Finance & Management Engineering Journal of Africa (https://damaacademia.com/fmeja/)
Volume 1, Issue 5, pp.22-46, May 2019
Published by: Dama Academic Scholarly & Scientific Research Society (www.damaacademia.com)

The Institutionalisation of Data Quality in the New Zealand Health Sector

Karolyn Kerr
The University of Auckland

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.

Keywords: *Institutionalisation, Data Quality, New Zealand Health Sector*

1.0 INTRODUCTION

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.

1.1 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

[1] http://mitiq.mit.edu

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.

1.2 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 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.

[2] www.iaidq.org
[3] http://mitiq.mit.edu/
[4] www.dmreview.com

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).

1.2.1 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).

1.2.2 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)*

1.2.3 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).

2.0 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).

2.1 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.

Governance comprises:	
stewardship, which is:	and custodianship, which is

• 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.	• 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.

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).

3.0 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, and 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.

Category	Dimension

Intrinsic	Accuracy Objectivity Believability Reputation
Accessibility	Accessibility Access security
Contextual	Relevancy Value-added Timeliness Completeness Amount of data
Representational	Interoperability Ease of understanding Concise representation Consistent representation

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.

Category	Dimension
Accuracy of Data	Believability
	Accuracy
	Objectivity
	Reputation
Relevancy of Data	Value Added
	Relevancy
	Timeliness
	Completeness
	Appropriate amount of data
Representation of Data	Interpretability
	Ease of Understanding
	Representational Consistency
	Concise Representation
Accessibility of Data	Accessibility
	Access Security
Eliminated Dimensions	Traceability
	Variety of Data Sources
	Ease of Operation
	Flexibility
	Cost-Effectiveness

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.

3.1 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:

Finance & Management Engineering Journal of Africa (https://damaacademia.com/fmeja/)
Volume 1, Issue 5, pp.22-46, May 2019
Published by: Dama Academic Scholarly & Scientific Research Society (www.damaacademia.com)

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).

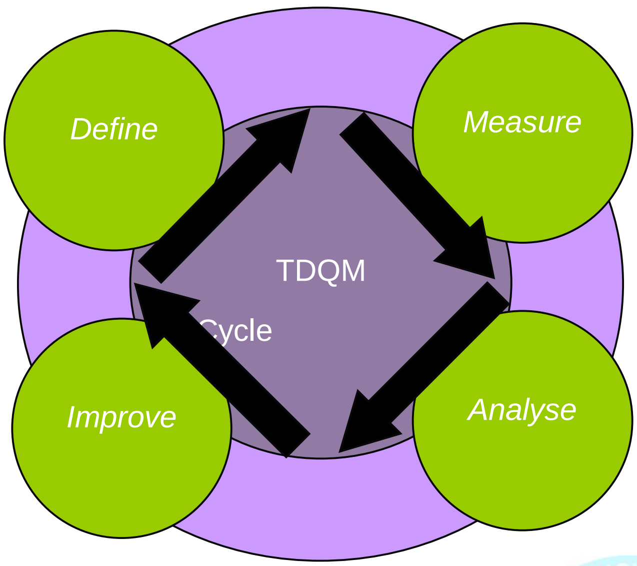

Figure 1: Components of the TDQM Cycle (Wang, 1998)

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).

3.2 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.

4.0 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.

4.1 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.

4.2 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

Table 4: The Assessment of the Wang & Strong (Wang & Strong, 1996) Data Quality Framework Using Criteria Defined by Eppler and Wittig (Eppler & Wittig, 2000).

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.

5.0 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).

5.1 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).

6.0 CONCLUSION

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).

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.

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.

References

Abate, M. L., Diegert, K. V., & Allen, H. W. (1998). *A hierarchical approach to improving data quality*. Retrieved 6.08.06, from http://www.dataquality.com/998abate.htm

Adelman, S. (2001). *Data Strategy Introduction*. Retrieved 18.8.06, from http://www.dmreview.com/article_sub.cfm?articleID=4291

Agmon, N., & Ahituv, N. (1987). Assessing data reliability in an information system. *Journal of Management Information Systems, 4*(2), 34-44.

Alshawi, S., Missi, F., & Eldabi, T. (2003). Healthcare information management: the integration of patient's data. *Logistics Information Management, 16*(3/4), 286-295.

Al-Shorbaji, N. (2001). *Health and medical informatics*. Cairo: World Health Organisation.

Argyris, C., & Schön, D. (1978). *Organisational Learning: A Theory of Action Perspective*. Reading: Addison-Wesley.

Arts, D. G. T., De Keizer, N. F., & Scheffer, G. J. (2002). Defining and improving data quality in medical registries: a literature review, case study, and generic framework. *Journal of the American Medical Informatics Association, 9*, 600-611.

