework Manual doesn’t match to the excel tools provided.

13. In the Data Quality Framework Manual on page 18 the example provided starting with – In contrast, for nursing registration form…is not accurate description of the nursing data collection.

14. Question 60 – the team suggest that the term ‘data holding’ to be replaced with ‘data collections’, because this term is widely used within NZHIS.

15. SIGS needs to be defined when it is mention for the first time in Criterion 64. Please note SIG definition is not provided in the glossary either.

16. In Criterion 65 – there was a discussion about what constitutes ‘raw data’. We assumed that ‘raw data’ is a hard copy. However, we think this question need to be more specific and address both hard and soft copy data therefore, we suggest that the question to be split into two questions:

   - Hard copy – paper based data and
   - Soft copy – data within the systems.

17. Question 66 is not clear therefore more clarification is required. What does the word ‘transit’ cover? Is it for data coming in to NZHIS or sent out from NZHIS. Does it cover only electronic data or both electronic and paper base data?

18. Suggestion:

   - Data quality assessment should be performed just before the annual data publication release.

19. From question 63 onwards in the excel tool the tables are not there, but this is a minor error.

20. In the Data Instrument Template page 6 Workforce Statistic Data Warehouse has to be changed to Workforce Data.
Data Quality Evaluation Framework Instrument Template

September 2004
Introduction

This document is a modification of the third version of the Canadian Institute for Health Information's (CIHI) Data Quality Manual.

To align the Data Evaluation Quality Framework (DQEF) to the New Zealand health context, a review of the status of national data collections was undertaken. This ascertained the current quality of data as perceived amongst data quality managers. Also, research revealed further rationale and identified issues relating to data quality. Focus group meetings were held with the Ministry’s data quality managers and data users, and drafts were distributed for feedback.

How to use this template

The Data Quality Framework Instrument Template was created to facilitate the computation of a data collection evaluation. As well, a Manual is available that specifically describes the content, administration, scoring, and interpretation of the instrument.

This instrument includes a pre-assessment checklist, a list of current Ministry of Health data collections, a hierarchical algorithm, a report template and a post-assessment checklist.

What to do:
1. Complete the pre-assessment checklist to ensure that the data collection is appropriate for assessment
2. Select from the list of Ministry of Health data collections and complete your details and date
3. Complete the dimension, characteristics and criteria table
4. Complete the report template
5. Complete the post-assessment checklist

Pre-assessment checklist

<table>
<thead>
<tr>
<th>Pre-assessment checklist</th>
<th>✓</th>
</tr>
</thead>
</table>
| 1. Stakeholders have been contacted.  
  a. A knowledgeable contact (data quality manager) has been identified to undertake assessment or to assist in assessment.  
  b. The owner of the collection knows that the collection is being assessed. | ✓ |
| 2. An appropriate time for assessment has been identified.  
  a. The Collection is ready for assessment (there is a production system to assess - not a pilot study or prototype). | ✓ |
| 3. A Data Quality Folder exists and metadata for the collection is available (if not, complete the Metadata template and DQEF Folder)  
  a. The purpose of the collection is known  
  b. All sources of data have been identified  
  c. The business owner of the collection has been identified  
  d. The manager of the collection has been identified  
  e. The data quality manager(s) have been identified  
  f. Dates are known for the last update of the data  
  g. A schedule of release dates is available | ✓ |
| 4. Current data quality practices for the collection have been identified  
  a. Documentation for data processes is available | ✓ |
| 5. Coverage statistics are known  
  a. The population of reference is known  
  b. The frame and frame maintenance procedures have been identified  
  c. External sources to validate frame are identified  
  d. Non-response statistics are known | ✓ |
Ministry of Health data collections

Please select the relevant data collection to be assessed from the following:

<table>
<thead>
<tr>
<th>HealthPAC Dunedin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Care Manager Database</td>
</tr>
<tr>
<td>Client Claims &amp; Payment System</td>
</tr>
<tr>
<td>Contract Management System</td>
</tr>
<tr>
<td>Output Collection Programme</td>
</tr>
<tr>
<td>PERORG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HealthPAC Wellington</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capitation Base Funding / Capitation Information Cleansing</td>
</tr>
<tr>
<td>Dental System</td>
</tr>
<tr>
<td>Maternity Data Warehouse</td>
</tr>
<tr>
<td>Operational Reporting Database</td>
</tr>
<tr>
<td>PROCLAIM</td>
</tr>
<tr>
<td>Special Authorities</td>
</tr>
<tr>
<td>WINCAPP</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NZHIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Medical Services Data Warehouse</td>
</tr>
<tr>
<td>Health Practitioner Index or National Provider Index</td>
</tr>
<tr>
<td>Hepatitis B Screening Register</td>
</tr>
<tr>
<td>Immunisation Data Warehouse</td>
</tr>
<tr>
<td>Laboratory Claims Data Warehouse</td>
</tr>
<tr>
<td>Maternity and Newborn Information System</td>
</tr>
<tr>
<td>Mental Health Information National Collection</td>
</tr>
<tr>
<td>Medical Warnings System</td>
</tr>
<tr>
<td>Mortality Collection</td>
</tr>
<tr>
<td>National Booking Reporting System</td>
</tr>
<tr>
<td>National Booking Reporting System Data Warehouse</td>
</tr>
<tr>
<td>National Cancer Registry</td>
</tr>
<tr>
<td>National Health Index</td>
</tr>
<tr>
<td>National Minimum Dataset</td>
</tr>
<tr>
<td>Pharmaceutical Information Database</td>
</tr>
<tr>
<td>Private Hospital Data Collection</td>
</tr>
<tr>
<td>Workforce Statistics Data Warehouse</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>National Screening Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast Screening Register</td>
</tr>
<tr>
<td>Cervical Screening Register</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other collections in original inquiry</th>
</tr>
</thead>
<tbody>
<tr>
<td>kidZnet/Kidslink</td>
</tr>
<tr>
<td>Waiting Times Fund System</td>
</tr>
</tbody>
</table>

Or other database (please specify):

Evaluation Date: _________ (dd) __________ (mm) ______ (yyyy)
Evaluator Name: ________________________________
Dimensions, characteristics, criteria

The algorithm’s base assesses 69 characteristics of data quality. These characteristics are given with specific guidelines, which are explained in detail in “Ministry of Health, Data Quality Evaluation Framework Manual”. The manual:

- describes each characteristic;
- gives criteria for assessing the characteristic;
- relates the criteria to an evaluation category (appropriate, marginal, not acceptable, and unknown);
- defines how often the characteristic is to be assessed.

Circle the applicable evaluation number in the right hand column from the detailed assessment information provided in sections 9 through 15.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Characteristic</th>
<th>Criteria and measurement</th>
<th>Evaluations (0,1-3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>Coverage</td>
<td>1 The population of reference is explicitly stated in all releases</td>
<td>0. Not Applicable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measurement: The population of reference is stated in all releases during the last year</td>
<td>1. Unknown</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Not Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 Known sources of under/over-coverage have been documented</td>
<td>0. Not Applicable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measurement: Known sources of coverage error are documented internally or externally as</td>
<td>1. Unknown</td>
</tr>
<tr>
<td></td>
<td></td>
<td>required</td>
<td>2. Not Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 The frame has been validated by comparison with external sources</td>
<td>0. Not Applicable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measurement: Frame has been compared to external sources within the last year to</td>
<td>1. Unknown</td>
</tr>
<tr>
<td></td>
<td></td>
<td>determine the presence of errors in the frame</td>
<td>2. Not Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 The rate of under coverage falls into one of the predefined categories</td>
<td>0. Not Applicable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measurement: Levels of under coverage fall within defined accepted levels</td>
<td>1. Unknown</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Not Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 The rate of over coverage falls into one of the predefined categories</td>
<td>0. Not Applicable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measurement: Levels of over coverage fall within defined accepted levels</td>
<td>1. Unknown</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Not Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Met</td>
</tr>
<tr>
<td>Capture and</td>
<td></td>
<td>6 Practices exist that minimise response burden</td>
<td>0. Not Applicable</td>
</tr>
<tr>
<td>Collection</td>
<td></td>
<td>Measurement: Practices are in place to minimise response burden</td>
<td>1. Unknown</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Not Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7 Practices exist that encourage co-operation</td>
<td>0. Not Applicable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measurement: There are practices in place that encourage co-operation</td>
<td>1. Unknown</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Not Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8 Practices exist that give support to data suppliers</td>
<td>0. Not Applicable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measurement: At least two methods by which support is given to the data suppliers</td>
<td>1. Unknown</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Not Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9 Standard manual and electronic data submission forms and procedures exist</td>
<td>0. Not Applicable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measurement: Data submission forms and procedures used in data collection are</td>
<td>1. Unknown</td>
</tr>
<tr>
<td></td>
<td></td>
<td>standardised and standards are enforced</td>
<td>2. Not Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10 Data capture and receiver quality control measures exist</td>
<td>0. Not Applicable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measurement: Any quality control measures are used by the data suppliers at the data</td>
<td>1. Unknown</td>
</tr>
<tr>
<td></td>
<td></td>
<td>capture stage and by the receivers at the submission stage</td>
<td>2. Not Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Met</td>
</tr>
<tr>
<td>Dimension</td>
<td>Characteristic</td>
<td>Criteria and measurement</td>
<td>Evaluations (0,1-3)</td>
</tr>
<tr>
<td>-----------</td>
<td>---------------</td>
<td>--------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>11</td>
<td>Data capture is at source i.e. health encounter</td>
<td>Measurement: Processes exists and are documented to ensure data is collected from the most appropriate source for its use</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
</tr>
<tr>
<td>12</td>
<td>Data is available at unit level</td>
<td>Measurement: Data is available and easily retrieved by appropriate users at unit level</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
</tr>
<tr>
<td>13</td>
<td>Data is collected at level that provides sufficient detail for its use</td>
<td>Measurement: The level of specificity of data collected is sufficient to meet the needs of the primary users</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
</tr>
<tr>
<td>14</td>
<td>Comparative data quality feedback is given to data suppliers</td>
<td>Measurement: Comparative data quality feedback is given to data suppliers</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
</tr>
<tr>
<td>15</td>
<td>Data quality issues as noted by users are fed back to data collection managers</td>
<td>Measurement: A mechanism is in place for users of data to feedback data quality issues to data collection managers</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
</tr>
<tr>
<td>Missing Records</td>
<td>16</td>
<td>The magnitude of missing records is mentioned in the data quality documentation</td>
<td>Measurement: Magnitude of missing records is reported in the data quality documentation at a level of detail relevant for most analysis</td>
</tr>
<tr>
<td>17</td>
<td>The number of records received are monitored and measured for unusual values</td>
<td>Measurement: Numbers of records received below the first level are monitored and measured over time for unusual values</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
</tr>
<tr>
<td>18</td>
<td>The magnitude of unit missing records falls into one of the predetermined categories</td>
<td>Measurement: Item response rate should be calculated for all variables that are used in analysis of the data</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
</tr>
<tr>
<td>Item (Partial) Non-Response</td>
<td>19</td>
<td>Item missing data is identified</td>
<td>Measurement: Item missing data can be identified for all core variables and is flagged when the variable is only conditionally required. A core variable is any variable that is routinely used in analysis.</td>
</tr>
<tr>
<td>20</td>
<td>The magnitude of item missing data falls into one of the predetermined categories</td>
<td>Measurement: Volume of variables with item missing data</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
</tr>
<tr>
<td>Measurement Error</td>
<td>21</td>
<td>The level of overall error falls into one of the predetermined categories</td>
<td>Measurement: % of perceived error test for appropriate values in pilot study</td>
</tr>
<tr>
<td>22</td>
<td>The level of bias is not significant</td>
<td>Measurement: No evidence of, or no reason to believe there is, a bias significant enough to affect the estimates to a noticeable degree</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
</tr>
<tr>
<td>23</td>
<td>The degree of problems with consistency falls into one of the predetermined categories</td>
<td>Measurement: % of perceived inconsistency</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
</tr>
<tr>
<td>Dimension</td>
<td>Characteristic</td>
<td>Criteria and measurement</td>
<td>Evaluations (0,1-3)</td>
</tr>
<tr>
<td>--------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td><strong>Edit and Imputation</strong></td>
<td>24 Validity checks are done for each variable at source and at each transfer of data</td>
<td>Measurement: All variables are checked for validity and any invalid data are at least flagged as invalid at source, and for each subsequent transfer between business units</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
</tr>
<tr>
<td></td>
<td>25 Edits and imputation rules are logical and consistent</td>
<td>Measurement: The edits and imputation rules are determined to be logical and obvious consistency checks are in place</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
</tr>
<tr>
<td></td>
<td>26 Edit reports for users are easy to use and understand</td>
<td>Measurement: Edit reports are easy to understand and in a usable format</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
</tr>
<tr>
<td></td>
<td>27 Any imputation that is done is automatically derived from edits</td>
<td>Measurement: All imputation done is automated</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
</tr>
</tbody>
</table>

| **Estimation and Processing** | 28 Documentation for all data processes are maintained | Measurement: All the processes that are run by the database personnel are adequately documented | 0. Not Applicable 1. Unknown 2. Not Met 3. Met |
|                               | 29 Documentation for all systems, programs or applications is maintained     | Measurement: The systems, programs or applications that are used with the database are documented | 0. Not Applicable 1. Unknown 2. Not Met 3. Met |
|                               | 30 Processing system has been tested after the last revision                 | Measurement: The systems are tested when changes are made                                 | 0. Not Applicable 1. Unknown 2. Not Met 3. Met |
|                               | 31 Raw data are saved in a secure location                                   | Measurement: The data that arrives from data providers is saved in a secure location      | 0. Not Applicable 1. Unknown 2. Not Met 3. Met |

| **Overall assessment of Accuracy Dimension** | | | |
|-----------------------------------------------|---------------------------------------------------------------|---------------------------------------------------------------|
| **Appropriate**                               | at least 5 characteristics are appropriate, neither of the other two are problematic | | |
| **Marginal**                                  | everything between appropriate and problematic                 | | |
| **Problematic**                               | any characteristic is problematic                              | | |
| **Unknown**                                   | cannot be assigned, unknown accuracy is problematic            | | |

