Improving Data Quality in Health Care

Karolyn Kerr
Simpl, New Zealand

Tony Norris
Massey University, New Zealand

INTRODUCTION

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. The approach must also encompass the rights of the patient to have their health data protected and used in an ethical way. This article describes the principles to establish good practice and overcome practical barriers that define and control data quality in health data collections and the mechanisms and frameworks that can be developed to achieve and sustain quality. The experience of a national health data quality project in New Zealand is used to illustrate the issues.

BACKGROUND

Tayi and Ballou (1998) define data as "the raw material for the information age." English (1999) builds on the idea of information as 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.

Klein and 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" (Loshin, 2001; Redman, 2001; Wang, Strong, & Guarascio, 1996). Data quality is 'contextual'; the user defines what is good data quality for each proposed use of the data, within its context of use (Pringle, Wilson, & Grol, 2002; Strong, Lee, & Wang, 1997). 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).

Data quality is now emerging as a discipline, with specific research programmes underway within universities, the most significant being that of the Sloan School of Management Information Quality Programme at the Massachusetts Institute of Technology (MIT)[1]. The field is based upon the well-established Quality Discipline, drawing on the work of Deming (1982) and the adaptation of the "*plan-do-check-ac*" cycle (the Deming Cycle). It also draws upon the "*quality is free*" concept of Crosby (1980) arising from the notion that doing things wrong is costly, and imports the ideas behind the Six Sigma approach and Total Quality Management (Juran & Godfrey, 1999) adapted to Total Data Quality Management (TDQM) and the management of information as a product (Wang, Lee, Pipino, & Strong, 1998).

The research programmes are developing ways to combine TDQM with the strategic direction of the organization, aligning the data quality requirements with overall goals. At present, there is little research published in this area, although some organizations do have data quality programmes with some strategic alignment to the business requirements.

Data quality is also becoming an increasingly important issue for health care providers, managers and government departments. The movement towards total quality management in health care to improve patient safety and health care efficiency is demanding high quality information. Further, evidenced based care requires the assimilation of large amounts of relevant and reliable research data available at the point of clinical decision making. Strategic prevention, national consistency of improvement practices, evolving data standards, and targeted improvements with increasing consumer involvement are all moving health care towards a TDQM model of data quality management.

Using the TDQM model, many of the methods developed in other industries can be useful for improving health care data. However, health data differ from data that arise in other fields in several ways. For example, the complex and multidisciplinary nature of medicine means that, unlike other sciences, health care data have no internationally standardized terminologies. This lack of standardization produces many homonyms, where the same term can mean different things depending on context, and synonyms, where there are several ways of expressing the same meaning. Added to these sources of ambiguity is the plethora of healthcare abbreviations which have no agreed format. The required longevity, privacy, and confidentiality concerns associated

with health data are also distinguishing characteristics. Combine these inherent features with the need to share and integrate the data across distributed and disparate entities such as the Ministry or Department of Health, regional health boards, hospitals, general practices, and individual specialists, and it is hardly surprising that the improvement of healthcare data quality is multifaceted and more involved than in many other domains.

THE DEVELOPMENT OF A DATA QUALITY EVALUATION FRAMEWORK

A common imperative in the health sector is the need to structure and improve the management of data quality within regional or national health data collections that aggregate patient health data and use the combined data for epidemiological and service planning purposes. Clinicians and managers recognize the need for assessment tools that indicate the level of data quality, identify where the problems are, and who should be accountable. This information can then be used to construct a strategy for quality improvement. This was recently the situation in New Zealand where previous work to develop the Ministry of Health information systems strategic plan, and a current state analysis of data quality, provided further support for the commencement of a programme of data quality improvement.

The starting point for any Data Quality Evaluation Framework (DQEF) in the health sector is the pioneering work of the Canadian Institute for Health Information (CIHI) (2003a, 2003b) The CIHI has developed a comprehensive data quality framework based on a hierarchy of quality attributes ranging from criteria at the base level that are grouped into characteristics, and then further reduced into a small number of high-level dimensions[2]. The CIHI framework is of course developed for Canadian conditions and must be adapted for local circumstances to ensure applicability. This can be done by engaging data custodians and stakeholders in an iterative sequence of focus group, interviews, and discussions that modify the data attributes at each level to reflect existing organizational structures and practice. The iterative process is most readily facilitated by action research using grounded theory (Strauss & Corbin, 1998) to structure the analysis of qualitative data through inductive coding and comparison.

In the New Zealand project, for example, analysis confirmed the utility of the Canadian quality dimensions of accuracy, relevancy, timeliness, usability, and comparability, but found it necessary to add a dimension of privacy and security to ensure that this aspect is explicitly managed to meet New Zealander's expectations for national data collections. The project also revealed that New Zealand's unique patient identifier, the National Health Index (NHI), is a powerful tool for the location of poor quality (e.g., redundant or incorrect) database entries.

The New Zealand project also revealed the advantages of the action research approach. Discussion and feedback, coupled with rigorous analysis, raises awareness of the importance of data quality, an issue which is frequently underrated at all levels by data users and managers. The approach similarly creates a common understanding of terminology and recognition of the importance that quality improvement is an ongoing venture needing continued support and clear lines of accountability. Thus, in developing and applying a data quality evaluation framework, it is important to:

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

THE DEVELOPMENT OF A DATA QUALITY IMPROVEMENT STRATEGY

Porter (1991) states that the function of a strategy is to integrate the activities of diverse functional departments to ensure consistency within an organization with explicit, reinforcing goals and policies when senior management cannot participate or monitor all decisions directly. It will be apparent from the above discussion that a Data Quality Evaluation Framework will not, by itself, automatically lead to required improvements and that a strategy, as defined by Porter, is critical to its successful implementation. Robson (1997) also notes that a successful strategy "exploits opportunities and fits the circumstances at the time" with a requirement, therefore, to undertake a systematic, skillful, accurate and realistic assessment of the opportunities and to re-evaluate them. The design and execution of a Data Quality Improvement Strategy (DQIS) in health care therefore begins with a current state analysis followed by an assess-

Related Content

Coverage Criteria for State-Based Testing: A Systematic Review
Sonali Pradhan, Mitrabinda Rayand Srikanta Patnaik (2019). *International Journal of Information Technology Project Management (pp. 1-20).*
www.irma-international.org/article/coverage-criteria-for-state-based-testing/215011

Building an Internet-Based Workflow System: The Case of Lawrence Livermore National Laboratories' Zephyr Project
Judith Gebauerand Heike Schad (1999). *Success and Pitfalls of Information Technology Management (pp. 108-119).*
www.irma-international.org/article/building-internet-based-workflow-system/33484

Challenges in Data Mining on Medical Databases
Fatemeh Hosseinkhah, Hassan Ashktorab, Ranjit Veenand M. Mehdi Owrang O. (2009). *Encyclopedia of Information Science and Technology, Second Edition (pp. 502-511).*
www.irma-international.org/chapter/challenges-data-mining-medical-databases/13621

Principles to Guide the Integration and Implementation of Educational Technology
Sara Dexter (2005). *Encyclopedia of Information Science and Technology, First Edition (pp. 2303-2307).*
www.irma-international.org/chapter/principles-guide-integration-implementation-educational/14603

IT Offshoring: Trust Views from Client and Vendor Perspectives
Hajer Kefi, Alya Mlaikiand Richard L. Peterson (2013). *Perspectives and Techniques for Improving Information Technology Project Management (pp. 113-130).*
www.irma-international.org/chapter/offshoring-trust-views-client-vendor/73231