Atkinson, C., Eldabi, T., Paul, R. J., & Pouloudi, A. (2001). *Investigating integrated socio-technical approaches to health informatics.* Paper presented at the 34th Hawaii International Conference on System Sciences.

Australian Department of Health and Aging. (2003). *Health Connect*. Retrieved 18.10.05, from http://www.healthconnect.gov.au/index.htm

Avison, D., & Myers, M. (1995). Information systems and anthropology: an anthropological perspective on IT and organisational culture. *Information Technology & People, 8*(3), 43-56.

Ball, M., & Douglas, J. (2002). IT, patient safety and quality care. *Journal of healthcare Information Management, 16*(1), 28-33.

Ballou, D. P., & Pazer, H. (1995). Designing information systems to optimise the accuracy-timeliness tradeoff. *Information Systems Research, 6*(1), 51-72.

Ballou, D. P., & Pazer, H. (2003). Modelling completeness versus consistency tradeoffs in information decision contexts. *Transactions on Knowledge and Data Engineering, 15*(1), 240-243.

Ballou, D. P., & Tayi, G. K. (1989). Methodology for allocating resources for data quality enhancement. *Communications of the ACM, 32*(3), 320-329.

Ballou, D. P., & Tayi, G. K. (1999). Enhancing data quality in data warehouse environments. *Communications of the ACM, 42*(1), 73-78.

Ballou, D. P., Wang, R., Pazer, H., & Tayi, G. K. (1998). Modelling information manufacturing systems to determine information product quality. *Management Science, 44*(4), 462-484.

Baskerville, R. (1999). Tutorial: Investigating information systems with action research. *Communications of the Association for Information Systems, 2*(19).

Baskerville, R., & Myers, M. (2004). Special issue on action research in information systems: making IS research relevant to practice - foreword. *MIS Quarterly, 28*(3), 329-335.

Baskerville, R., & Wood-Harper, A. T. (1996). A critical perspective on action research as a method for Information Systems research. *Journal of Information Technology, 11*(3), 235-246.

Baskerville, R., & Wood-Harper, A. T. (1998). Diversity in information systems action research methods. *European Journal of Information Systems, 7*, 90-107.

Bates, D. W. (2002). The quality case for information technology in healthcare. *BMC Medical Informatics and Decision Making, 2*(7).

Bauer, K. (2004). *The power of metrics: KPIs - the metrics that drive performance management*. Retrieved 18.08.06, from http://www.dmreview.com/article_sub.cfm?articleID=1009207

Berndt, D. J., Fisher, J. W., Hevner, A. R., & Studnicki, J. (2001). Health care data warehousing and quality assurance. *Computer, 34*(12), 56-65.

Black, N., Barker, M., & Payne, M. (2004). Cross sectional survey of multicentre clinical database in the United Kingdom. *British Medical Journal, 328*, 1478-.

Bowen, P., Fuhrer, D. A., & Guess, F. M. (1998). Continuously improving data quality in persistent databases. *Data Quality, 4*(1).

Bowers, B. (1988). Grounded Theory. In B. Sarter (Ed.), *Paths to Knowledge: Innovative Research Methods for Nursing* (pp. 33-59). New York: National League for Nursing.

Broder, C. (2004). *Case Study: Data repository gets to the heart of quality improvement*. Retrieved 18.08.06, from www.ihealthbeat.org/index.cfm?Action=dspItem&itemID=103350

Bryant, A. (2002). *Grounding systems research: re-establishing grounded theory.* Paper presented at the 35th Hawaii International Conference on System Sciences.

Canadian Institute for Health Information. (2003a). *Data Quality Framework.* Retrieved 18.8.06, from http://secure.cihi.ca/cihiweb/dispPage.jsp?cw_page=quality_e

Canadian Institute for Health Information. (2003b). *Earning Trust. Key Findings and Proposed Action Plan from the Data Quality Strategies Study*.

Carey, R. G., & Lloyd, R. C. (2001). *Measuring Quality Improvement in Healthcare. A Guide to Statistical Process Control Applications*. Milwaukee: Quality Press.

Carson, C. S. (2000). *What is data quality? A distillation of experience*. Retrieved 18.08.06, from http://www.thecre.com/pdf/imf.pdf

Carson, C. S. (2001). *Toward a framework for assessing data quality*. Retrieved 18.08.06, from http://www.imf.org/external/pubs/cat/longres.cfm?sk=3991.0

Champagne, F. (2002). *The ability to manage change in health care organisations*: Commission of the Future of Health Care in Canada.

Chengalur-Smith, I., Ballou, D. P., & Pazer, H. (1999). The impact of data quality information on decision making: an exploratory analysis. *IEE Transactions on Knowledge and Data Engineering, 11*(6), 853-864.