<p>| <strong>Timeliness</strong>                                | <strong>Data Currency At The Time of Release</strong>                      | | |
|-----------------------------------------------|---------------------------------------------------------------|---------------------------------------------------------------|
| 32 The difference between the actual date of release and the end of the reference period is reasonably brief | Measurement: The difference between the actual date of release and the end of the reference period is reasonably brief | 0. Not Applicable 1. Unknown 2. Not Met 3. Met |
| 33 The official date of release was announced in advance of the release in the schedule of publications | Measurement: The official date of release for the schedule of publication of data was planned for was announced at least six months in advance of the release | 0. Not Applicable 1. Unknown 2. Not Met 3. Met |
| 34 The official date of release was met       | Measurement: Data was released on or before the official date of release, if not, reasons have been documented | 0. Not Applicable 1. Unknown 2. Not Met 3. Met |</p>
<table>
<thead>
<tr>
<th>Dimension</th>
<th>Characteristic</th>
<th>Criteria and measurement</th>
<th>Evaluations (0,1-3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Documentation Currency</td>
<td>36 The recommended data quality documentation was available at the time of data or report release</td>
<td>Measurement: Data quality documentation was available and linked to data or report at data or report release</td>
<td>0. Not Applicable</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1. Unknown</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Not Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Met</td>
</tr>
<tr>
<td>Overall assessment of Timeliness Dimension</td>
<td>Appropriate = both characteristics are appropriate</td>
<td>Marginal = everything in-between appropriate and problematic</td>
<td>Problematic = either characteristic is scored as problematic</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unknown = is not an option and is scored as problematic</td>
<td></td>
</tr>
<tr>
<td>Comparability</td>
<td>Data Dictionary Standards</td>
<td>37 Data elements are evaluated in comparison to the appropriate data dictionary</td>
<td>0. Not Applicable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measurement: The data elements have been evaluated against the appropriate data dictionary and evaluation documented</td>
<td>1. Unknown</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Not Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td>38 Data elements conform to the appropriate data dictionary</td>
<td>0. Not Applicable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measurement: All elements conform to the data dictionary or variations / mappings documented</td>
<td>1. Unknown</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Not Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td>39 Common data elements across the Ministry are defined consistently</td>
<td>0. Not Applicable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measurement: All data elements conform to a consistent definition standard</td>
<td>1. Unknown</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Not Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Met</td>
</tr>
<tr>
<td>Standardisation</td>
<td>40 For any derived variable the original variable is also maintained on the main database</td>
<td>Measurement: No original variables are permanently deleted from the main database</td>
<td>0. Not Applicable</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1. Unknown</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Not Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Met</td>
</tr>
<tr>
<td>Linkage</td>
<td>41 Each coded field in the collection uses accepted health sector coding table standards</td>
<td>Measurement: All coded fields use valid and appropriate code tables with version used and variations documented</td>
<td>0. Not Applicable</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1. Unknown</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Not Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td>42 Data are collected using a consistent time frame and format</td>
<td>0. Not Applicable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measurement: A common format for date is used within the collection and documented</td>
<td>1. Unknown</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Not Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td>43 Health sector standardised codes are used to uniquely identify institutions</td>
<td>0. Not Applicable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measurement: Acceptable institution code is available on the database and code table used is documented</td>
<td>1. Unknown</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Not Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td>44 Health sector standardised codes are used to uniquely identify persons</td>
<td>0. Not Applicable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measurement: NHl or HPI are used</td>
<td>1. Unknown</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Not Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Met</td>
</tr>
<tr>
<td>Equivalency</td>
<td>45 The impact of problems related to relationships (one to one, one to many, many to many) falls into one of the predetermined categories</td>
<td>Measurement: All codes are being converted properly and the impact is known. Equivalency has been investigated</td>
<td>0. Not Applicable</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1. Unknown</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Not Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td>46 Methodology and limitations of mappings are documented</td>
<td>0. Not Applicable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measurement: Mappings are adequately documented</td>
<td>1. Unknown</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Not Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Met</td>
</tr>
<tr>
<td>Historical Comparability</td>
<td>47 Trend analysis is used to examine changes in key data elements over time</td>
<td>Measurement: Trend analysis has been performed for key data elements as appropriate to the collection</td>
<td>0. Not Applicable</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1. Unknown</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Not Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Met</td>
</tr>
</tbody>
</table>
### Overall assessment of Comparability Dimension

- **Appropriate** – all characteristics are appropriate
- **Marginal** – everything between appropriate and problematic
- **Problematic** – any one characteristic is problematic
- **Unknown** – cannot be assigned. Unknown comparability is problematic

### Usability

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Characteristic</th>
<th>Criteria and measurement</th>
<th>Evaluations (0,1-3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accessibility</td>
<td>50 An official subset of micro-data is defined, created, made available, and frozen per release</td>
<td>Measurement: A micro data sub set of the database/registry is frozen per release</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
</tr>
<tr>
<td>Documentation</td>
<td>51 Standard tables and analysis are produced per release</td>
<td>Measurement: Commonly used standard tables and analysis are made available at each release</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
</tr>
<tr>
<td>Interpretablity</td>
<td>52 Products are defined, catalogued, and/or publicised</td>
<td>Measurement: The database/registry is listed in Ministry of Health dissemination systems for each release</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
</tr>
<tr>
<td>Usability</td>
<td>53 Reports are available through HADIE (^15)</td>
<td>Measurement: Reports are available through HADIE at time of release</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
</tr>
<tr>
<td>Usability</td>
<td>54 Data quality documentation exists per subset release</td>
<td>Measurement: Data quality documentation for users exists per subset release, is accessible and maintained</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
</tr>
<tr>
<td>Usability</td>
<td>57 A mechanism is in place whereby key users can provide feedback to, and receive notice from, the product area</td>
<td>Measurement: Product area contact information is included with any major release and if major users are encouraged to use the contact information to provide feedback on any limitations they may discover</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
</tr>
</tbody>
</table>

---

\(^15\) HADIE, the Health and Disability Sector Information Exchange will be a single point of access providing authenticated users with access to all data collections available to them, and this will include comprehensive and up to date data descriptions (metadata). (From MOH 2003 ISSP – Chapter 4 – data strategy)
<table>
<thead>
<tr>
<th>Dimension</th>
<th>Characteristic</th>
<th>Criteria and measurement</th>
<th>Evaluations (0,1-3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall assessment of Usability dimension</td>
<td>58 Revision guidelines are available per subset release Measurement: Database specific revision guidelines are available and applied per subset release</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overall assessment of Usability dimension</td>
<td>59 Client and stakeholder liaison mechanisms are in place to keep informed of developments Measurement: Client and stakeholder liaison mechanisms are in place to keep informed of developments</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
</tr>
<tr>
<td></td>
<td>60 The database/registry can adapt to change Measurement: Database/registry has demonstrated an ability to adapt to an important emerging issue, to a new technical standard, or to a major data quality limitation</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
<td></td>
</tr>
<tr>
<td></td>
<td>61 The mandate of the data holding fills a health information gap Measurement: The mandate of the data holding fills a health care information gap</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
<td></td>
</tr>
<tr>
<td></td>
<td>62 The level of usage and types of users of the data holding is monitored Measurement: The level of usage and types of users of the data holding is monitored</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
<td></td>
</tr>
<tr>
<td></td>
<td>63 User satisfaction is periodically solicited Measurement: Client satisfaction assessments are conducted on a regular basis and a tool has been developed for consistent assessment</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overall assessment of Relevance dimension</td>
<td>59 Client and stakeholder liaison mechanisms are in place to keep informed of developments Measurement: Client and stakeholder liaison mechanisms are in place to keep informed of developments</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
</tr>
<tr>
<td></td>
<td>60 The database/registry can adapt to change Measurement: Database/registry has demonstrated an ability to adapt to an important emerging issue, to a new technical standard, or to a major data quality limitation</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
<td></td>
</tr>
<tr>
<td></td>
<td>61 The mandate of the data holding fills a health information gap Measurement: The mandate of the data holding fills a health care information gap</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
<td></td>
</tr>
<tr>
<td></td>
<td>62 The level of usage and types of users of the data holding is monitored Measurement: The level of usage and types of users of the data holding is monitored</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
<td></td>
</tr>
<tr>
<td></td>
<td>63 User satisfaction is periodically solicited Measurement: Client satisfaction assessments are conducted on a regular basis and a tool has been developed for consistent assessment</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overall assessment of Relevance dimension</td>
<td>59 Client and stakeholder liaison mechanisms are in place to keep informed of developments Measurement: Client and stakeholder liaison mechanisms are in place to keep informed of developments</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
</tr>
<tr>
<td></td>
<td>60 The database/registry can adapt to change Measurement: Database/registry has demonstrated an ability to adapt to an important emerging issue, to a new technical standard, or to a major data quality limitation</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
<td></td>
</tr>
<tr>
<td></td>
<td>61 The mandate of the data holding fills a health information gap Measurement: The mandate of the data holding fills a health care information gap</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
<td></td>
</tr>
<tr>
<td></td>
<td>62 The level of usage and types of users of the data holding is monitored Measurement: The level of usage and types of users of the data holding is monitored</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
<td></td>
</tr>
<tr>
<td></td>
<td>63 User satisfaction is periodically solicited Measurement: Client satisfaction assessments are conducted on a regular basis and a tool has been developed for consistent assessment</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
<td></td>
</tr>
<tr>
<td>Security and Privacy</td>
<td>Legislation</td>
<td>64 All data has been classified in accordance with SIGS Measurement: Classification is carried out, reviewed regularly and documented</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
</tr>
<tr>
<td></td>
<td>65 Policies around archiving of raw data have been developed Measurement: Policies around archiving raw data are adhered to</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
<td></td>
</tr>
<tr>
<td></td>
<td>66 Unauthorised changes to data is prevented while the data is in transit Measurement: Message integrity checks and encryption is used to maintain data integrity and confidentiality while in transit</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
<td></td>
</tr>
<tr>
<td></td>
<td>67 Data privacy is maintained in accordance to legislations and available standards Measurement: There is evidence of meeting requirements of appropriate legislation. It is possible to audit the collection</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
<td></td>
</tr>
<tr>
<td></td>
<td>68 The identity of the user is assured before allowing access to data Measurement: Appropriate authentication mechanisms (e.g. digital certificates, tokens, smart cards, reusable passwords, etc) are implemented</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
<td></td>
</tr>
<tr>
<td>Dimension</td>
<td>Characteristic</td>
<td>Criteria and measurement</td>
<td>Evaluations (0,1-3)</td>
</tr>
<tr>
<td>-----------</td>
<td>---------------</td>
<td>--------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Data is not made available or disclosed to unauthorised individuals, entities, or processes</td>
<td>0. Not Applicable 1. Unknown 2. Not Met 3. Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measurement: Adequate access control mechanisms have been implemented and access granted based on the data access and release policies</td>
<td></td>
</tr>
<tr>
<td>Overall assessment of Security and Privacy dimension</td>
<td>Appropriate – All Characteristics are appropriate</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Marginal – One or more characteristics are not met</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Problematic – Any characteristic is scored as problematic</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown – unknown is not an option and is also scored as problematic</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Introduction
The purpose of the [collection name] collection is to …

Summary of Assessment by Dimension
Explain in words …

Accuracy
[Example] … problems noted in the accuracy dimension are …

Timeliness
[Example] … there are no noted problems with timeliness …

Comparability
[Example] … problems with comparability are …

Usability
[Example] … there are no noted problems with usability …

Relevance
[Example] … problems with relevance are …

Privacy and Security
[Example] … there are no noted problems with privacy and security …
Detailed [collection name] collection assessment

Use this section to complete the detailed analysis for each criterion. These results are then summarised in the sections above.

