The Development of a Data Quality Framework and Strategy for the New Zealand Ministry of Health

Karolyn Kerr

Department of Information Systems and Operations Management, University of Auckland,

Private Bag 92 019, Auckland, New Zealand

Karolyn_kerr@moh.govt.nz

Abstract

This research in progress aims to develop a data quality strategy for the Ministry of Health. Data quality requirements are increasing as a wider range of data becomes available and the technology to mine the data shows the value of data that is 'fit for use'. A data quality framework is the initial step towards an organisation wide data quality strategy that aligns with the sectors existing strategies and policies. The framework development builds on an existing framework developed by the Canadian Institute for Health Information, and takes into account current data quality literature and recognised Total Quality Management Principles. Further piloting of the framework through assessment of national health collections will provide the information on which to base an effective data quality strategy for the Ministry of Health.

1. Introduction

Organisations are becoming more and more dependant on data, virtually everything the modern organisation does both depends upon and creates enormous quantities of data. To meet the needs of the organisation, a comprehensive data management program is essential [1]. Further, Levitin and Redman (1998) [2] see the need for management science for data, as data are different from other resources and require different management techniques.

Organisational structures needed to address the above issues often do not exist. A contributing factor may be that there is no simple means to place economic value on data resources. A comprehensive, organisation wide data quality strategy, can address many of these issues.

As with most organisations, no formal data quality strategy is yet documented for the New Zealand Ministry of Health. The proposed Data Quality Framework will inform the overall development of a data quality strategy for the Ministry of Health. All outcomes of the development of the strategy will be documented for each collection in a 'Data Quality Documentation Folder' to enable internal data users access to all information about a collection in one location. Chapter headings in this folder include:

- Outline of collection, including uses of data and original purpose for the collection
- Roles and responsibilities
- Results of assessment of collection using data quality framework
- Action plan for quality improvement
- Current and previous data quality initiatives
- Meta data around collection
- Correspondence relating to collection

The documentation folder will be both paper and web based. This will build on existing work on data quality at the Ministry, pulling together documentation and current practice and making this information immediately available in one location. To determine an appropriate framework we need to consider the meaning of data quality and previous work in this field. This is outlined below.

2. Data Quality

Klein and Rossin (1999) [3] note there is no single definition of data quality accepted by researchers working in the discipline. Data quality is a multidimensional concept [3] as data itself is multidimensional [4]; [5]. Modern definitions of data quality have a wider frame of reference and many more attributes than the obvious characteristics of accuracy. Strong et al (1997) [6] take a consumer (people or groups who have experience in using organisational data to make business decisions) focused view that quality data is 'data that is fit for use', and this view is widely adopted

by the literature {[7];[8, 9];[6]. Redman (1999) comes to the following definition based on Joseph Juran [5];

> *"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* [10]".

Tayi and Ballou (1998) [11] confer, noting that data considered appropriate for one use may not possess sufficient quality for another use, citing the trend towards multiple uses of data through data warehouses.

Wang et al (1997) [7] see the consumer as integral to the meaning of data quality and consider that consumers have a much broader quality conceptualisation than information systems professionals realise. They went on to analyse what data quality means to consumers through a survey of data quality attributes and their dimensions as defined by consumers. This resulted in a comprehensive framework of data quality from data consumers' perspectives built into a hierarchical framework of data quality [7].

Therefore, a data quality strategy in an organisation must consider the end user and allow that user to define the level of quality required to make the data useful. Requirements may be different for corporate data than for local data. First steps in any improvement process must be to identify the uses made of the data and by whom. A data quality strategy also needs to look forward to the future potential uses of the data.

Larry English notes the emerging discipline of 'Enterprise Data Quality Management' (EDQM) whereby the organisation develops and adopts a set of consistent technology processes, which institutionalise data quality as a strategic asset, and business processes to make it a consistent competitive advantage [12]. This is similar to Total Quality Management (TQM) principles [12]. Data (or information) quality is recognised as one aspect of the TQM movement.

3. Data Quality Frameworks

Seminal works in the area of quality have defined various extensive frameworks to review systems within organisations. The frameworks all seek to assess areas where poor quality workmanship or inefficiencies that 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 [7]. The framework can go beyond the individual elements of data quality assessment, becoming integrated within the processes of the organisation. Willshire and Meyen (1997) [13] 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 (2000) [14] add that a framework should not only evaluate, but also provide a scheme to analyse and solve data quality problems with their proactive management.