Chung, W. Y., Fisher, C., & Wang, R. (2002). *What skills matter in data quality?* Paper presented at the Seventh International Conference on Information Quality, Boston.

Codd, E. F. (1986). *An evaluation scheme for database management systems that are claimed to be relational.* Paper presented at the Second International Conference on Data Engineering, Los Angeles.

Committee on Quality of Health Care in America. (2000). *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington DC: National Academy Press.

Council of the European Union. (1995). *Article 6. European Union Data Protection Directive*. Retrieved 3.11.05, from http://www2.warwick.ac.uk/fac/soc/law/elj/jilt/1996_1/special/directive/#a2.1

Croft, G., & Williams, J. (2005). The RCP Information Laboratory (iLab): breaking the cycle of poor data quality. *Clinical Medicine, 15*(1), 47-49.

Crosby, P. (1980). *Quality is Free*. New York: Penguin Group.

Cunningham, N. (2001). A model for the role of information systems in strategic change within health care organisations. *Organisation Development Journal, 19*(1), 93-108.

Davenport, T. H. (1998). *Working Knowledge: How Organisations Manage What They Know*. Boston: Harvard University.

Davidson, A. J., Melinkovich, P., Beatty, B., Chandramouli, V., Hambridge, S. J., Phibbs, S. L., et al. (2003). Immunisation registry accuracy. Improvement with progressive clinical application. *American Journal of Preventive Medicine, 24*(3), 276-280.

Davidson, B. (2004). Developing data production maps: meeting patient discharge submission requirements. *International Journal of Healthcare Technology and Management, 6*(2), 223-240.

Davidson, C., & Voss, P. (2002). *Knowledge Management*. Auckland: Tandem.

Dravis, F. (2003). *Data Quality Strategy*. Retrieved 14.8.06, from http://www.firstlogic.com/_gics10643/DQ/article.asp?articleID=32

Dravis, F. (2004a). *Data quality strategy: A step by step approach.* Paper presented at the Ninth International Conference on Information Quality, Boston.

Dravis, F. (2004b). *IQ Maturity, it will be smaller than your shoe size*. Retrieved 18.08.06, from http://www.firstlogic.com/resourcelibrary/bestpractices/iqview/iqview0104.asp

De Lusignan, S., Stephens, P. N., Adal, N., & Majeed, A. (2002). Does feedback improve the quality of computerised medical records in primary care? *Journal of the American Medical Informatics Association, 9*, 395-401.

Deming, W. E. (1982). *Out of the Crisis*. Cambridge: Massachusetts Institute of Technology.

Detmer, D. E. (2003). Building the national health information infrastructure for personal health, health care services, public health, and research. *BMC Medical Informatics and Decision Making, 3*(1).

Dey, I. (1993). *Qualitative Data Analysis. A User Friendly Guide for Social Scientists*. London: Routledge.

Dubois, L. (2003). *What Works: Curing the blown budget in data consolidation*. Retrieved 11.07.05, from www.tdwi.org/research/displayaspx?ID=6877

Dugas, M., Hoffman, E., Janko, S., Hahnewald, S., Matis, T., & Uberla, K. (2001). *XML-based visual data mining in medicine.* Paper presented at the MEDINFO 2001, Amsterdam.

Eckerson, W. W. (2002). *Data Quality and the Bottom Line*. Retrieved 26.08.06, from http://www.adtmag.com/article.asp?id=6321

Effken, J. (2002). Different lenses, improved outcomes: a new approach to the analysis and design of health care information systems. *International Journal of Medical Informatics, 65*, 59-74.

Elrod, P. D. (2002). The death valley of change. *Journal of Organisational Change, 15*(3), 273-291.

English, L. (1998, May 1998). *Plain English on data quality: Data quality point 6: Institute training on data quality*. Retrieved 23.08.06, from http://www.dmreview.com/master.cfm?NavID=198&EdID=619

English, L. (1999a). *Improving Data Warehouse and Business Information Quality*. New York: John Wiley & Sons.

English, L. (1999b, October 1999). *Information quality in the knowledge age*. Retrieved 21.08.06, from http://www.dmreview.com/master.cfm?NavID=198&EdID=1452

English, L. (1999c, March 1999). *Plain English on data quality: Data quality point 13: Education and self-improvement*. Retrieved 21.08.06, from http://www.dmreview.com/master.cfm?NavID=198&EdID=117

English, L. (2003, June 9, 2003). *Plain English about information quality: How to save $576 925 000 through IQ management*. Retrieved 21.08.06, from http://www.dmreview.com/article_sub.cfm?articleId=6823

Eppler, M. J., & Wittig, D. (2000). *A Review of Information Quality Frameworks from the Last Ten Years*. Retrieved 14.08.06, from http://www.knowledgemedia.org/modules/pub/view.php?id=knowledgemedia-2&page=1

Fine, L. G., Keogh, B. E., Cretin, S., Orlando, M., & Gould, M. M. (2003). How to evaluate and improve the quality and credibility of an outcomes database: validation and feedback study on the UK cardiac surgery experience. *British Medical Journal, 326*, 25-28.