Accuracy Dimension

Coverage

Criterion 1: The population of reference is explicitly stated in all releases
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 2: Known sources of under/over-coverage have been documented
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 3: The frame has been validated by comparison with external sources
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 4: The rate of under coverage falls into one of the predefined categories
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 5: The rate of over coverage falls into one of the predefined categories
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Characteristic Assessment – Coverage:
Description here …

Capture and Collection

Criterion 6: Practices exist that minimize response burden
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 7: Practices exist that encourage co-operation
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 8: Practices exist that give support to data suppliers
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 9: Standard manual and electronic data submission forms and procedures exist
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 10: Data capture and receiver quality control measures exist
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 11: Data capture is at the most appropriate source i.e. health encounter
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 12: Data are available at unit level
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Characteristic Assessment – Capture and Collection:
Description here …

Criterion 13: Data is collected at a level that provides sufficient detail for its use
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 14: Comparative data quality feedback is given to data suppliers
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 15: Data quality issues as noted by users are fed back to data collection quality managers
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Characteristic Assessment – Capture and Collection:
Description here …

Missing Records
Criterion 16: The magnitude of missing records is mentioned in the data quality documentation
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 17: The numbers of records received are monitored to detect for unusual values
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 18: The magnitude of missing records falls into one of the predetermined categories
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Characteristic Assessment – Unit Non-Response:
Description here …

Missing data values
Criterion 19: Missing data values are identified
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 20: The magnitude of missing data values falls in one of the predetermined categories
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Characteristic Assessment – Item (Partial) Non-Response:
Description here …

Measurement Error
Criterion 21: The level of overall error falls into one of the predetermined
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …
Criterion 23: The degree of problems with consistency falls into one of the predetermined categories
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Characteristic Assessment – Measurement Error:
Description here …

Edit and Imputation

Criterion 24: Validity checks are undertaken for each variable at source and at each transfer of data
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 25: Edits and imputation rules are logical and consistent
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 26: Edit reports for users are easy to use and understand
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 27: Any imputation that is undertaken is automatically derived from data
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Characteristic Assessment – Edit and Imputation:
Description here …

Estimation and Processing

Criterion 28: Documentation for all data processes are maintained
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 29: Documentation for all systems, programs or applications is Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 30: Processing system has been tested after the last revision
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 31: Raw data are saved in a secure location
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Characteristic Assessment – Estimation and Processing:
Description here …

Overall Assessment of the Accuracy Dimension

Summary description for accuracy here …
Timeliness Dimension

Data Currency at the Time of Release

Criterion 32: The difference between the actual date of release and the end of the reference period is reasonably brief
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 33: The official date of release was announced in advance of the release
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 34: The official date of release was met
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 35: Database/registry methods are regularly reviewed for efficiency
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Characteristic Assessment – Data Currency At The Time of Release:
Description here …

Documentation Currency

Criterion 36: The recommended data quality documentation was available and attached at the time of data or report release
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Characteristic Assessment – Documentation Currency:
Description here …

Overall Assessment of the Timeliness Dimension

Summary description here …

Comparability Dimension

Data Dictionary Standards

Criterion 37: Data elements are evaluated in comparison to the appropriate Data Dictionary and documented
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 38: Data elements conform or are mapped to the appropriate Data Dictionary
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 39: Common data elements across the Ministry are defined consistently
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Characteristic Assessment – Data Dictionary Standards:
**Standardisation**

Criterion 40: For any derived variable, the original variable is also maintained on the main database
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

**Characteristic Assessment – Standardisation:**
Description here …

**Linkage**

Criterion 41: Each coded field in the collection uses accepted health sector coding table standards
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 42: Data are collected using a consistent time frame and format
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 43: Sector standardised codes are used to uniquely identify institutions
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 44: Sector standardised codes are used to uniquely identify persons
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

**Characteristic Assessment – Record Linkage:**
Description here …

**Equivalency**

Criterion 45: The impact of problems related to relationships (one to one, one to many, many to many) falls into one of the predetermined categories
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 46: Methodology and limitations of mappings are documented
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

**Characteristic Assessment – Equivalency:**
Description here …

**Historical Comparability**

Criterion 47: Trend analysis is used to examine changes in key data elements over time
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …
Criterion 48: The impact of historical changes is known
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 49: Other previous surveys or collections relating to this collection are known
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Characteristic Assessment – Historical Comparability:
Description here …

Overall Assessment of the Comparability Dimension
Summary description here …

Usability Dimension

Accessibility

Criterion 50: An official subset of micro-data is defined, created, made available, and frozen per release for users where appropriate
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 51: Standard tables and analysis are produced per release
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 52: Products are defined, catalogued, and/or publicised
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 53: Reports are published through HADIE
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Characteristic Assessment – Accessibility:
Description here …

Documentation

Criterion 54: Data quality documentation exists per subset release
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 55: Database/registry methods documentation exists for internal purposes per subset release
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 56: A caveat accompanies any official release
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Characteristic Assessment – Documentation:
Interpretability

Criterion 57: A mechanism is in place whereby key users can provide feedback to, and receive notice from, the product area
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 58: Revision guidelines are available and applied per subset release
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Characteristic Assessment – Interpretability:
Description here …

Overall Assessment of the Usability Dimension

Summary description here …

Relevance Dimension

Adaptability

Criterion 59: Client and stakeholder liaison mechanisms are in place to keep informed of developments
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 60: The database/registry can adapt to change
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Characteristic Assessment – Adaptability:
Description here …

Value

Criterion 61: The mandate of the data holding fills a health information
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 62: The level of usage and types of usage of the data holding is monitored
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 63: User satisfaction is periodically solicited
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Characteristic Assessment – Value:
Description here …

Overall Assessment of the Relevance Dimension

Summary description here …
Privacy and Security

Legislation

Criterion 64: All data has been classified in accordance with SIGS (Security in the Government Sector Guidelines)
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 65: Policies around archiving raw data have been developed
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 66: Unauthorised changes to data is prevented while the data is in transit
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Characteristic Assessment – Legislation:
Description here …

Disclosure

Criterion 67: Data privacy is maintained in accordance with legislation and available standards
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 68: The identity of the user is assured before allowing access to data
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Criterion 69: Data is not made available or disclosed to unauthorised individuals, entities or processes
Met (3), Not Met (2), Unknown (1), Not Applicable (0)
Description here …

Characteristic Assessment – Disclosure:
Description here …

Overall Assessment of the Privacy and Security Dimension

Summary description here …
### Post-assessment checklist

<table>
<thead>
<tr>
<th>Post-assessment checklist</th>
<th>✓</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. DQF Evaluation Instrument is complete.</td>
<td></td>
</tr>
<tr>
<td>2. Feedback to all stakeholders has been made.</td>
<td></td>
</tr>
<tr>
<td>3. The Data Quality Folder has been updated.</td>
<td></td>
</tr>
<tr>
<td>4. The date for the next assessment has been set.</td>
<td></td>
</tr>
<tr>
<td>5. Action Points from the evaluation have been noted in the Data Quality Folder.</td>
<td></td>
</tr>
</tbody>
</table>
Introduction

Purpose
The purpose of the Principles for Data Quality in National Health Data Collections is to ensure the proactive management of data quality throughout the health sector and is part of the sector wide Data Quality Improvement Strategy. Any new national health data collection will undergo an assessment of technical specifications, collection and management processes prior to implementation. Existing collections undergoing significant change will also be reviewed in line with the principles.

The principles provide a framework to assess new data collections as early as possible in their project phase. The principles assist the Clinical Analysis Team within NZHIS and project managers responsible for the implementation of new national data collections or to changes to existing collections. The assessment of a new data collection by applying the principles will highlight the data points at risk of producing poor quality data.

Scope
The current scope of the principles is restricted to all national health data collections under the custodianship of the New Zealand Health Information Service (NZHIS) as defined by the draft Data Stewardship Policy. All projects working on new data collections and significantly changing existing collections should refer to the principles for guidance on data quality.

Checklist Structure
This document provides a set of requirements, which allows the assessor to evaluate the status of data collections in terms of data quality needs. Seven areas of interest are examined:

1. management requirements;
2. service level agreements;
3. organisational specifications;
4. process definitions;
5. data collection, processing, and usage;
6. data quality management, measurement, measures, and monitoring;
7. technical requirements (data quality related).

Alignment
The principles have been developed in alignment with the draft Data Quality Improvement Strategy, the Ministry of Health Data Quality Evaluation Framework, and the draft New Zealand Health Information Strategy.

This document provides the principles for data collections in different fields of action during a project life cycle. The assessment schema to evaluate the data quality status of data collections is to be defined by the Clinical Analysis Team of NZHIS.
### Glossary

**Business Rule**
The business driven needs of a data consumer are described by business rules. It is a logical expression, which joins different business objects and helps to divide data into correct and faulty.

**Content Rule**
Measurable expression of a business rule on a data collection.

**Data Quality Metric**
Objective, measurable variable, which is sensitive to different specifications of quality attributes. The fulfilment of content, key, and matching rules is measured by Data Quality Metrics.

**HPI**
Health Practitioner Index. Unique identifier for health care providers.

**IT**
Information Technology

**JDBC**
Java Database Connectivity is a part of Java programming language, which provides a standardised interface to SQL compatible databases.

**Key Rule**
Measurable expression of a business rule on a data collection.

**Matching Rule**
Measurable expression of a business rule on a data collection.

**Normalisation**
Normalisation in databases helps avoid data redundancies (e.g. double entries or errors), and simplifying database maintenance. Six levels of normalisation are possible. But higher normalisation means lower performance to the database.

**NHI**
National Health Index. National health unique identifier for patients.

**Metadata**
Data about data or information about a collection

**ODBC**
Open Database Connectivity is a product-neutral interface between applications and database servers.

**Referential integrity**
Database built-in dependencies among tables, e.g. prevents deleting dependent information. Related topics are entity integrity, domain integrity, and user defined integrity.

**SQL**
Structured Query Language is the basic language to work with relational databases.

**TDQM**
The original TDQM cycle is adapted for the ongoing process of improvement of data collections.

The terms data and information are used synonymously in this document.
Use of the Principles

Who Will Use the Principles?
The development of a data quality plan will be produced for proposed new national health data collections and will follow these principles. Development of a data quality plan is to be undertaken by Project Managers when their project aim is to develop new collections or significantly alter existing collections. Project Managers should use the principles with the assistance of the NZHIS Clinical Analysis Team. The project sponsor has overall responsibility for ensuring these principles are applied throughout the project.

When Should the Principles be Used?
The Principles are applied every time a major part of the project changes and requirements as defined by the principles are impacted. These can be changes e.g. in the scope, the technical environment, and in the organisation. Project Managers should be aware of the requirements defined by the principles before the project starts to be able to define the related tasks. Project Managers and the Clinical Analysis Team define milestones in the project plan when the principles are applied, for example when:

- defining the organisational needs;
- the development of the business case;
- the development of the project plan;
- defining the technical environment;
- in the testing phase of the project;
- at the end of the project – the principles should be a part of the project review.

The Principles should be applied regularly throughout the project to make data quality risks manageable.

How Are the Principles to be Applied?
The document delivers more than a checklist. To make data quality manageable the requirements should be extracted into a checklist with an evaluation schema. At every decided milestone the responsible Project Manager should assess the new planned data collection against the defined requirements.

One might come to the conclusion that not every question can be applied to a specific project, but this decision should be in agreement with the NZHIS Clinical Analysis Team. The NZHIS Clinical Analysis Team can decide to review the application of the principles.
Data Quality Risk Assessment of a New Data Collection

Project Management Requirements

Project managers need to adhere to health sector policies such as the Privacy, Authentication and Security (PAS) guidelines, production security, business continuity etc. Data quality requirements need to be taken into account throughout the whole lifecycle of a project.

The clarification of the project’s business case plays a major role in the planning phase. Project managers need to apply with the following checklist during the planning phase:

Checklist:

- The project clearly defines the scope, the business perspective, the vision, and gives a scenario of the implemented change.16
- Clinicians and technical experts are consulted by the project team.
- All stakeholders come to an agreement about the data fields, which are stored in the data collection through an appropriate consultation process with the Health Sector.
- The project managers know about the basic requirements of data quality and they are aware of the Data Quality Evaluation Framework requirements for the new data collection.
- Resources such as budget, persons, and time are allocated to data quality in the project.
- The NZHIS Clinical Analysis team are consulted in the project-planning phase and have reviewed the project plan.
- The project delivers a Data Quality Activity Plan (template provided).
- The project managers frequently carry out a risk assessment with respect to data quality by applying the Data Quality Principles for National Health Collections. This is defined as milestones in the project plan.

Service Level Agreements

Service Level Agreements (SLA) describe how data are delivered with respect to textual, logical, chronological, contextual, and quality dependent requirements. These parameters are not independent of each other. Quality is highly related to the expectations of data consumers and these expectations have to be measurable and formulated in a SLA. The SLA should reflect the will of the participants to commit to data quality.

Checklist:

- A SLA is documented and agreed upon by the process participants.
- The SLA defines the data (fields) delivery textual, logical (description of dependencies), chronological, and contextual.
- The SLA defines the data quality requirements to organisation, processes, data quality measurements, and derived actions as described in this paper.

16 What do we want to do? What are the driving factors? What may be future needs? What are the benefits and who benefits of the project? Who are the stakeholders and drivers?
Organisational Specifications

To secure a high level of data quality in the data collection some certain roles have to be defined. These specifications should be defined for every data field. To keep the effort on a pragmatic level this should be done in information groups\(^\text{17}\) to which data fields belong.

Data Officer / Data Owner

The Data Officer is in charge of all issues belonging to certain data elements or information groups. The role maintains the data to ensure it is kept up-to-date in changing environments. The role of the Data Officer is normally identical with the Data Owner. Differentiations are artificial and belong mostly to organisational reasons, for example, a department owns data and implements a person in charge as Data Officer. Over the whole organisation the Chief Information Officer may have this role, however in the Ministry of Health this is often the role of the directorates.

Data Collector

The Data Collector is the source of data, which is later processed and used. This source can be a person or another trusted automated information source. The basics of good data quality are laid at this stage. Poor input data can often only be improved by applying considerable effort in time and cost. Therefore the Data Collector plays an important role in the data delivery chain and in the management of data quality.

Data Custodian

The Data Custodian processes the data, which were collected by the Data Collector. The Data Custodian has to secure data and information integrity with respect to the needs of the Data Consumer. For national health data collections the NZHIS is the Data Custodian.

Data Consumer

The Data Consumer is the user of collected and processed data. The consumer defines the requirements of the data and data information groups from a business perspective. In the business context a Business Manager is a special type of a Data Consumer. Everyone, who uses information deduced out of data, is a Data Consumer.

Checklist:

- The roles Data Owner / Data Officer, Data Collector, Data Custodian, and Data Consumer are defined and agreed upon by the process participants.
- The tasks and responsibilities of the predefined roles are clearly described and agreed upon.
- The process participants take part in basic data quality training.
- A data quality team of data collectors, custodians, and consumers is implemented to exchange data quality needs and experiences.

\(^{17}\) Data fields, which belong to the same logical information, are arranged in an information group. An address is an information group with the data fields ‘name, street, city, …’.
Process Definitions

Poor data quality is mainly related to the process landscape. The processes are not known, the known processes are not documented and accessible, or known and documented processes are not followed. All these reasons can lead to poor data quality. Therefore a project is required to ensure the definition and implementation of data quality processes.

Processes

The following processes are essential to implement and to document for good data quality. They should be explicitly described in the Data Quality Activity Plan and be implemented within the projects creating new national health data collections.