In developing a data quality framework for the International Monetary Fund, Carson (2001) [16] notes that an assessment tool for data quality needs to have the following characteristics:

- comprehensive in coverage of the dimensions of quality and the characteristics that might represent quality
- balanced between rigour desired by an expert and the bird's eye view desired by a general data user
- structured but flexible enough to be applicable across a broad range of data collections
- lead to transparent results
- arrived at by drawing on best practice.

Both Willshire and Meyen (1997) [13] and Carson (2001) [16] found that a framework needs both objective and subjective attributes to be considered, using both objective and subjective measurement metrics, in order to reflect the contextual nature of data quality and the many potential users of that data.

Further steps to development include defining data quality attributes, the collection, measurement and analysis of the attributes. An evaluation should use at least two different methods to adequately determine the quality of data using both quantitative and qualitative measurements. The final step is to identify, evaluate and select effective remedies. Use of the framework is a step-by-step process. Steps include:

1. modelling existing data
2. defining data quality attributes
3. evaluating existing data quality levels
4. determining data quality priorities
5. identify remedies
6. re-measure and iterate

An extensive review of data quality frameworks from the last ten years [14] 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 study found that existing data quality frameworks are often domain specific and either strong on objective or subjective measurements, but thus far no framework has been developed that is strong on both measurements at the same time. Frameworks also often fail to analyse the interdependencies between the various criteria within the framework.

Therefore, Eppler and Wittig (2000) [14] suggest the following should be included in any new framework development:

- 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
- a framework that is at the same time theoretical and practical.

To assess the quality of organisational data, Wang and Strong (1996) [7], defined a data quality framework that contained 20 quality dimensions. These were later reduced to 15 and assembled into four categories, as shown in Table 1. The framework has been validated across a range of industrial and government locations. Wang and Strong (1996) [7] suggest several ways in which this hierarchical framework can be applied, including the use of a questionnaire to measure perceptions of data quality, development of quality improvement methods to improve data quality and as a checklist during data requirements analysis. The appropriateness of this framework was tested in a study of the US healthcare industry, in which respondents confirmed the importance of all 15 dimensions [17]. Included in their assessment of frameworks is the Wang and Strong (1996) [7] framework noted above.

This framework was assessed using the following criteria – clear definitions, contextual positioning, and mutually exclusive and collectively exhaustive criteria, concise, provide examples, and provide tools. Table 2 shows the conclusions made by Eppler and Wittig (2000) [14] when assessing the Wang and Strong (1996) [7] framework, finding that overall, the framework is generic and balanced. The Wang and Strong (1996) [7] framework was the only one out the seven assessed that included a means of measurement, tools to assist with using the framework and offers 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.

Taking into account the above literature, a data quality framework for the Ministry of Health can be defined as:

"A point in time assessment and measurement tool, integrated into organisational process, providing a benchmark for the effectiveness of any future data quality improvement initiatives and a standardised template for information on data quality both for internal and external users. The framework takes into account the Ministry's Data Governance Policy, Information Systems Strategic Plan, and WAVE".

4. Framework Development for the New Zealand Ministry of Health

New Zealand's public and private health care providers forward data on health care provision and claims based data to two of the business units (HealthPAC and NZHIS) within the Ministry of Health. Further data is collected through the cervical screening and breast screening registers. There are currently 30 national health information systems including data warehouses. The Public Health Intelligence Unit undertakes both regular and one-off surveys, related to public health topics.

Health Payments, Agreements and Compliance (HealthPAC) services include:

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

The New Zealand Health Information Service (NZHIS) is a group within the Ministry of Health responsible for the collection and dissemination of health-related data.

Inconsistencies often exist across an organisations collections and business units in relation to data quality improvement initiatives, business rules, coding standards and tables, and data definitions. Data moving from one collection to another in particular can provide considerable interoperability problems.

4.1 Uses of the Framework

The primary use of the framework is to assist in the assessment of cross-organisational data quality. Where consistencies in data quality problems are found, improvement measures that provide the best return on investment are more likely to occur where improvements impact across all or most of

the national collections. Data quality problems will be highlighted through the output report following assessment of a collection using the framework. Analysis of these reports from all the national collections will provide a prioritised list of 'to do' projects.