Fisher, C., Chengalur-Smith, I., & Ballou, D. P. (2003). The impact of experience and time on the use of data quality information in decision making. *Information Systems Research*, *14*(2), 170-188.

Fletcher, D. (2004). Achieving data quality: new data from a paediatric health information system earns trust of its users. *Journal of AHIMA, 75*(10), 22-26.

Fox, C. J., & Levitin, A. V. (1996). Data and Data Quality. *Encyclopaedia of Library and Information Science, 57*(Supplement 20), 100-122.

Fry, R. (2002). *Appreciating your past in order to transform the future*. Retrieved 26.9.05, from http://www.clevelandshrm.com/

Fuller, B., & Redman, T. C. (1994). *Data Quality lessons learned at Telecom Australia.* Paper presented at the QUALCON 94 Conference, Melbourne.

Gendron, M. S., & D'Onofrio, M. J. (2001). *Data Quality in the Healthcare Industry*. Retrieved 21.08.06, from http://www.dataquality.com/901GD.htm

Glaser, B. G., & Strauss, A. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine Publishing.

Glaser, J. P. (2002). *The Strategic Application of Information Technology in Health Care Organisations* (2nd ed.). San Francisco: Jossey-Bass.

Gribben, B., Coster, G., Pringle, M., & Simon, J. (2001). Non-invasive methods for measuring data quality in general practice. *New Zealand Medical Journal, 114*, 30-32.

Haan, C., Adams, M., & Cook, R. (2004). Improving the quality of data in your database: lessons from a Cardiovascular Centre. *Joint Commission Journal on Quality and Safety, 30*(12), 681-688.

Hagland, M. (2003). Reduced errors ahead. *Healthcare Informatics*(August), 31-40.

Hamel, G., & Prahalad, C. K. (1994). *Competing for the Future. Breakthrough Strategies for Seizing Control of Your Industry and Creating the Markets of Tomorrow*. Boston: Harvard Business School Press.

Hovenga, E. J. S., Kidd, M. R., & Cesnik, B. (1996). *Health Informatics. An Overview*. Melbourne: Churchill Livingstone.

Huang, K.-T., Lee, Y. W., & Wang, R. (1999). *Quality Information and Knowledge*. New Jersey: Prentice-Hall.

Institute of Medicine. (2000). *To Err is Human. Building a Safer Health System*. Washington: National Academy Press.

Juran, J. M., & Godfrey, A. B. (1999). *Juran's Quality Handbook* (5 ed.). New York: McGraw-Hill.

Kamira, R. (2003). *Kaitiakitanga: Introducing useful indigenous concepts of governance.* Paper presented at the Health Informatics New Zealand Conference, Auckland.

Karr, A. F., Sanil, A. P., & Sacks, J. (2000). *Workshop Report, Affiliates Workshop on Data Quality*. Morristown, New Jersey: National Institute of Statistical Sciences.

Kenney, N., & Macfarlane, A. (1999). Identifying problems with data collection at a local level: survey of NHS maternity units in England. *British Medical Journal, 319*, 619-622.

Kerr, K. (2004). *The Electronic Health Record in New Zealand*. Retrieved 12.8.06, from http://hcro.enigma.co.nz/website/index.cfm?fuseaction=articledisplay&FeatureID=040304

Kerr, K., & Norris, T. (2004). *The development of a healthcare data quality framework and strategy.* Paper presented at the Ninth International Conference on Information Quality, Boston.

Klein, B., & Rossin, D. F. (1999). Data errors in neural network and linear regression models: An experimental comparison. *Data Quality, 5*(1), 25.

Kmietowicz, Z. (2004). Data collection is poor because staff don't see the point. *British Medical Journal, 328*, 786.

Kovac, R., Lee, Y. W., & Pipino, L. L. (1997). *Total Data Quality Management: The case of IRI.* Paper presented at the 1997 Conference on Information Quality, Boston.

Langley, P. (1994). *Doing Social Research: a guide to coursework*. Lancashire: Causeway Press Ltd.

Lee, R. D., Conely, D. A., & Preikschat, A. (2000). eHealth 2000: health care and the internet in the new millennium. *WIT Capital Industry Report*.

Lee, Y. W. (2004). Crafting rules: context reflective data quality problem solving. *Journal of Management Information Systems, 20*(3), 92-119.