Data Collection Process

The Data Collection Process describes which and how data are collected at the source and which dependencies among data collections exist. In particular, the relationship between the new data collection to existing leading national collections like the NHI has to be taken into account. The Data Collection Process also includes technical processes like regular backups of data. It is important that main stakeholders can influence the Data Collection Process therefore the input of clinicians in the clinical coding process is highly important for the quality of coded data.

Data Correction Process

The collection of incorrect data can’t always be avoided. Therefore, a process for correction of data needs to be implemented. This includes the access, the access authorisation, and data protection and privacy issues.

Data Quality Process (working on TDQM cycle)

The Data Quality Process describes how data quality is monitored, measured and who is responsible for several steps like Define, Measure, Analyse, and Improve (including Report) in this process cycle. At every point in the quality cycle define where, when, who, and how action is taken. The process for already trusted information is described.

Change Process

Changes in such areas as legislation, new systems, new required data elements, extension of database, and the migration of data, can lead to new requirements for data collection. The new data collection has to be part of the organisation’s overall change management policy e.g. the 1 July change process. Changes in the data collection must be published for all process participants. Changes in leading / reference data collections, e.g. coding tables and classification systems, and national data collections such as the NHI, which directly influence the data collection, have to be observed.

Escalation Process

The escalation process guarantees all process participants a secure way to deal with up coming data quality problems in data collections. It describes how decisions are achieved in the case of data quality conflicts among the process participants.

Process Control

All processes are subject to improvement and change. Requirements of the NZHIS concerning processes are to be applied.

Checklist:

The following checklist applies for all above defined processes.

- The process fulfils the defined requirements.
- The process is appropriately described and documented and the documentation is accessible to relevant persons.
- For ease of understanding processes are visualised.
- The process is agreed upon the process participants.
• A process owner is defined, is named, and is known to the process participants.
• Documentations for all data processes are maintained.

Data Collection, Processing, and Usage
To avoid data redundancies new data collections are to be checked against existing and trusted leading / reference data collections. Leading data collections can be internal or external. Existing national data collections (e.g. NHI) are taken into account.

The implementation and definition of a ‘trusted system’ and data collection is essential for an integrated compilation of national data collections. This avoids data redundancies, eases the maintenance of the data collections, and lowers the costs of data holding.

Checklist:

Data Collection / Data Collector
• Existing leading / reference collections like the NHI or the HPI data collection are the only source for the data held by them. Data held by NHI or HPI are not to be held by the new data collection (for example, demographic data).
• Reference systems are taken as a source of data where appropriate.
• Data collection is from the most appropriate source for its intended uses.
• Data collectors are sufficiently trained and are aware of the importance of data quality of the collected data for the Sector and the Ministry of Health.
• Critical data are defined and collected using peer review where appropriate.
• Data collection quality control measures are installed.
• Data Collectors can easily correct mistakes identified during data collection.
• Mistakes, deviations or missing data are removed, changed or replaced in the source collection / system. All changes are followed up in the dependent collections / systems.
• The data collector takes care that reference collections meet the expectation of the data consumer.
• Data are stored in a reasonable timeframe following collection.

Data Processing / Data Custodian
• Data Custodians are sufficiently trained to manage the collections / systems.
• Clearly defined processes, which regulate, document, and monitor data processing are in place.
• Changes in data contents and / or structures can be traced to the originator of these changes.
• Processing of data assures data and information consistency.
• Processes for saving, transforming and enriching of data are clearly described.

Data Usage / Data Consumer
• Data Consumers have clearly defined their requirements for data quality.
• Data Consumers are aware that their requirements concerning data quality were taken into account.
• There is a communication plan for the development and implementation phases of the project.
• There are mechanisms in place to keep stakeholders informed of developments in the production phase.
• Data Collectors, Custodians and Consumers have come to a joint agreement concerning the measurement criteria for monitoring data quality.
• To encourage co-operation and data quality feedback, Data Collectors, Custodians and Consumers regularly meet to discuss their issues.
• Data quality feedback issues as noted by users are fed back to the Data Custodians.
Data Quality Management, Measurement, Measures, and Monitoring

Even though data models and data elements are well described in their early life cycle the management of data quality, its measurements, and agreements upon these measurements, are essential for the data collection in the future. The understanding of data and their logical relationship often diminishes and / or changes over time. Thus the requirements for data quality measurements have to be defined as early as possible and be as flexible as possible in the project stage to meet future expectations.

The requirements for data quality should be described in a comprehensible way and measurement of data quality should be undertaken in a reproducible way. The measurement of data quality is strongly correlated to the needs of the data consumers and can differ from user to user on the same data collection. Hence requirements are to be defined from a business point of view. These requirements are described as ‘Business Rules’ in this document. Business Rules can check a data collections’ content, matching, or unique key definition. It may be necessary to define content, key, and matching rules for data collections.

The management of data quality means the measurement of data quality, the deduction of measures (reactive as well as preventive measures), the verification of errors, the formulation of measures, and the controlling of measures for their effectiveness.

All activities applied to data elements on an atomic level may also be applied to metadata. Metadata describe to some extent the requirements of data elements and are therefore an important partial criterion for the quality of data elements.

Basic Definitions

Reconciliation Methods
Reconciliations can be done by:

- internal consistency checks within one data collection (intrinsic data quality);
- checks against internal or external reference / leading collections (e.g. if data are held redundantly);
- checks among data collections.

Visual checks are not an appropriate way to reconcile data.

TDQM cycle
The ongoing process of improvement of data quality and the measurement of data quality is illustrated in the TDQM cycle:

Define Define the objects, data quality metrics, data quality criteria, and data quality goals.

Measure Quantify the results, execute the measurement, report and document the results.

Analyse Assess by error verification and prioritisation, and measure controlling.

Improve Improve data quality by measure coordination.

Profiling
In large data collection, where visual checks are no longer an appropriate way to identify new errors, structured profiling techniques should be used due to several reasons:

- changes in the IT landscape and/or organisation landscape requirements;
- identified data conflicts;
- identified data ambiguities;
- checking data for business rules.

Unknown data model
- checking data for (unique, primary) keys;
- checking data for dependencies.

Different analysis methods are widely used:
- plausibility checks;
- data mining;
• descriptive statistics;
• data segmentation (modelling).

Checklist:

Frequency

• Frequency of checks, analysis, and profiling of data is defined and documented.
• Data are frequently checked / analysed / profiled appropriate with respect to their usage and agreed upon the process participants.

Data (Quality) Checks

• Quality of data collection is measured regularly with the Ministry of Health Data Quality Evaluation Framework.
• Validity checks are done for each variable at source and at each transfer to data with respect to the needs of the data consumer.
• Reconciliations to leading / reference systems are carried out regularly at source with respect to the needs of the data consumer.
• Data validation and reconciliation are described with respect to the questions who, where, when, and how are they done.
• Measurement of data quality
  o Data quality is measured regularly with respect to the needs of the data consumer and expressed in the formulation of business rules.
  o Business rules to check data are defined (e.g. in types such as content, key, and matching rules), documented, and accessible.
  o Business rules are applied for data checks and data cleansing at source.
  o Data collection tests account for defined business rules.
  o Data collection tests are carried out frequently appropriate to the needs of the data consumers.
  o A standard set of data quality metrics is defined.
• Profiling of the data collection is frequently undertaken due to agreement upon the process participants.

Measures and Monitoring

• The data collector, the data custodian, and the data consumer verify data quality problems.
• Identified data quality problems are addressed through reactive and / or preventive measures.
• Measurement of data quality, verification of possible errors, measures, and controlling of measures are steps in an implemented TDQM cycle.
• Derived reactive and preventive measures are documented.
• Derived reactive and preventive measures are subject to a follow up process and proved for their effectiveness.

Technical Requirements (Data Quality Related)

All national data collections should act as an integrated and consistent database for decisions. Up to now there is no overall data model in place, which describes the interactions and dependencies among the data collections. It is essential to the quality of the whole data collection, that the logical relationship among the data collections are reflected within the database e.g. by referential integrity.

Checklist:

Systems

• Systems managing data collections are tested after the last revision.
• Changes to the data collections are explicitly named in the release documentation.
Data Model and Metadata Model
- NZHIS data integrity rules (e.g. referential integrity) are assured in the database.
- Common data elements across the Ministry are defined consistently.
- Codes as of the NHI or HPI are used to uniquely identify persons.
- The history of data changes is recorded and is accessible.
- Metadata sufficiently define the data structures and are available to the whole of the sector.
- Data are normalised to an appropriate level.
- Derived variables are maintained on the main database as well as the original variable.

Data Flow Diagram
- A detailed documented data flow diagram, which shows all interfaces and collections related to a process, is available, accessible, and up-to-date.
- The critical points and limitations in data flow are known.
- The critical points and limitations in data flow are continuously supervised.

Linkage to other (National) Data Collections
- Linkages and dependencies to other (national) data collections are known, documented and the documentation is accessible to the whole of the sector.
- Redundant collection and the storage of data already existing in leading national data collections (such as NHI) is avoided.
- If redundant collection and storage of data is necessary, reconciliation processes are defined and carried out.

Interfaces
- Data structures of interfaces for the transfer of data are clearly defined, described, and agreed upon by the process participants.
- Interfaces are preferably electronic. The use of manual interfaces is justified.
- Standard manual and electronic data submission forms and procedures are used.
- The processes regarding manual interfaces are described in a comprehensible way and agreed upon by the process participants.
- Interfaces are described with respect to textual, logical, chronological, contextual, and quality dependent requirements.

Data Access
- Data can be easily accessed in accordance with security and privacy legislation
- Data are accessible by using standard tools e.g. ODBC, JDBC.
- Data access is supervised by authentication processes and logon history files.

Data Correction
- Data load / update / delete processes (manual as well as electronic) are clearly described.
- Data can be easily corrected.
- Corrections to data are traced and log files are implemented.
- Log files are stored and regularly reviewed.
Documentation

All requirements to the Data Quality Principles for National Health Collections are to be documented and the documentation should be available to all interested stakeholders. Some documents are over and above those required as part of standard project documents.

From a data quality perspective the relevant documents are:
- Business Case;
- Project Plan;
- Service Level Agreements;
- Process Descriptions;
- Organisation Model;
- Data Quality Activity Plan (template available, to be developed with the NZHIS Clinical Analysis Team);
- Technical Documentation;
- Data Dictionary.

Summary

The Data Quality Principles for National Health Data Collections provides a set of requirements, which allow the project manager to evaluate the status of new data collections in terms of data quality needs.

Seven areas of interests are examined:

- Management Requirements;
- Service Level Agreement;
- Organisational Specifications;
- Process Definitions;
- Data Collection, Processing, Data Modelling and Usage;
- Data Quality Management, Measurement, Measures, and Monitoring
- Technical Requirements (data quality related).

Not all requirements will be applicable to all data collections. It is up to the assessor to select the ‘right’ applicable requirements.
Appendix 7: Data Quality Activity Plan Template

Data Quality Activity Plan (Template)
**Background**

**Purpose**

This document provides an Activity Plan for instilling good Data Quality in the development and implementation phases for a new data collection.

The Data Quality Activity Plan is not identical with the ‘Data Quality Principles for National Health Data Collections’, which goes much further into the project’s details, but the guidelines will assist in the development of the action plan.

Assuming a typical project outline, data quality activities run throughout the project, as illustrated in Figure 1 below. The sketch shows the importance of data quality for the production phase of a product / new data collection.

![Figure 1: The project phases in relation to data quality.](image)

**Alignment**

Ensure the business processes for the new collection consider the Data Quality Evaluation Framework, from which data quality will be assessed following implementation of the collection.

**Scope**

Please explain what is not delivered in the Data Quality Activity Plan. Please explain which other documents contain data quality related topics. Refer to the ‘Data Quality Principles for National Health Data Collections’ (e.g. requirements for standard project documentation like the business case, or the business plan).

The Data Quality Activity Plan should define all activities undertaken to ensure a high quality data in all phases of the development of a new data collection. Hence this document will change over the project lifetime. Please adjust the scope of this document to this task.

**Documentation**

**Data Flow Process Documentation**

The data flow process documentation provides the standard that defines what takes place within each process. It utilises a cross-functional data flow process that details the roles and the technology responsible for each process step.

**Overview of Process Documentation**

Documentation Presentation

**Data Flow Chart**

The main goal of a data flow chart (Figure 2) is to illustrate the action points where data quality should be addressed. The focus hereby should be on the business perspective rather than the IT point of view.
Please provide a generic overall data flow chart. The data flow chart could be illustrated like this:

![Data Flow Chart](image)

Figure 2: Data Flow Chart

Please provide within the data flow chart all data sources, all data users, all additional / reference data etc. used.

This Data Flow Chart can be used to illustrate activities to achieve a good data quality.

**Data Collection, Data Processing and Data Usage**

Please clearly define who is responsible for collecting, processing and using the data. This is important to know for all phases of the project including the production phase.

**Business Rules**

Measurements of data quality should be undertaken on the basis of the business rules. These business rules should be specified through a common agreement among data consumers, data custodians, and data collectors. Use these business rules to carry out:

- internal consistency checks;
- reconciliations;
- profiling;
- descriptive statistics;
- data validation (e.g. at the source).

Please identify in each of the following sections the critical points and the points where to measure data quality on a field level on a data flow chart. To illustrate it you could use the data flow chart provided (Figure 3).
Data Collection

Clearly outline the data collection processes.

Data Processing

This section provides information from technical point of view.

Data Usage

Clearly define the proposed uses for the data – who are the consumers and what are their requirements?

Data Access and Data Correction

For daily work, but also for trouble shooting reasons, the data should be easily accessible (e.g. by an online system) at an appropriate point in the process. If a problem with data is found late in the process chain, data consistency over the whole process should be secured.

Wherever data have to be assessed in the collection step, the processing step, or the usage step, the processes have to be specified.

Data Access

Please describe where and how data are accessible.
Please describe the authentication process.