A consistent assessment of quality over time will allow for the analysis of the effectiveness of data quality interventions, with assessment undertaken pre and post the intervention.

The framework will also be used to inform new data collections on what processes need to be instilled in the development phase that will contribute to the collection of high quality data. Prevention is better than re-work.

5. Aim of the Research

The aim of the New Zealand Data Quality Evaluation Framework (DQF) project is to deliver:

> *"A Data Quality Framework that allows for the consistent and accurate assessment of data quality in all national health data collections held by the Ministry of Health, which will enable improved decision making and policy development in the health sector".*

The framework is being developed to provide a common, objective approach to assessing the data quality of all health information databases and registries. The framework enables the identification and measurement of major data quality issues, standardises information on data quality for users and helps to identify priorities, which in turn leads to continuous improvements.

6. Methodology

The Canadian Institute for Health Information (CIHI), similar in function to the NZHIS, have developed a framework based on Statistics Canada guidelines and methods, information quality literature and the principle of Continuous Quality Improvement. Eppler and Wittig's (2000) [14] research on the Wang and Strong's 1996 study [7] noted above, further informed the Canadian frameworks' development.

Following an assessment of the literature on data quality frameworks, and assessment against Eppler and Wittig's (2000) [14] evaluation criteria, the CIHI Framework was found to be robust. The researcher met with the developers of the CIHI Framework to discuss the feasibility of adjusting the CIHI Framework to a New Zealand health environment. The researcher attended the Massachusetts Institute of Technology Information Quality Conference 2003, in Boston, and the Massachusetts Institute of Technology Information Quality Course, (Information Quality Certification Programme, Course One) at the Sloan School of Management.

The CIHI Framework was assessed for completeness and relevance against current Ministry IT & IM strategy documents. These include the Information Systems Strategic Plan (ISSP) and the WAVE Report (Working to Add Value through E-information). Compliance with New Zealand legislation was also considered.

Proposed additions and changes to the CIHI Framework that take into account the above policies, along with the assessment of the existing framework, were included for discussion at two focus groups with internal Ministry staff. Focus groups were used in an effort to bring together business units who appeared to have similar issues with data quality, but no formal infrastructure was in place to co ordinate quality initiatives. The 'Ministry Data Quality Team' was formed to specifically look at ways of improving quality in a consistent way across the organisation. The terms of reference for the group state the objectives as being to:

- Educate and create awareness of the advantages of using quality data in decision-making.
- Coordinate data quality improvement initiatives across the Ministry.
- Assist in the development of a data quality framework for the Ministry of Health.
- Assist in the development of an organisation wide data quality strategy

Membership of the group was selected for representation from across the Ministry and its separate business units. The proposed framework was sent to all participants of the group. A presentation to

the group was made prior to the focus groups to ensure all participants had a common understanding of the purpose of the framework and the outcome goals of the focus groups. The group participated in two focus groups of two hours each. A member of the strategic Information and Technology team (the researcher) led the focus groups and an administrator was present to make audio recordings and to later transcribe the recordings, noting also the interaction between group members on discussion points.

7. Internal and External Piloting

This research is still in progress with piloting of the framework and its user manual on all data collections held by the Ministry. One of the main considerations of the pilot study is the clarity and ambiguity of the language used in the Framework and its manual. It is important the framework is used consistently across collections and misinterpreted meanings will affect this.

The Draft Framework is currently being piloted on national health collections within NZHIS on the Mortality database. Some data for the Mortality collection is provided to NZHIS by the Department of Internal Affairs from Births, Deaths and Marriages, with additional data being supplied by Practitioners and Coroners and from existing Ministry of Health data collections. This means that much of the collection process is not under the control of NZHIS.

An assessment of the HealthPAC Capitation Based Funding, (CBF) system is also taking place. The Capitation Based Funding System's primary function is to allocate funding on a population basis according to funding formulae.

External assessment, on a health related data collection managed outside the Ministry of Health, is being undertaken on a collection held by the A+ Network Centre for Best Patient Outcomes. The aims of the Centre are to assist clinicians to improve patient outcomes through the development of a generic tool to help manage care delivery.