Lee, Y. W., Pipino, L. L., Strong, D. M., & Wang, R. Y. (2004). Process-embedded data integrity. *Journal of Database Management, 15*(1), 87-103.

Lee, Y. W., & Strong, D. M. (2003). Knowing why about data processes and data quality. *Journal of Management Information Systems, 20*(3), 13-39.

Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). AIMQ: A methodology for information quality assessment. *Information and Management, 40*(2), 133-146.

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

Levitin, A. V. (1998). Data as a resource: properties, implications, and prescriptions. *Sloan Management Review, 40*(1), 89-101.

Levitin, A. V., & Redman, T. C. (1993). A model of the data (life) cycles with application to quality. *Information and Software Technology, 35*(4), 217-223.

Lewin, K. (1951). *Field Theory in Social Science*. New York: Harper and Brothers.

Liener, F., Gaus, W., Haux, R., & Knaup-Gregori, P. (2003). *Medical Data Management*. New York: Springer-Verlag.

Long, J. A. (2004). *Where to start? A preliminary data quality checklist for emergency medical services data.* Paper presented at the 9th International Conference on Information Quality, Boston.

Long, J. A., & Seko, C. E. (2002). *A new method for database data quality evaluation at the Canadian Institute for Health Information.* Paper presented at the 7th International Conference on Information Quality, Boston.

Lorence, D. P., & Jameson, R. (2001). Adoption of information quality practices in US healthcare organisations. A national assessment. *International Journal of Quality and Reliability Management, 19*(6), 737-756.

Lorenzi, N., Riley, R., Blyth, A., Southon, G., & Dixon, B. (1997). Antecedents of the people and organisational aspects of medical informatics: review of the literature. *Journal of the American Medical Informatics Association, 4*(2), 79-93.

Lorenzoni, L., Da Cas, R., & Aparo, U. L. (1999). The quality of abstracting medical information from the medical record: the impact of training programmes. *International Journal of Quality in Health Care, 11*(3), 209-213.

Loshin, D. (2001). *Enterprise Knowledge Management. The Data Quality Approach*. California: Academic Press.

Loshin, D. (2003). *Knowledge Integrity: Assessment in the golden years*. Retrieved 23.9.05, from http://www.dmreview.com/article_sub.cfm?articleId=6803

Loshin, D. (2004). *Issues and opportunities in data quality management coordination*. Retrieved 11/04/05, from http://www.dmreview.com/article_sub.cfm?articleId=1000836

Lynes, D. A. (1998). *Analysis of power dialects occuring between health professionals and clients*. Lancaster: University of Lancaster.

Malcom, A. (1998). *Poor data quality costs 10% of revenues, survey reveals*. Retrieved 18.10.05, from http://computerworld.co.nz/news.nsf/UNID/CC256CED0016AD1ECC25684C000E0278?OpenDocument&Highlight=2,Poor,data,quality,costs

Marks, R. G., Conlon, M., & Ruberg, S. J. (2001). Paradigm shifts in clinical trials enabled by information technology. *Statistics in Medicine, 20*, 2683-2696.

McFadden, F. R., Hoffer, J. A., & Prescott, M. B. (1999). *Modern Database Management* (Fifth ed.). Reading: Addison-Wesley.

McGrath, G. M., & More, E. (2001). *Data integration along the health care supply chain: the Pharmaceutical Extranet Gateway Project.* Paper presented at the 34th Hawaii International Conference on Systems Sciences.

Menke, J., Broner, C., Campbell, D., McKissick, M., & Edwards, -. B., Joy. (2001). Computerised clinical documentation system in the pediatric intensive care unit. *DMC Medical Informatics and Decision Making, 1*(3).

Miles, M., & Huberman, M. (1994). *Qualitative Data Analysis. An Expanded Sourcebook* (2nd ed.). Thousand Oaks, California: Sage.

Miller, H. (1996). The multiple dimensions of information quality. *Information Systems Management, Spring*, 79-82.

Ministry of Health. (1996). *Health Information Strategy for the Year 2000*. Wellington.

Ministry of Health. (2000). *The New Zealand Health Strategy*. Wellington.

Ministry of Health. (2002). *He Korowai Oranga – Maori Health Strategy*. Wellington.

Ministry of Health. (2003a). *DHB CIO Health Sector Survey*. Wellington.

Ministry of Health. (2003b). *Health and Independence Report*. Wellington.

Ministry of Health. (2003c). *Ministry of Health ISSP Project*. Wellington.

Ministry of Health. (2004). *Draft National Collections Stewardship Framework*. Wellington.

Ministry of Health. (2005). *The Health Information Strategy for New Zealand*. Wellington.

Mintzberg, H. (1987). Crafting Strategy. *Harvard Business Review, 65*(4), 66-.