Data Manipulation

Please describe where and how data are manipulated. Please describe the log file process.
Change Process

Please describe how changes in the future e.g. due to law changes, new systems, new required data elements, extension of database, and migration of data, will be managed.

How will subsequent changes be incorporated into the data collection? Describe the process for making changes.

Data Quality Process

For an ongoing good level of data quality it is essential react to data quality problems. This can be secured by implementing the TDQM Cycle of define, measure, analyse and improve (Figure 4), which should be executed regularly.

The following sketch illustrates this cycle with the major actions in every point.

Test Phase

Although dealing with the implementation of a new data collection or changes to existing systems seems to be a mainly IT driven topic, the business experts (clinicians as well as the data user) have to be involved in all phases, including the test phase of the project. In this test phase the data quality requirements for the data collections should be checked.

Please define on the basis of the specified business rules which results are expected to verify good data quality. The data users should sign the results of the tests (and / or any person interested in the results e.g. the clinicians).

Evaluation of the Testing

How will the testing be evaluated, who will do the evaluation and at what point?
Audit Process

Assess / audit the new data collection with the ‘Data Quality Principles for National Health Data Collections’ as a guideline.

Define clearly how the audit will be undertaken, how often, and by whom. What is the method for providing feedback to data suppliers?
Purpose of the Document

This document offers representatives from the health and disability sector the opportunity to consider and comment on the structure of the proposed Data Quality Improvement Strategy, which is outlined below. The document is the basis for four regional workshops on the subject and for written comments. Questions that will be discussed at the workshops are interspersed throughout this document to enable workshop participants to solicit opinion from their organisations before attending the workshops, and guidance for those providing written comments.

We would be interested in feedback from anyone who works in health care management and data management, anyone who uses data, data quality managers, clinicians, clinical coders, administration managers and any other interested parties within the health and disability sector.

About the Consultation

This consultation aims to:

- provide information to the health and disability sector on the proposed Data Quality Improvement Strategy (and its place in the national health information management context);
- encourage feedback on the proposed strategy, validating and directing attention to areas that need amendment;
- plan and inform the feasibility of implementing the strategy in the health and disability sector.

Feedback received from this consultation will be analysed and incorporated into a revised Data Quality Improvement Strategy document. A summary of the feedback will also be available to those who participated in the initial consultation process. A further opportunity to provide written comment on the final draft document will be provided before the revised strategy is finally implemented.

Below is an outline of the proposed consultation timeline.

<table>
<thead>
<tr>
<th>Consultation Type</th>
<th>Where</th>
<th>When</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regional workshops</td>
<td>Auckland</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Waikato</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wellington</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Christchurch</td>
<td></td>
</tr>
<tr>
<td>Written submissions accepted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Written submissions on final strategy proposal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Release of Data Quality Improvement Strategy document for implementation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Why Contribute?

Poor quality data are a significant barrier to effective service delivery and planning and limits the ability to monitor policy implications. Providing high quality information assists in providing efficient and cost-effective care and improving health outcomes, patient safety and customer satisfaction. The Data Quality Improvement Strategy seeks to provide the health and disability sector with a common understanding and philosophy for data quality.

The Process to Date

The proposed strategy was developed from:

- a review of the current literature on data quality theory;
- a review of current practice within the Ministry and the health and disability sector;
- discussions with international colleagues on current practice in data quality in health care overseas;
- the development of a cross-Ministry data quality group (MDQT)
- two focus group discussions on the development and implementation of the Data Quality Evaluation Framework (DQEF);
• the implementation of the DQEF within NZHIS;
• a workshop with Ministry staff to review the proposed programme and develop priorities for data quality improvement;
• discussions with sector stakeholders.

This work has provided the Ministry with the information required to plan extensive but feasible improvements through a phased approach.

**What You Need To Do**

Please review the consultation document. Questions relating to the principles of the proposed Data Quality Improvement Strategy are interspersed throughout the document to prompt your thinking. Please consider your answers to these questions and any further comments before attending one of the regional workshops. A full list of the questions can also be found at the end of the document.

If you prefer to make a written submission rather than attend a regional workshop, emailed or hard copy submissions will be accepted up to May 2 2005. Respond in the space provided below each question and send the whole document as an attachment in a return email to: karolyn_kerr@nzhis.govt.nz

Your feedback is appreciated, and we look forward to working with those of you who are able to attend the workshops.
Executive Summary

The Data Quality Improvement Strategy seeks to improve health outcomes by making high quality information available to all those who make decisions at any point in the health and disability sector. From the clinician to the policy developer, we all depend on, and need to trust, this information reflecting reality. Health care providers and consumers are demanding access to more accurate, high quality information for better decision-making.

The Strategy aligns closely with the action zones outlined in the draft Health Information Strategy for New Zealand. The sharing of data between providers requires a good understanding of the context in which data was collected and all who use the data require accurate data quality information.

High performing health care organisations distinguish themselves by backing up their words with concrete actions. They make quality a continuous, living process of drilling down into data, feeding information back, tracking progress and making the needed changes. The Data Quality Improvement Strategy will provide the health and disability sector with a common understanding and philosophy related to data quality and seeks to instil Total Data Quality Management (TDQM) throughout the sector. TDQM adapts the widely used Deming Quality Cycle of defining, measuring, analysing and improving (Deming, 1982) to manage data throughout its entire life cycle. Data quality improvement relates to all processes applied to data, from its collection through to the information product derived from that data.

The Data Quality Improvement Strategy includes the development of processes to prevent systemic data quality errors in all existing and proposed new national health collections managed by the Ministry of Health. While the scope of this Strategy is limited to the national collections, health care organisations may wish to use the Strategy for the management of all their data collections. Business Units and Directorates within the Ministry of Health will adopt the Strategy to ensure their data, when fed into the national collections, meets quality requirements. The Strategy is expected to expand beyond the national collections in future years following a review to discover what improvement initiatives were most effective.

The Strategy provides an outline for managing data to prevent data quality errors at the source of collection and to maintain the original meaning of the collected data as it moves throughout the health and disability sector. This requires viewing data quality from a holistic perspective – going beyond an assessment of accuracy to consider also the relevance, timeliness, comparability, usability and security and privacy of data.

As data quality affects everyone in the sector, the whole sector is responsible for maintaining and improving data quality. The role of the Ministry is one of leadership and support, while data collectors need to employ all possible processes to ensure only high quality data are collected, using agreed national and international standards, where available. Data quality needs to be the responsibility of high level managers in an organisation to ensure the entire organisation makes the required changes for improvement.

All too often data quality is seen as something that is the responsibility of informatics staff alone and is often seen with disinterest by clinicians and managers, despite being so critical to the quality of the decisions they make (Data Remember, 2001).

The Health Information Standards Organisation (HISO) will support data quality through the endorsement of nationally agreed standards following extensive sector consultation.

The initial projects proposed in the Data Quality Improvement Strategy will provide for better information on our data quality and for prioritising improvements. Projects that impact the sector include the development and implementation of sector wide best practice guidelines for data quality management processes and the development and implementation of an accreditation process for ‘good data suppliers’ through the Quality Health New Zealand accreditation process and/or the Royal Colleges process. It is expected that data quality information gathered across the sector will enable a collaborative working party to develop the best practice guidelines to further enhance the improvement of our data quality.
Vision: Where do we want to go?

Future State Vision

The Data Quality Improvement Strategy has the following vision.

*High quality health information will be accessible to all appropriate users through the application of total quality data management principles across the health sector. Data suppliers and custodians will consistently assess data quality, with continual improvements made in accordance with the assessment results.*

Continuous quality improvement will ensure that the Ministry of Health has an extensive understanding of the health and disability sector’s data quality needs and issues and is able to provide leadership on ground-breaking data quality processes. Encouraging trust in the data that is collected and held throughout the sector will ensure that data are used to its greatest possible benefit. Data that are highly utilised for a variety of reasons will improve in quality. The Data Quality Improvement Strategy will ensure consistency across the health and disability sector in data management practices and will provide coherent direction towards Total Data Quality Management (TDQM). The strategy seeks to provide the sector with a common understanding and skill set related to data quality.

Extensive data mining, combining currently disparate collections, will provide more comprehensive information and extend our knowledge of our populations’ health, impacting health outcomes through better decision-making and assessment of the results of those decisions. Policy developers and decision-makers within the Ministry of Health and the health and disability sector will have high quality information available to better inform their decisions. The ability to accurately predict service requirements through better information reduces the risk to health care organisations that they fund inappropriate services for their population. Managers, clinicians and researchers will also have the ability to assess the outcomes of those decisions through improved data and information management.

The Data Quality Improvement Strategy seeks to provide:

- better decision-making to target areas of improvement in national health data collections, with a transparent prioritisation decision tool for data quality improvement projects;
- improved relationships with data suppliers, developing a whole-of-sector responsibility for data quality;
- improved awareness of a data quality culture throughout the sector;
- improved understanding of the processes involved in developing an information product;
- minimum data quality requirements for existing and new national collections;
- minimum requirements for regular operational data quality initiatives across all national health collections, with TDQM principles applied to provide feedback and support to data suppliers on persistent data quality errors;
- the development of best practice guidelines;
- education and support to data suppliers;
- the potential to develop an accreditation process for data suppliers.

The Data Quality Evaluation Framework (DQEF) will be used on a regular basis to assess the quality of data in each national health collection. The DQEF will also be used to assess the progress of data quality after improvements have been initiated, thus confirming their effectiveness. Initiatives are to be based on a measure of priority derived from the critical impacts of the health collection on the Ministry’s work.

New proposed collections will undergo an assessment of the technical specifications and collection and management processes to ensure good data quality processes are instilled in the development stages.

Improved relationships with data suppliers through feedback loops and education programmes within the Ministry and out in the sector will further improve data quality supplied from the sector. Data quality will be the responsibility of the whole sector. Responsibilities within each organisation type (that is, the Ministry, DHBs and PHOs) will be clearly assigned to ensure appropriate accountability.

Within the guidelines of the Privacy Act, data will be mined from collections to provide the best information. This will be made possible through the assessment and improvement of business process, rules and data definitions to ensure consistency and accurate mapping across the Ministry and the health and disability sector.
Aims of the Strategy

The Data Quality Improvement Strategy seeks to instil Total Data Quality Management (TDQM) throughout the health and disability sector. This is a continuous cycle of proactive management of data to ensure only high quality health information is available to all. The Data Quality Improvement Strategy views data as a product and this is fundamental to TDQM. The formulation of the information product can be managed at every point along its transition, much like a manufacturing process, from collection at source to its inclusion in an information product.

The Data Quality Improvement Strategy aims to:

- ensure the flow of information through the sector does not change the information’s meaning;
- prevent poor quality data entering the system through appropriate business rules, validation checks, entity checks, etc, thereby reducing the costs resulting from rework due to poor quality data;
- educate and assist data suppliers to send in only good quality data by applying the TDQM principles that see an information product managed at every point along its transition, from data collection at source to its final amalgamation into an information product;
- educate Ministry of Health staff on TDQM principles and apply those principles in all processes related to the custodianship of national data.

Data Quality in the New Zealand Health Sector

A current state analysis of perceived data quality levels and ‘business as usual’ improvement initiatives within the Ministry of Health was undertaken late in 2003. Results show inconsistencies across the national health data collections in almost all areas of data management. The analysis was limited to national collections managed by the Ministry of Health. Inconsistencies were found in the areas of:

- collection documentation
- knowledge of data production processes
- knowledge of current data quality procedures.

In the Ministry’s Information Systems Strategic Plan (ISSP), Ministry staff stated the most common reasons for inadequate quality in existing data collections were:

- multiple sources of data;
- data entry errors and insufficient validation;
- ambiguous nature of data;
- insufficient definition of data (metadata);
- acquisition of faulty data from external sources;
- combining good data with bad data.

This ISSP and the current state analysis on data quality support the development of a more structured and strategic approach to data quality management.

Scope of the Data Quality Improvement Strategy

Scope Inclusions

The Data Quality Improvement Strategy includes the development of processes to prevent systemic data quality errors in all existing and proposed new national health collections managed by the Ministry of Health. While the scope of this Strategy is limited to the national collections, health care organisations may wish to use the Strategy for the management of all their data collections. Business units and directorates within the Ministry of Health will adopt the Strategy to ensure their data, when fed into the national collections, meets quality requirements.

The Data Quality Improvement Strategy is expected to expand beyond the national collections in future years following a review to discover what improvement initiatives were most effective.

A national collection, as defined in the draft Stewardship of National Collections Framework is:
A long term collection of nationwide data or reference data set, of which the Ministry (NZHIS) is the custodian on behalf of the sector, and which is used for analysis and information provision to achieve improvement in the services and capabilities of the publicly funded health and disability sector.

The strategy’s scope includes processes along the continuum of data collection at source through to the publication of information products. Therefore, as sources of the data, health care providers are included in the strategy’s scope. Not all data runs through the steps highlighted below, but this is the general flow of data from its source through to the development of an information product.

<table>
<thead>
<tr>
<th>Coding Classification Terminology</th>
<th>Transactions Databases</th>
<th>Information Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Source</td>
<td>Transfer to Ministry</td>
<td>Warehouses</td>
</tr>
</tbody>
</table>

All business units within the Ministry are included in the Strategy’s scope.

The strategy also identifies the importance of educating data suppliers about the Ministry’s expectations on collection processes as well as clearly defining the roles and responsibilities across the health and disability sector. Health software vendors are also included as their co-operation is paramount.

All publicly funded health care provision requires reporting to NZHIS and this data are included in the scope.

**Scope Exclusions**

Data suppliers are expected to manage their own data quality initiatives, aligning them with the Data Quality Improvement Strategy. However, the Ministry of Health will provide support, leadership and guidance. Strategies should be managed within existing funding available.