A template has been developed to assist data managers to assess the effectiveness of the framework, its user manual and the proposed 'Data Quality Documentation Folder' for each collection and to document their findings. Areas formally assessed through semi structured interviews and a formal questionnaire include:

- The language used in the framework
- The language and examples provided in the user manual
- The length of time required to complete the assessment using the framework
- The value of the information provided from using the framework, as found by various users of the data
- The applicability of the dimensions, characteristics and criteria for the collection being assessed
- The contents of the data quality folder.

An assessment of the framework on a yet to be implemented collection, the Mental Health Workforce System, will be undertaken. The framework will be used as check list for ensuring data quality in instilled in the collection processes prior to implementation and an assessment made of frameworks applicability to this type of use.

8. Future Work

The development of the framework is an iterative approach. The pilot study will provide valuable practical information, as noted by the Canadians following the implementation of their first framework. Change management is required to ensure those working on data quality accept the ethos that prevention is better than rework and the improvement of data quality is everyone's job across the organisation.

Using the Ministry Data Quality Team to assess the usefulness of the framework has provided an internal user perspective on the aspects of data quality that are important to all types of users.

Further assessment of the framework by external data users, such as researchers at the Centre for Best Patient Outcomes, will be helpful in improving the assessment of the framework further.

A programme of work will be undertaken to improve the objective metrics used within the framework. Currently many of the metrics associated with each criterion are subjective assessments, made by those who manage the collections. While this is a valid form of measurement, the robustness of the framework will be improved through the addition of relevant objective metrics. The metrics will be based on current literature on data quality metrics, trend analysis of historical data, current key performance indicators for data suppliers as outlined in their contracts with the Ministry of Health, and on legislative requirements.

The summary information gained from assessments of all collections will be collated to form a prioritised list of data quality improvement initiatives across the Ministry. Ongoing assessment using the framework will provide information on the success of initiatives.

The framework is a tool that will be used as part of a data quality strategy that covers the entire organisation. The data quality strategy development will be informed, in part, by the outcomes of the framework assessment of collections, highlighting areas of need. The strategy will follow the guidelines found in the Ministry of Health Information Systems Strategic Plan, which include tasking data quality to those at a strategic level in the organisation.

While the Ministry value data there is further potential for the use of this data. Building ‘trust’ in the data throughout the health sector will ensure the data is used to its highest possible benefit. Data that is highly utilised for a variety of reasons improves its quality. Through extensive data mining, combining currently disparate collections will provide far more granular information, knowledge and wisdom on the state of our nations health.

Table 1: Data quality categories and dimensions (Strong et al., 1997, p.104)

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

9. Bibliography

1. Pautke, R.W. and T.C. Redman, *The Organisation's Most Important Data Issues*. 2001.
2. Levitin, A.V., *Data as a resource: properties, implications, and prescriptions*. 1998, Sloan Management Review.
3. Klein, B. and D.F. Rossin, *Data errors in neural network and linear regression models: An experimental comparison*. Data Quality, 1999. **5**(1): p. 25.

4. Fox, C.J. and A.V. Levitin, *Data and Data Quality.* Encyclopedia of Library and Information Science, 1996. **57**(Suppliment 20): p. 100-122.

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

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

7. Wang, R.Y., D.M. Strong, and L.M. Guarascio, *Beyond Accuracy: What data quality means to data consumers*. 1996, Total Data Quality Management Programme.

8. Department of Defence, U.S, *DOD Guidelines on Data Quality Management*. 2003, Defence Information Systems Agency. p. 28.

9. Brown, S., *Data quality: Relatively critical and critically relative.* DM Review, 2002.

10. Redman, T.C., *Data Quality. The Field Guide.* 2001, Boston: Digital Press. 241.

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

12. English, L., *Achieving enterprise wide data quality*. 2000, Trillium Software System. p. 18.

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

14. Eppler, M.J. and D. Wittig, *A Review of Information Quality Frameworks from the Last Ten Years*. 2000, Net Academy.

15. Porter, M.E., *Towards a dynamic theory of strategy.* Strategic Management Journal, 1991. **12**: p. 954.

16. Carson, C.S., *Toward a framework for assessing data quality*. 2001, International Monetary Fund.

17. Gendron, M.S. and M.J. D'Onofrio, *Data Quality in the healthcare industry*. 2001, Data Quality Journal.