Mitchell, S., & Brown, A. (2002). *Performance Measurement Data: Is it fit for use? Data Quality Initiatives at CIHI*. Retrieved 23.9.05, from http://secure.cihi.ca/cihiweb/en/downloads/quality_perforamancepaper_e.pdf

Moody, D. L., & Shanks, G. G. (1998). What makes a good data model? A framework for evaluating and improving the quality of entity relationship models. *The Australian Computer Journal, 30*(3), 97-110.

Mumford, E. (2001). Action Research: Helping Organisations to Change. In E. M. Trauth (Ed.), *Qualitative Research in IS: Issues and Trends* (pp. 298). London: Idea Publishing Group.

Murray, C., Lopez, A., & Wibulpolprasert, S. (2004). Monitoring global health: time for new solutions. *British Medical Journal, 329*, 1096-1100.

New Zealand Health Information Service. (2003). *Current State Analysis of Data in the National Health Collections*. Wellington: Ministry of Health.

NHS Information Authority. (1998). *An Information Strategy for the Modern NHS. A national strategy for local implementation*. Retrieved 26.9.05, from http://www.nhsia.nhs.uk/def/pages/info4health/1.asp

NHS Information Authority. (2003). *Electronic record development and implementation programme. Lessons learned topic summary: Data quality*. Retrieved 3.11.05, from http://www.nhsia.nhs.uk/erdip/pages/publications/ERDIPDataQualitySummary_3.pdf

Nord, G. D., Nord, J. H., & Xu, H. (2005). An investigation of the impact of organisation size on data quality issues. *Journal of Database Management, 16*(3), 58-71.

Olson, J. E. (2003). *Data Quality. The Accuracy Dimension*. San Francisco: Morgan Kaufmann Publishers.

One Six Sigma Glossary. Retrieved 11.07.05, from http://www.onesixsigma.com/tools_resources/glossary_i.php

Orfanidis, L., Bamidis, P., & Eaglestone, B. (2004). Data quality issues in electronic health records: an adaption framework for the Greek health system. *Health Informatics Journal, 10*(1), 23-36.

Orr, K. (1998). Data quality and systems theory. *Communications of the ACM, 41*(2), 66-71.

Pandit, N. R. (1996). The creation of theory: A recent application of the grounded theory method. *The Qualitative Report, 2*(4), 1-14.

Patel, V., Kushniruk, A., Yang, S., & Yale, J.-F. (2000). Impact of a computer-based patient record system on data collection, knowledge, organisation and reasoning. *Journal of the American Medical Informatics Association, 7*(6), 569-585.

Pauleen, D. J., & Yoong, P. (2004). Studying human-centred IT innovation using a grounded action learning approach. *The Qualitative Report, 9*(1), 137-160.

Pautke, R. W., & Redman, T. C. (2001). The Organisation's Most Important Data Issues. In G. Mario, M. G. Piattini, C. Calero & M. F. Genero (Eds.), *Information and Database Quality*. Dordrecht: Kluwer.

Penchas, S. (2003). Complexity in health care systems. *Complexus, 1*, 149-156.

Pettigrew, A., Ferlie, E., & McKee, L. (1992). *Shaping Strategic Change. Making Change in Large Organisations. The Case of the National Health Service*. London: Sage Publications.

Pierce. (2004a, 25 July). Health care Rx: Information. *The Seattle Times*.

Pierce, E. (2003). *Pursuing a career in information quality: the job of the data quality analyst.* Paper presented at the 8th International Conference on Information Quality, Boston.

Pierce, E. (2004b). *Job wanted: the data quality analyst*. Retrieved 6.08.06, from http://www.tdan.com/i028hy03.htm

Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM, 45*(4), 211-218.

Plsek, P., & Greenhalgh, T. (2001). The challenge of complexity in health care. *British Medical Journal, 232*, 625-628.

Plsek, P., & Wilson, T. (2001). Complexity, leadership, and management in health care organisations. *British Medical Journal, 323*, 746-749.

Plummer, A. A. (2001). *Information systems methodology for building theory in health informatics: the argument for a structured approach to case study research.* Paper presented at the 34th Hawaii International Conference on System Sciences.

Porcheret, M., Hughes, R., Evans, D., Jordan, K., Whitehurst, T., Ogden, H., et al. (2004). Data quality of general practice electronic health records: the impacts of a program of assessments, feedback and training. *Journal of the American Medical Informatics Association, 11*, 78-86.

Porter, M. (1991a). Towards a dynamic theory of strategy. *Strategic Management Journal, 12*, 95-117.

Porter, M. (1998). What is strategy? In D. Ulrich (Ed.), *Delivering Results: A New Mandate for Human Resource Professionals*: Harvard Business School Press.

Porter, M. E., & Olmestead Teisberg, E. (2004). Redefining competition in health care. *Harvard Business Review, June*, 65-76.