**New Zealand Health Strategies: The Context of the Data Quality Improvement Strategy**

Further to the WAVE (Working to Add Value through E-information 2001) recommendations, the New Zealand Health Information Strategy will provide a context for increasing our information systems capability to share information across funders and providers. The Data Quality Improvement Strategy aligns closely and supports the action zones outlined in the draft Health Information Strategy for New Zealand. The sharing of data between providers requires a good understanding of the context in which data was collected and all who use the data require accurate data quality information.

The focus on improved information requires improving data quality management to address all of the strategy’s action zones, particularly to improve data quality on the National Health Index (NHI), and subsequently reducing duplication.

The following strategies and policies also define a need for active data quality management, including better access to information:

- Primary Care Strategy;
- Improving Quality (IQ): A systems approach for the New Zealand health and disability sector
- New Zealand Disability Strategy.

Data quality improvement is an integral stream across many health strategies. The Data Quality Improvement Strategy seeks to meet the requirements outlined in the above documents through a whole of sector approach.
International Data Quality Strategies in Health

A review of international data quality improvement strategies in health care, including the National Health Service (NHS) in the United Kingdom, the Canadian Institute for Health Information (CIHI), HealthConnect Australia and the United States Department of Health and Human Services identified similarities between the strategies. All the reviewed international strategies note the complexity of data quality improvement in their multi-level, multi-dimensional initiatives. Principles are based on the need to manage data proactively to ensure ongoing integrity by preventing data quality problems, using a systematic Total Data Quality Management (TDQM) approach.

There is also commonality of role expectations – the data suppliers are responsible for the quality of the data they provide to central government, while central government is required to provide leadership and assistance to data suppliers by developing sector-wide standards and best practice guidelines.

The NHS in particular outlines clearly and in detail substantial work that is required by any provider to ensure good data quality. They developed an accreditation scheme that was initially thought to be all that would be required to ensure the supply of good quality data was supplied. The scheme is extensive and was found to be very successful but did not sufficiently identify the responsibilities of the supplier to ensure the supply of good data quality – central government was still monitoring more than it was leading. This led to the more extensive guideline development on the principles and theories of data quality supported within the NHS. Several trusts have developed and published data quality strategies on their websites that align to the NHS core strategy requirements.

The CIHI also briefly discusses accreditation for enhancing collaboration with data suppliers. They have undertaken extensive work on data quality through collaborative work with experienced statisticians from Statistics Canada and base their theories on research by the Massachusetts Institute of Technology in the United States.

Consultation Questions 1–2

Question 1
Do you agree with the aims (Section 1.2) for the Data Quality Improvement Strategy?

Yes  No

Please comment:

Question 2
Do you agree with the scope (Section 1.4) of the Data Quality Improvement Strategy? For example, should the Strategy be restricted to the national health data collections?

Yes  No

If no, what changes would you suggest?
Strategy Approach

What is Data Quality?
Modern definitions of data quality have a wider frame of reference and many more attributes than the obvious characteristics of accuracy. Data quality takes a consumer-focused view (consumers being people or groups who have experience in using organisational data to make business decisions) that quality data are ‘data that is fit for use’. Data quality is ‘contextual’; the user defines good data quality for each proposed use of the data, within its context of use. Therefore:

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 (Wang & Strong 1996).

It is particularly important to note that a data quality improvement strategy is not just ‘information technology’ or ‘information systems’ focused. Although such a focus may provide insight and tools to help improve data quality, data quality improvements cannot be attained merely through information technology; the problem is one of processes and people. As noted in Ward and Peppard:

… Clearly, technology on its own, no matter how leading edge is not enough. (Ward and Peppard, 2002).

The first steps in any improvement process must be to identify how the data are used and by whom. Improvements also need to look to the future potential uses of the data.

Total Data Quality Management (TDQM)
Total Data Quality Management (TDQM) seeks to deliver high quality information products to information consumers and to facilitate the implementation of an organisation’s overall data quality policy. The TDQM cycle adapts the widely used Deming Quality Cycle (Deming 1982).

Figure 1: Components of the TDQM Cycle

TDQM aligns closely with the theories discussed in ‘Improving Quality (IQ): A systems approach for the New Zealand health and disability sector’ (Minister of Health 2003).

Alternative solutions to data quality management often involve a projects approach. This approach targets specific errors within a collection or an entire collection but often does not devise solutions to prevent systemic problems. The foundation of the Data Quality Improvement Strategy is institutionalisation of TDQM principles. Institutionalisation means encouraging data quality management to be accepted as ‘business as usual’; a regular process that is undertaken at all levels in the sector.
The Data Quality Improvement Strategy views data as a product, and this is fundamental to TDQM. In order to manage information as a product, organisations should:

- know their customers/consumers of the information and their information needs;
- manage the information as if it were the product of a well-defined information process, which includes considering the technology and the organisational culture;
- manage the entire life cycle of their information products;
- make management accountable for managing their information processes and resulting products.

An information product within the Ministry of Health could be any one of the national health collections, such as the National Minimum Data Set (NMDS) or the National Cancer Registry. In DHBs, information products could take the form of collections used for reporting purposes (that is, to NZHIS) or collections used for claiming purposes (that is, to HealthPAC).

The four steps of the TDQM cycle – defining, measuring, analysing and improving are discussed below in more detail.

**Defining**

Defining what data quality means to your organisation, then the level of data quality required by those who use the information, can be a difficult task. There are several requirements as listed below.

- Define the functionalities of the information product. This is the information needed by the information consumers.
- Define the characteristics of the information product. Information product mapping shows us the ‘critical path’ that data takes as it develops into an information product and helps define the information product.
- Define the basic unit, components and their relationships for the information product through an entity relationship (ER) model. The ER model outlines the relationships between different data entities. For new data collections, Quality Entity Relationships (QER) can define where data quality can be incorporated into the database design (Wang et al 2001).
- Define the data requirements from the perspectives of information product suppliers, manufacturers, consumers and managers.
- Prioritise the data quality dimensions for the information product. This could include ranking, weighting or applying a trade-off to decide what dimensions are of the highest priority for an individual or group of information products.

**Measuring**

The actual level of data quality found in data collections should be measured using both objective and subjective tools to continually and consistently monitor ‘where we are at’. The Data Quality Evaluation Framework (DQEF), as discussed below, enables us to look at data quality from a wider perspective than just accuracy using both subjective and objective measurements.

Statistical Process Control (SPC) measures subjectively monitor outliers within a collection by analysing trends, particularly where historical data are available (Carey & Lloyd, 1995).

**The Data Quality Evaluation Framework**

The aim of the Data Quality Evaluation Framework (DQEF) is to make consistent and accurate assessments of data quality in all national health data collections held by the Ministry of Health. The framework can also be applied to data collections outside the Ministry.

Seminal works in the field of quality have defined various extensive frameworks that all seek to assess areas where poor quality workmanship or inefficiencies may reduce the workflow of an organisation. At its most basic, the DQEF is a tool for assessing data quality within an organisation. The framework can go beyond the individual elements of data quality assessment to become integrated within the processes of the organisation.
The DQEF for the New Zealand health and disability sector is defined as:

**A point in time assessment and measurement tool, integrated into organisational process, providing a benchmark for the effectiveness of any future data quality improvement initiatives and a standardised template for information on data quality both for internal and external users.**

The New Zealand DQEF has been developed to provide a standardised approach to assessing the data quality of all health information databases and registries. It is used to identify and measure major data quality issues, standardise information on data quality for users and identify priorities for data quality improvement, all of which perpetuates improvements.

**The Data Quality Dimensions**

Six data quality dimensions are outlined in the DQEF, with accuracy being the most extensive. The six data quality dimensions are:

- Accuracy
- Relevancy
- Timeliness
- Comparability
- Usability
- Security and Privacy

Together the dimensions offer a more detailed interpretation of data quality. The relationships between the dimensions of data quality and health care data are outlined in Figure 2 below.

![Data Quality Dimensions and Health Care Data](image)

**Figure 2: Data Quality Dimensions and Health Care Data**

Each dimension is broken down into its characteristics. The framework requires these characteristics to be optimised using TDQM techniques.

The data quality dimensions in the DQEF provide the fundamental foundations of data quality and should be viewed in their entirety. A DQEF should not omit any one of the six dimensions. While different data may emphasise the dimensions differently and each characteristic under a dimension may not always be applicable, the six dimensions have an equal importance in ensuring that TDQM principles are fulfilled.

**Analysing**

The processes around data collection and management and information systems themselves should be analysed to identify data quality problems and where such problems commonly originate.
Improving

Data quality improvements are based on an organisation’s priorities and are informed by the previous three steps of the TDQM cycle.

What are the priorities for change? These will be different for each organisation, but the following questions should always be considered.

- What are the uses of the data, now and in the future?
- Do the identified errors impact on clinical decision-making?
- What improvements will provide us with the best return on investment?
- What improvements can we afford to make in the immediate/medium/long term?
- Do we have the skills available in-house or through outsourcing to make the improvements?

An assessment should be made of the effectiveness of any improvement. This can be done using the DQEF. When applied consistently, the framework provides feedback on the effectiveness of data quality improvement initiatives. An organisation should continue to use existing data quality measures where these are found to be effective. A core set of data quality improvement activities should be in place for all collections, but for projects where the assessment phase highlights a need, such activities can focus more specifically on developing and improving data collection processes.

Less Rework – Prevention is Better Than Cure

The strategy suggests the TDQM approach to improving data quality. Alternatives to this approach could include the traditional approach of fixing identified errors in the data once they are found. Up to now data quality management has taken a less preventative approach, apart from the inclusion of efforts such as business rules and format restrictions on data entry. While this kind of work will always be required, the volume of data now generated and the increasingly various uses made of that data, mean that identifying and fixing errors is becoming increasingly difficult and ineffective.

Extensive work can prevent systematic and repeated data quality problems. Mere human error leading to problems with accuracy may be more difficult to prevent but can be addressed through targeted education. TDQM looks at the processes that data flows through before ending in an information product, making sure that none of those processes change the initial meaning of the data. Some of the time currently spent daily on fixing errors already in data collections would be better channelled towards the systematic prevention of errors occurring in the first place.

Roles and Responsibilities

All of Sector Responsibility

Data quality improvement is a cultural change where all those who work in the sector take responsibility for collecting data of the highest possible quality. This requires a skilled workforce, where all are aware, through education, of the implications of poor quality data and make changes to work practices that reflect this understanding. Good information technology systems provide the tools staff require to collect, manage, and maintain data of high quality.

The Data Quality Improvement Strategy outlines the need for clear roles and responsibilities in all organisations in the health and disability sector, with overall responsibility for data quality occurring at the highest level of the organisation, where change can be effected. For data quality initiatives to succeed in any organisation there must be full commitment from senior management. Further, the accountability for data quality levels needs to sit with senior managers and could be measured through Key Performance Indicators (KPIs).

There are three defined roles relating to data quality – the data custodian (for example, Ministry of Health, shared services agencies, DHBs, PHOs), the data collector (for example, DHBs, PHOs, GPs, and NGOs) and the data consumer (for example, researchers, policy developers, and funding and performance managers).
Another key requirement is that data quality is the responsibility of the data supplier. In the health care environment, this is particularly important. Once the data has left the supplier, it has moved from the source, the only place where the ‘truth’ can be verified. It would be unethical and not best practice, for organisations that hold data centrally to make assumptions on that data.

**Health Information Standards Organisation (HISO)**

The Health Information Standards Organisation (HISO) plays an important role in data quality by endorsing national e-health standards. The Minister of Health established the HISO in June 2003 to champion and facilitate the development of New Zealand Health Information Standards. The HISO provides a forum where key sector interests can be represented to plan, develop and promote the health information standards needed to implement e-health. E-health involves increasing the quantity and type of transactions that can be undertaken electronically.

The development of agreed health information standards will provide for better quality information. Standards allow for the introduction and improvement of health care delivery projects that have an information management component and reduce compliance costs for data suppliers. The ability to transfer data throughout the sector is greatly increased, reducing the requirement for subsequent mapping of data to different coding or classification standards when reaching its destination.

Data suppliers and custodians are required to use HISO endorsed health information standards (including the endorsed version of the standard) where available.

**The role of the Ministry of Health**

As the custodian of the data collected throughout the health and disability sector for the national health collections, the Ministry of Health holds the mandate to manage that data. The Ministry is responsible for maintaining the integrity of the data it receives. The Chief Advisor, Health Information Strategy and Policy owns the Improvement Strategy and is responsible for any subsequent review requirements.

**The role of NZHIS**

NZHIS is to be the sector and Ministry leader in data quality, providing expert advice and leading through example. NZHIS will:

- undertake independent research and consult with experts in areas such as data collection, data analysis and data management
- maintain ongoing contact with users and participants in conferences and workshops in order to objectively assess the current and future information needs of the sector
- review the quality of information products before they are disseminated
- treat data quality as integral to every step of the development of information, including its creation, collection, maintenance and dissemination (TDQM)
- make its statistical and analytic information products widely available and broadly accessible
- identify critical information with known quality problems that have significant negative impacts on the health and disability sector, adhering to information privacy legislation.

**The Data Quality Team at NZHIS**

The manager of the Data Quality Team at NZHIS reports directly to the group manager, as do managers from the Business Intelligence Unit, Clinical Coding Services, the Operations Team and the Chief Advisor, Health Information Strategy and Policy. Each team has considerable impact on the quality of data management for the health and disability sector, and these teams work closely together. However, all require expert advice if the organisation and the sector are to make significant and effective improvements in data quality levels.