Porter, S. (1991b). A participant observation study of power relationships between nurses and doctors in a general hospital. *Journal of Advanced Nursing, 16*, 728-735.

Porter, S. C., & Mandl, K. D. (1999). *Data quality and the electronic medical record: a role for direct parental data entry.* Paper presented at the American Medical Informatics Association Symposium, Washington DC.

Pretzlik, V. (1994). Observational methods and strategies. *Nursing Researcher, 2*(2), 13-21.

Price, R. J., & Shanks, G. (2005). *Empirical refinements of a semiotic information quality framework.* Paper presented at the 38th Hawaii International Conference on Systems Sciences.

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

Redman, T. C. (1994). Data Quality for telecommunications. *IEEE Journal on Selected Areas in Communications, 12*(2), 306-312.

Redman, T. C. (1996a). *Data Quality in the Information Age*: Artech House.

Redman, T. C. (1996b). Why Care About Data Quality. In *Data Quality in the Information Age* (pp. 303): Artech House.

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

Ribiere, V., LaSalle, A., Khorramshahgol, R., & Gousty, Y. (1999). *Hospital information systems quality: A customer satisfaction assessment tool.* Paper presented at the 32nd Hawaii International Conference of System Sciences.

Rigby, M. (1998). Integrated record keeping as an essential aspect of a primary care led health service. *British Medical Journal, 317*, 579-582.

Robson, W. (1997). *Strategic Management and Information Systems* (Second ed.). London: Pitman Publishing.

Rosenhead, J. (1998). *Complexity theory and management practice*. Retrieved 20.08.06, from http://human-nature.com/science-as-culture/rosenhead.html

Rothenburg, J. (1996). *Meta data to support data quality and longevity.* Paper presented at the First IEEE Metadata Conference, Silver Spring, MD.

Rusin, R. S., & Williams, A. M. (2001). Challenges in incorporating a quality strategy into information systems development efforts. *Topics in Health Information Management, 21*(4), 21-29.

Ryan, K. (2004). Universal electronic health records: a qualitative study of lay perspectives. *NZFP, 31*(3), 149-154.

Saba, V. K., & McCormick, K. A. (2001). *Essentials of Computers for Nurses. Informatics for the New Millennium* (Third Edition ed.). New York: McGraw-Hill.

Sanderson, H., Adams, T., Budden, M., & Hoare, C. (2004). Lessons from the central Hampshire electronic health record pilot project: evaluation of the electronic health record for supporting patient care and secondary analysis. *British Medical Journal, 328*, 875-878.

Schultze, U. (2001). Reflexive Ethnography of Information Systems Research. In E. M. Trauth (Ed.), *Qualitative Research in IS: Issues and Trends* (pp. 78-103). London: Idea Group Publishing.

Segev, A. (1999). Data quality challenges in enabling ebusiness transformation. Paper presented at the Seventh International Conference on Information Quality, Boston.

Seiner, B. (1997). *Improving data quality through information stewardship.* Paper presented at the Seventh International Conference on Information Quality, Boston.

Shankaranarayan, G., & Cai, Y. (2005). Supporting data quality management in decision-making. *Decision Support Systems*.

Shankaranarayan, G., Ziad, M., & Wang, R. Y. (2003). Managing data quality in dynamic decision environments: an information product approach. *Journal of Data Management, 14*(4), 14-32.

Shortliffe, E. H., & Barnett, G. O. (2000). Medical Data: Their Acquisition, Storage and Use. In E. H. Shortliffe, L. E. Perrecult, G. Wiederhold & L. M. Fagan (Eds.), *Computer Applications in Health Care and Biomedicine* (2nd ed., pp. 37-69): Springer.

Sidthorpe, B., Glasgow, N., & Longstaff, D. (2004). *Complex adaptive systems: A different way of thinking about health care systems*. Retrieved 25.8.05, from http://www.anu.edu.au/aphcri/Publications/Background_paper_stream1.pdf

Skinner, B. F. (1963). Behaviourism at fifty. *Science, 140*, 951-958.

Smith, M. K. (2001). *Peter Senge and the learning organisation*. Retrieved 30/8/05, from http://www.infed.org/thinkers/senge.htm

Spradley, J. P. (1980). *Participant Observation*. New York: Holt, Rinehart and Winston.

Stacey, R. D. (1993). *Strategic Management and Organisational Dynamics*. London: Pitman.

Stackpole, B. (2001). *Keeping info clean*. Retrieved 3.11.05, from http://www.eweek.com/article2/0,1759,1250280,00.asp

Stewart, D. W., & Shamdasani, P. N. (1990). *Focus Groups: Theory and Practice*. California: Sage.