The Data Quality Team will supply the Ministry of Health and the sector with expert advice by employing a subject matter expert. The subject matter expert will, with the assistance of the Data Quality Team, apply consistent TDQM practices to ensure:
• data suppliers are informed of major data quality issues on a monthly basis
• data suppliers are provided with expert advice on possible solutions to data quality problems
• regular ‘business as usual’ strategies are in place that assess the quality of data on all national health collections, with effective processes to make prioritised improvements where necessary and possible, in a timely manner and aligning with existing changes processes
• prevention mechanisms are in place, where possible, to confirm that data quality is not degraded once it is received within the Ministry. This includes providing advice to projects developing new collections
• the provision and monitoring of KPIs to suppliers. This allows the sector to understand the whole of sector minimum requirements for data quality outcomes. The KPIs provide for comparative performance ratings and can be used as a diagnostic tool for learning purposes
• provision of a ‘train the trainers’ education Strategy for the whole of the sector
• data quality caveats are supplied to anyone using data to help them understand the limitations of the data.

The Business Intelligence Unit at the Ministry, the outlet for information from national collections, will ensure appropriate caveats are included when providing information to customers.

The Role of the Health Sector

National governance

The role of the governance group is to oversee the continuing implementation of the Data Quality Improvement Strategy by:

• championing the Strategy and educating others within their organisations about the Strategy
• communicating to the rest of the governance board issues that arise in the sector with regard to implementing the Strategy
• providing feedback to the rest of the governance board on local effective data quality improvement practices.

This group will be part of a larger National Collections Governance group(s)under discussion to oversee the implementation of the New Zealand Health Information Strategy.

Data collectors and suppliers

Health care providers, in their role as data collectors, are responsible for providing the sector with information that reflects reality. Health care providers actively collect, hold and use clinical and administrative data, which is then disseminated throughout the sector. The data collector has the pivotal role. Providing poor quality data from source causes a ripple effect down the line.

The data are often under the custodianship of the CIO in the IT department. Some health care providers already employ data quality coordinators/analysts within their IT departments. In other cases, this role is partially filled by many different roles within the organisation.

Accountability within organisations for data quality would most likely sit with the CIO or practice manager, although there is an argument for it being the responsibility of the quality manager where available, due to the ‘people and process’ nature of data quality work. Dual accountability may be the most appropriate solution. Managers of departments that collect data should be required to ensure the quality of that data are sufficient before submitting it to the DHB or Ministry of Health.

Publicly funded health care providers may wish to work towards accreditation as good data suppliers, once the Data Quality Improvement Strategy is in place.
Clinical coders

A clinical coder is a health care worker whose primary responsibility is the translation of narrative descriptions of diseases, injuries and procedures into alphanumeric codes, according to a predefined health classification system (Clinical Coder National Competency Standards, 1996). New Zealand has adopted the use of the International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Australian Modification (ICD-10-AM).

New Zealand implemented 3rd edition of ICD-10-AM on 1st July 2004, which is one version behind Australia. The Mortality and Private Hospital Collections and the New Zealand Cancer Registry also utilize ICD-10-AM. In addition the New Zealand Cancer Registry uses the International Classification of Diseases for Oncology (ICD-O), for classifying the morphology of neoplasms. DSM-IV is used in some DHBs for the coding of mental health information.

There are approximately 200 clinical coders working in the New Zealand health sector. All public hospitals are required under the Memorandum of Understanding between the Ministry of Health and District Health Boards (DHBs) to submit patient data for every inpatient and day-patient discharge to the National Minimum Dataset (NMDS) at NZHIS. The data provided from each DHB are used to produce national health statistics.

Coded medical record data are used within health care facilities, at a regional and national level, for the following reasons:

• casemix management
• clinical research
• health service planning and evaluation
• quality assurance activities
• funding and resource allocation
• role delineation of health care facilities
• epidemiological studies of disease distribution and trends
• education of health care professionals

NZHIS has identified the following principles for guiding the development of a data quality strategy for clinical coding practitioners.

Utilization of a triple tiered systems model:

At sector level, this is best achieved using a broad coding quality analysis tool. One example is Performance Indicators for Clinical Coding (PICQ) - this contains a series of indicators that analyse admitted patient morbidity data coded with ICD-10-AM and is based on Australian Coding Standards and coding conventions. PICQ identifies possible coding errors and could be used on a national level to regularly monitor coding quality and identify areas where addition education and training may be required.

Second tier is at the DHB Coding Unit level where data quality responsibilities are managed using either data quality software or the traditional face-to-record general audit process, or ideally a combination of both. The Australian Coding Benchmark Audit (ACBA) is already used in some Coding Units. This tool has been designed on the philosophy of continuous quality improvement and is recommended as a user-friendly system.

Third tier is at the individual coder level - all coders, regardless of qualification and experience, benefit from having their work individually audited as a means of checking progress, and to identify specific problem areas and coding difficulties. Results can then be compared over time as a measure of success and ongoing competence.

Create synergy between the three tiered systems model throughout the sector
Capture multiple dimensions of quality within the overall strategy
Include coding practice for all coding practitioners in specialized collections
Provide efficient mechanisms for data reporting and collection of data quality information
Link all data quality results to education initiatives/Strategies
Health Sector Data Quality Management

In order to instil TDQM throughout the health sector, organisations should work towards the following:

- Taking part in a sector peer review/audit process
- Meeting Key Performance Indicators (KPIs)
- Implementing an in-house data quality education programme
- Developing and implementing a local data quality improvement strategy
- Organising regular meetings with a cross-organisational data quality group
- Alignment with Best Practice Guidelines
- Work towards the accreditation process requirements.

Peer review/audit

It is appropriate that the sector monitor its own performance in relation to data quality practices and outcomes. A peer review and auditing system will provide the sector with the ability to assess data supplier practices and provide advice and learning between organisations. A cyclic peer review process, where suppliers review different suppliers each year, ensures the advice and learning is widely disseminated throughout the sector.

Peer review offers valuable information on data management processes. Those data suppliers that achieve excellent data quality results may provide the sector with valuable learning and the review process may provide the mechanism for the dissemination of good practices throughout the sector.

Key Performance Indicators (KPIs)

KPIs provide for comparative performance for learning purposes. The Data Quality Team at NZHIS will assess the performance of data suppliers monthly against the KPIs and feedback will be given in the form of a report to data suppliers. The report will contain an assessment of any data errors found. Data suppliers will be assisted, where possible, to rectify and prevent persistent, regular data errors. Data quality issues should be resolved/addressed by suppliers within a reasonable time frame. Key Performance Indicators (KPIs) will be included in the District Annual Plan requirements for 2005/06 to provide data suppliers with guidelines for minimum levels of quality.

In-house education programme

Data suppliers should provide a formal and regular education programme for data collectors and coders within their organisation. NZHIS will assist by offering a ‘train-the-trainers’ programme, with this education then disseminated throughout the suppliers’ organisations. Data collectors include clinicians at all levels and in all disciplines, as well as administrative and coding staff.

Local data quality improvement programme/strategy

The development and implementation of organisational data quality improvement programme/strategy should work towards preventing errors through TDQM. Such improvement strategies should align, where possible, with the national Strategy of TDQM, using the six data quality dimensions found in the Ministry’s DQEF as a guideline for assessment and prevention requirements. A template suggesting the content of a data quality improvement Strategy is included in Appendix 1.

Cross-organisational data quality team

A cross-organisational data quality team is needed to facilitate the discussion of data quality issues from an organisation-wide perspective. The team should include representatives of business units that collect, manage or use data, as well as clinicians and clinical coders.

The role of the team is to:

- encourage discussion across business units on data affecting the whole of the organisation
• create partnerships between IT staff and business areas
• provide management with the impetus to make changes when the same message is coming from all the business units
• provide feedback to each member’s business unit on data quality improvements or issues
• provide input into data quality policy and strategy development as well as discussion around proposed projects and the potential impact on the different business units
• identify critical information on known data quality problems that have significant negative impacts on the organisation.

The organisation’s data quality manager or an equivalent role should facilitate this team.

Best practice guidelines

The education and support of data suppliers in their collection processes requires extensive work on behalf of the Ministry, in combination with the sector, to develop applicable and achievable best practice guidelines. The information gained from existing and new experience and knowledge of data quality will be used to provide detailed guidelines to data managers on possible ways to improve data quality.

The best practice guidelines should provide its users with pragmatic, tested solutions to data quality problems commonly found in health care organisations. The guidelines should consider TDQM practices. While health care organisations will experience similar problems with their data quality, each organisation should use the sections of the guidelines that best suit their needs, not all may be applicable.

Accreditation

The Ministry, through the Quality Health New Zealand and/or the Royal Colleges accreditation process, will provide sector organisations with clear guidelines on how to achieve accreditation as good data suppliers. The accreditation process will be developed in consultation with the sector following the development and testing of the Best Practice Guidelines. Those organisations that have been able to implement extensive data quality programmes, as outlined above, will be accredited as good data suppliers. This may lead to a reduction in the need for peer review and audit processes.
Consultation Questions 4–9

Question 3
Do you agree with the six data quality dimensions (Section 2.4.2) used within the Data Quality Evaluation Framework (DQEF)?
Comments: [Yes No]

Question 4
Do you agree with the roles and responsibilities outlined for:
Comments:
The national governance group (Section 2.10.1)
Comments:
NZHIS (Section 2.9.1)
Comments:
Data collectors and suppliers (Section 2.10.2)
Comments:
HISO (Section 2.8.2)
Comments:
Clinical coders (Section 2.10.3)
Comments:

Question 5
Do you agree with the development and implementation of a cyclic peer review process, where suppliers review different suppliers each year?
Comments: [Yes No]

Question 6
Do you agree that the sector should develop and work towards Best Practice Guidelines (Section 2.12) within the next two years?
Comments: [Yes No]

Question 7
Do you agree to the development and implementation of a national Accreditation programme (Section 2.13)?
Comments: [Yes No]

Question 8
Who should manage the Accreditation programme (Section 2.13)? For example, should the accreditation process be managed through the Quality Health New Zealand and/or Royal Colleges accreditation processes?
Implementation: How do we get there?

Improving data quality is not a simple process, and no standard checklist is available to cover all situations. Each organisation needs to find appropriate solutions for their needs. The detailed projects of the Ministry’s Data Quality Improvement Programme will highlight future improvement requirements.

Phase one – 2004/05

For the financial year 2004/05, the following projects will be undertaken within the Ministry of Health. The projects all aim to improve business processes and provide better information on the data quality of the national health collections. This information will inform subsequent projects within the Ministry and the sector, improving our ability to target areas of most need.

The Data Quality Improvement Programme addresses the need to improve data quality in developing new collections, the movement of data across collections and the appropriate access to the information gained. The programme outlines requirements for new and existing collections held throughout the health and disability sector, as this data will potentially move around the sector. This work should therefore also be undertaken throughout the Ministry of Health Directorates and Business Units, where data are collected and managed, with priority given to data that eventuates in a national collection.

<table>
<thead>
<tr>
<th>Project 1</th>
<th>Data Quality Evaluation Framework</th>
<th>Justification</th>
<th>Evaluation of implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implement the DQEF and the data quality dimensions</td>
<td>• Provides an initial assessment of current data quality status&lt;br&gt;• Provides data quality information to inform the strategy development</td>
<td>All assessments documented</td>
<td></td>
</tr>
<tr>
<td>Develop a priority scale for data quality improvements</td>
<td>The process of deciding what to do first should be logical, and transparent</td>
<td>Priority scale for improvements in use on data quality assessments</td>
<td></td>
</tr>
<tr>
<td>Develop the business processes to support the framework implementation as ‘business as usual’</td>
<td>We need to continually monitor our data quality and the impact of improvement projects</td>
<td>Ensure the framework is used regularly to assess quality&lt;br&gt;Ensure action plans for improvement are developed and implemented</td>
<td></td>
</tr>
<tr>
<td>Project 2</td>
<td>Customer Satisfaction</td>
<td>Justification</td>
<td>Evaluation of Implementation</td>
</tr>
<tr>
<td>----------------</td>
<td>------------------------</td>
<td>---------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Develop a list of priority customers</td>
<td>To ensure we provide the best possible service to our customers</td>
<td>A priority list with justification is documented</td>
<td></td>
</tr>
<tr>
<td>Assist priority customers to ascertain their data quality requirements where necessary</td>
<td>Some customers may not know what they need</td>
<td>Clear requirements for customers are known, and improvement projects target these requirements where possible</td>
<td></td>
</tr>
<tr>
<td>Develop a standardised assessment tool for customer satisfaction to be used regularly in a 'business as usual' system to ensure improved customer delivery</td>
<td>We need to consistently monitor whether we are delivering to our customers and whether we are improving over time</td>
<td>Customers are regularly surveyed to assess their satisfaction with service delivery</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Project 3</th>
<th>Minimum Requirements for New Collections</th>
<th>Justification</th>
<th>Evaluation of Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Develop minimum data quality requirements for new collections</td>
<td>NZHIS currently accepts the management of new collections without input into their data quality requirements at the development stage</td>
<td>A tool is developed and used to inform new collections project managers of the requirements for data quality measures on all NZHIS national collections</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Project 4</th>
<th>Statistical Process Control (SPC) Measures and KPIs</th>
<th>Justification</th>
<th>Evaluation of Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Develop appropriate Statistical Process Control (SPC) measures</td>
<td>To understand our data quality levels To ensure improvement initiatives have an effect Analysis of historical data will enable the development of appropriate KPIs for data suppliers</td>
<td>Accurate KPIs are developed and accepted by the sector</td>
<td></td>
</tr>
<tr>
<td>Develop the business processes required to support the implementation of SPC measures</td>
<td>To ensure TDQM processes are in place</td>
<td>The expected KPIs are assessed monthly against the actual data supplied by DHBs The NZHIS Data Quality Team reports data quality information to the sector on a monthly basis</td>
<td></td>
</tr>
</tbody>
</table>

**Phase two – 2005/06**

The work undertaken in phase one will provide the health and disability sector with extensive information on our current data quality. Business processes developed by the phase one projects will be in place at NZHIS to support the sector in making changes to their own data quality practices and in developing their own improvement strategies.