Storey, V. C., & Wang, R. Y. (1994). *Modelling quality requirements in conceptual database design.* Paper presented at the International Conference on Information Quality, Massachusetts.

Strauss, A. (1987). *Qualitative Analysis for Social Scientists*. Cambridge: Cambridge University Press.

Strauss, A., & Corbin, J. (1998). *Basics of Qualitative Research*. Thousand Oaks: Sage Publications.

Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. *Communications of the ACM, 40*(5), 103-110.

Suh, B., & Han, I. (2003). The impact of customer trust and perception of security control on the acceptance of electronic commerce. *International Journal of Electronic Commerce, 7*(7), 135-162.

Susman, G., & Evered, R. (1978). An assessment of the scientific methods of action research. *Administrative Science Quarterly, 23*(4), 582-603.

Sutcliffe, K., & Weber, K. (2003). The high cost of accurate knowledge. *Harvard Business Review, May 2003*, 74-82.

Tayi, G. K., & Ballou, D. P. (1998). Examining data quality. *Communications of the ACM, 41*(2), 54-57.

TechRepublic. (2005). *Information Management Trends in 2005*. Retrieved 8.8.05, from www.techrepublic.com

Thiru, K., Hassey, A., & Sullivan, F. (2003). Systematic review of scope and quality of electronic patient record data in primary care. *British Medical Journal, 326*, 1070-1072.

UK Audit Commission. (2004). *Information and Data Quality in the NHS*. London: Audit Commission.

United Kingdom Department of Health. (2004). *A Strategy for NHS Information Quality Assurance*.

United States Department of Defence. (2003). *DOD Guidelines on Data Quality Management*: Defence Information Systems Agency.

United States Department of Health and Human Services. (2002). *HHS Guidelines for ensuring and maximising the quality, objectivity, utility, and integrity of information dissemination to the public*. Retrieved 6.11.05, from http://www.hhs.gov/infoquality/part1.html

Urquhart, C. (2001). An encounter with grounded theory: Tackling practical and philosophical issues. In E. M. Trauth (Ed.), *Qualitative Research in IS: Issues and Trends* (pp. 298). London: Idea Group.

van Bemmel, J. H., & Musen, M. A. (1997). *Handbook of Medical Informatics*. Houston: Springer-Verlag.

van Gool, K., Lancsar, E., Viney, R., Hall, J., & Haywood, P. (2002). Diagnosis and prognosis of Australia's health information for evidence-based policy. *Journal of Health Services Research Policy, 7*(Suppl 1), S1:40-S41:45.

van Manen, M. (1997). *Researching Lived Experience: Human Science for an Action Sensitive Pedagogy* (2nd ed.). London, Ontario: Althouse Press.

Wand, Y., & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM, 39*(11), 86-95.

Wang, R. Y. (1998). A product perspective on total data quality management. *Communications of the ACM, 41*(2), 58-63.

Wang, R. Y., Lee, Y. W., Pipino, L. L., & Strong, D. M. (1998). Manage your information as a product. *Sloan Management Review*(Summer), 95-105.

Wang, R. Y., Storey, V. C., & Firth, C. P. (1995). A framework for analysis of data quality research. *IEE Transactions on Knowledge and Data Engineering, 7*(4), 623-640.

Wang, R. Y.& Strong, D. M. (1996). Beyond accuracy: what data quality means to data consumers. *Journal of Management Information Systems, 12*(4), 5-33.

Wang, R. Y., Ziad, M., & Lee, Y. W. (2001). *Data Quality*. Massachusetts: Kluwer Academic Publishers.

Ward, J., & Peppard, J. (2002). *Strategic Planning for Information Systems* (Third ed.). Chichester: John Wiley & Sons.

WAVE Advisory Board. (2001). *From Strategy to Reality. The WAVE (Working to add Value through E-information) Project. Health Information Management and Technology Plan* (Project Report). Wellington: Ministry of Health.

Weed, L. (1989). *New premises and tools for medical care and medical education.* Paper presented at the International Symposium of Medical Informatics and Education, British Columbia, Canada.

Willshire, M. J., & Meyen, D. (1997). A process for improving data quality. *Data Quality*, *3*(1).

World Health Organisation. (1988). *Informatics and Telematics in Health. Present and potential uses*. Geneva.

Wu, J. (2004). *Empowering the information enterprise: Business processes are the foundation of an information enterprise*. Retrieved 3.11.05, from http://www.dmreview.com/article_sub.cfm?articleID=1003984

Zeffane, R., Cheek, B., & Meredith, P. (1998). Does user involvement during information systems development improve data quality? *Human Systems Management*, *17*, 115-121.

Zmud, R., Lind, M., & Young, F. (1990). An attribute space for organisational communication channels. *Information Systems Research*, *1*(4), 440-457.