The Ministry, with increased information on the quality of the data in the national collections, will begin to make Ministry-wide improvements to enable increased sharing of data, improving the Ministry's ability to understand health information for policy implications.
## Ministry of Health

<table>
<thead>
<tr>
<th>Project 5</th>
<th>Business Rules and Data Dictionaries</th>
<th>Justification</th>
<th>Evaluation of Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Review business rules and data dictionaries from a Ministry-wide perspective</td>
<td>To ensure a common understanding of data elements no matter where they are used</td>
<td>Whole of Ministry documentation of data dictionaries and business rules, available to all Ministry staff</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Project 6</th>
<th>Review NZHIS team roles</th>
<th>Justification</th>
<th>Evaluation of Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Review the roles and responsibilities of the NZHIS Data Quality Team</td>
<td>To make the best use of our resources and investigate the possibility of moving to a more preventative focus Review of phase one projects – what works, provides good data quality information and prevents data quality errors</td>
<td>The Data Quality Team actively works on strategic prevention of data quality problems, provides sector leadership and education, regularly provides the sector with accurate reports on data quality issues and assists with the prevention of persistent errors from suppliers</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Project 7</th>
<th>Standard Operational Minimum Requirements for All Collections</th>
<th>Justification</th>
<th>Evaluation of Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Develop minimum standard operational data quality initiatives that are feasible and appropriate to apply to all national collections</td>
<td>There is variation between collections on what we do to ensure and improve data quality</td>
<td>All collections have preventative and improvement measures in place All collections are regularly assessed using the DQEF</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Project 8</th>
<th>Information Product Maps</th>
<th>Justification</th>
<th>Evaluation of Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Develop data production maps for all major information products outlining flow of data around the organisation and possible areas of data quality issues</td>
<td>To examine the flow of data that ends in an information product to detect points where data quality may be impacted</td>
<td>All information products have documented data production maps, and quality touch points are identified Where appropriate, improvement projects are in place, targeting areas where data quality could be impacted</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Project 9</th>
<th>Data Quality Software</th>
<th>Justification</th>
<th>Evaluation of Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analyse the availability of data quality assessment software or the potential for in-house development</td>
<td>Some processes can be automated with reports generated, saving time on manual work</td>
<td>A report is provided to NZHIS senior management on the available software, with a cost-benefit analysis of the applications considered appropriate</td>
<td></td>
</tr>
</tbody>
</table>
### Project 10  
**Metadata for Data Quality**

<table>
<thead>
<tr>
<th>Justification</th>
<th>Evaluation of Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Develop minimum documentation requirements for each collection to ensure that the metadata repository contains adequate information for data quality.</td>
<td>To ensure the metadata repository is adequate for maintaining data quality throughout the data life cycle and provides all the data quality information required for users. The repository needs to be highly accessible to be of use.</td>
</tr>
</tbody>
</table>

### Project 11  
**Education Programme**

<table>
<thead>
<tr>
<th>Justification</th>
<th>Evaluation of Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Develop a comprehensive education programme for the Ministry and the sector, co-ordinated by one dedicated full-time employee at NZHIS.</td>
<td>An education programme that is responsive to data quality information provided by the DQEF and SPC. A train-the-trainers approach requires the sector to disseminate the training provided.</td>
</tr>
</tbody>
</table>

### The Health Sector

### Project 12  
**Cross-Organisation Data Quality Team**

<table>
<thead>
<tr>
<th>Justification</th>
<th>Evaluation of Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organise a Data Quality Team, which includes stakeholders from across the organisation.</td>
<td>To bring about discussion among business units and provide information on data quality requirements across the organisation. To create partnerships between IT staff and business areas and between business areas. Can provide impetus to management to make change when the same message is coming from all the business units.</td>
</tr>
</tbody>
</table>
Phase three – 2006/07

Ministry of Health and the health and disability sector

The following two projects are large and require considerable input from the whole health and disability sector to ensure applicable guidelines are produced for the sector. The guidelines and the accreditation scheme will be developed through sector working groups, with secretariat support from the Ministry. The guidelines will be developed from the data quality information available through the improved assessment, analysis and management of data quality across the sector following implementation of the first two phases of the Data Quality Improvement Programme as outlined above.

<table>
<thead>
<tr>
<th>Project 13 Best Practice Guidelines</th>
<th>Justification</th>
<th>Evaluation of Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Develop best practice guidelines for data quality management processes</td>
<td>Guidelines for data collectors ensure we all have the same understanding for data quality requirements and a common definition/interpretation of data quality processes across the sector</td>
<td>Guidelines are developed from the data quality information available through the improved assessment and analysis of data quality across the sector A wide range of sector and Ministry stakeholders are involved Documented best practice guidelines are available and used by the sector</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Project 14 Health sector data quality management</th>
<th>Justification</th>
<th>Evaluation of Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Develop DHB/PHO data quality programmes that include: peer review/audit best practice guidelines meet KPIs in-house education programme local data quality improvement programme/strategy cross-organisational data quality group completion of accreditation process</td>
<td>Peer review rather than a centrally led audit that does not support moving responsibility for data quality to the whole of the sector – provides for dissemination of good practice throughout sector Provides for whole of organisation and strategic view of data quality</td>
<td>Organisations have most of the requirements in place for an extensive strategic data quality programme, with input across the organisation and including management</td>
</tr>
</tbody>
</table>
Consultation Questions 9-13

Question 9
Do you agree with the phased implementation approach (Section 3.1)?

Please Comment:

Question 10
Do you agree with the Ministry and Sector projects proposed in phase three (Section 3.3)?

Please Comment:

Question 11
Is the proposed programme of work feasible for your organisation within three years (Sections 3.2, project 12 and Section 3.3, projects 13 and 14)?

Please Comment:

Question 12
If you were only able to implement some of the changes suggested by project 14 (section 3.3), which would you undertake first?

Question 13
What incentives would your organisation require to make improvements to your data quality programme?
Implications for the Ministry and the sector

Implementation Plan

The implementation of the Data Quality Improvement Programme will start within the Ministry. The change in culture around data quality requires the Ministry to provide sector leadership and support for improvement at the data supplier level. Much of the proposed work to be undertaken within the Ministry will provide the health and disability sector with data quality information that will enable the sector to target the best solutions that will create the biggest wins. The initial projects aim to develop ongoing, "business as usual", regular assessments of data quality within collections to feed back this information to data suppliers.

The programme outlines specific directions for data quality improvement for the sector for the next 3 years. Improvement programmes within organisations need not be limited to the suggested improvements and local data quality strategy development may highlight specific areas of local need.

High-level time lines are included in the projects section for each phase. At the end of phases two and three of the programme will be evaluated for effectiveness and implementation progress.

Funding Implementation

Finding the resources for improvements to data quality management in the sector rests with organisations. It is likely that projects related to the action zones in the Health Information Strategy for New Zealand will require funding through business cases. Data quality improvements should be outlined within the same business case as part of the project requirements.

Financial Impact

Costs are likely where new staff are required or a change in work practices for existing staff means training is required. Increased costs are most likely in the first year of programme implementation. Costs may be offset in the future by the ability to prevent errors and therefore reduce rework. Costs will be further reduced through the prevention of systemic errors in new collections if a thorough assessment of potential data quality risks is made and effectively addressed before implementation.

Return on investment is difficult to analyse in data quality, as improvements are likely to impact a wide range of information consumers throughout the sector. Improved data on funding and planning requirements are, however, likely to reduce waste and provide increased certainty for strategic service planning.

It may be more applicable to measure the impact of not making improvements. When considering the cost of not making improvements, consider that poor data quality:

- compromises decision-making with potentially serious impacts on clinical decision making
- increases operational costs due to time and resources required to detect and correct errors
- directly affects the patients’ perception of the organisation
- reduces employee morale.

Monitoring and evaluation

The Ministry of Health will implement and monitor the effectiveness of the Data Quality Improvement Programme. Monitoring will take place through the district annual plans and KPI measurements of data submitted to NZHIS.

The governance group will help monitor the implementation of the programme throughout the sector.

Peer review will provide feedback to organisations within the sector from a peer perspective, with a report to be supplied to the Ministry of Health on the outcomes of the review.
Organisational change requirements

The amount of organisational change required depends on current practices. Some organisations already have extensive data quality improvement programmes and staff members who are accountable for data quality. All that may be required is a shift in focus to align with the national Data Quality Improvement Programme that means the organisation meets KPIs and for an increased emphasis on training data collectors, including clinical staff. Incremental change is a pragmatic process that allows changes to be assessed on an individual organisations basis for effectiveness.

The development of a cross-organisational data quality group will increase awareness of responsibilities across the organisation without impacting significantly on resources. However, increased awareness of data quality issues and commitment to improvement will require human resources with appropriate skills and some training may be required. The initial impact will be an increase in workload that may require the redirection of staff currently working on ‘find and fix’ projects. It is expected that improvements in data management processes will subsequently reduce the ‘find and fix’ workload, therefore increased staff counts should not be required where a data quality team is already in place.
## Stakeholder analysis

### Ministry stakeholders

<table>
<thead>
<tr>
<th>Stakeholder</th>
<th>Information Interest/Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minister and Associate Ministers of Health</td>
<td>Accurate information and advice from Ministry officials</td>
</tr>
<tr>
<td>Ministry of Health Directorates Management Teams</td>
<td>Decision-making Policy development and analysis Assessment of DHB performance</td>
</tr>
<tr>
<td>Policy Analysts</td>
<td>Policy development and analysis</td>
</tr>
<tr>
<td>Policy Advisors</td>
<td>Advice to the Minister, Ministry and sector</td>
</tr>
<tr>
<td>NZHIS Clinical Analysis Team</td>
<td>Assessment of health outcomes from national collections Reporting requirements to WHO Yearly publications of health indicators, including mortality analysis Information to the Minister, researchers, DHBs, the Ministry directorates Data quality caveats to understand limitations of data Metadata of all national collections</td>
</tr>
<tr>
<td>Data Quality Team</td>
<td>Information from data suppliers on collection processes Ensure business processes support data suppliers to provide good data Metadata of all national collections Data standards and data dictionaries Data quality requirements of customers Data quality caveats to customers</td>
</tr>
<tr>
<td>Business Intelligence Unit</td>
<td>Information following data analysis on collections on request from customers Data quality caveats and contextual information to customers Administrative data to initiate improvement of business processes to ensure a high-quality service Management of the metadata repository</td>
</tr>
<tr>
<td>Clinical Coding Team</td>
<td>Codes for Mortality, Cancer Registry and Private Hospital data Metadata of all national collections International coding standards</td>
</tr>
<tr>
<td>HealthPAC Business Analysts</td>
<td>Persistent issues with the supply of claims data Provision of data to NZHIS NZHIS used standards, data dictionaries and metadata Fraudulent claims Accidental over/underpayment of claims</td>
</tr>
<tr>
<td>National Screening Unit Data Quality Manager</td>
<td>Information from data suppliers about collection processes Data standards Data quality requirements of customers Data quality caveats to customers Ensure business processes support data suppliers to provide good data</td>
</tr>
<tr>
<td>HIS</td>
<td>Current standards in use Priorities for standards endorsement to support data quality Dissemination and support of endorsed standards</td>
</tr>
</tbody>
</table>
## Sector stakeholders

<table>
<thead>
<tr>
<th>Stakeholder</th>
<th>Information Interest/Requirement</th>
</tr>
</thead>
</table>
| Clinicians                   | Access to clinical data to support decisions  
                                 | Data quality caveats to understand the limitations of the data  
                                 | Education as data collectors                                                                 |
| HBs                          | Strategic decision-making on funding and planning requirements                                         |
| CEOs                         | Management of data supplied from business units  
                                 | Management of information systems (new and existing) and clinical and administration data  
                                 | Support for business units that collect and store data  
                                 | Strategic analysis of future information requirements                                         |
| CIOs                         | Management of data supplied from business units  
                                 | Management of information systems (new and existing) and clinical and administration data  
                                 | Support for business units that collect and store data  
                                 | Strategic analysis of future information requirements                                         |
| Quality Managers             | Assessment of all aspects of quality across the organisation requiring accurate clinical and administrative data  
                                 | Provision of quality improvement initiatives  
                                 | Meeting hospital accreditation requirements                                                                 |
| Business Units               | Collection of data for internal and external reporting requirements                                  |
| Information Analysts         | Assessment of clinical and administrative data for health outcomes                                  |
| Clinical Coders              | Contextual information for effective and specific coding of primary and secondary diagnoses  
                                 | Current international data standards                                                                 |
| Shared Services Agencies     | Information Analysts  
                                 | Funding and performance measurement for regions  
                                 | Health outcomes analysis for regions  
                                 | Strategic directions for service requirements                                                                 |
| PHOs/IPAs                    | CEOs/Managers  
                                 | Funding and performance measurements for catchment area  
                                 | Health outcomes analysis for catchment area  
                                 | Reporting requirements  
                                 | Claims data requirements                                                                 |
| Practice Managers            | Administrative data  
                                 | Clinical data management – possibly used for decision support  
                                 | Claims data requirements  
                                 | Reporting data requirements                                                                 |
| Privacy Commissioner         | Appropriate management of the privacy of health data throughout the sector including access restrictions |
| Health and Disability        | Accurate and complete documentation of clinical practice  
                                 | Tracking of alterations to data  
                                 | Protection of the patients information for appropriate users only                                                                 |
| Commissioner                | ACC Information Analysts  
                                 | Ability to merge data to provide health outcomes data                                                                 |
| State Services Commission    | Alignment with e-government standards and guidelines                                                                 |
| Statistics New Zealand      | Alignment with standards used in demographic data to enable benchmarking                           |
| Researchers                  | Data quality caveats to understand the limitations of the data  
                                 | Contextual information on data collection Metadata                                                                 |
10 References


Kmietowicz, Z. (2004). Data collection is poor because staff don’t see the point. *British Medical Journal, 328*, 786.


Redman, T. C. (1996b).* Why Care About Data Quality*. In *Data Quality in the Information Age* (pp. 303): Artech House.


