MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-928

Taming the Data Quality Dragon – A Theory and Method for Data Quality by Design

Jens H. Weber[a,b], Morgan Price[a,b], Iryna Davies[b]

[a] *Department of Computer Science, University of Victoria, Victoria, BC, Canada*
[b] *Department of Family Practice, University of British Columbia, Vancouver, BC, Canada*

Abstract

A lack of data quality (DQ) is often a significant inhibitor impeding the realization of cost and quality benefits expected from Clinical Information Systems (CIS). Attaining and sustaining DQ in CIS has been a multi-faceted and elusive goal. The current literature on DQ in health informatics mainly consists of empirical studies and practitioners' reports, but often lack a holistic approach to addressing DQ 'by design'. This paper seeks to present a general framework for clinical DQ, which blends foundational engineering theories with concepts and methods from health informatics. We define an architectural viewpoint for designing and reasoning about DQ. We introduce the notion of DQ Probes for monitoring and assuring DQ during system operation. The concepts presented have been validated in a real-world case study.

Keywords:
Information Management [L01.399]; Engineering [J01.293].

Introduction

The issue of clinical DQ has many facets. There is not yet a single commonly accepted standard definition of this concept; there however has been growing consensus on a number of aspects that have to be taken into account when discussing DQ in a health informatics context. Weiskopf and Weng have mapped DQ issues reported in the CIS literature to five quality attributes (*completeness*, *correctness*, *currency*, *plausibility* and *concordance*) [1]. The published literature on these DQ attributes is primarily of empirical nature or focuses on a particular practical aspect of defining, measuring or improving DQ for a particular purpose [2]. The lack of a fundamental theory and method for engineering CIS for DQ presents an impediment to the design and evolution of better health systems. The purpose of this paper is to define such a theory and method, referred to as *DQ by Design* (DQbD).

Methods

The DQbD method is based on the engineering paradigm of *functional decomposition*, in which complex system behaviour is broken down into more elementary discrete functions. We incorporated the idea of *Design by Contract (DbC)* from component-based software engineering and extended it to enable assertions over DQ concerns. We introduced and defined the concept of *DQ Probes* as a way to make DQ related assertions in pre- and post-conditions associated with CIS data processing functions. Our theory and method have been validated in several CIS data projects, including our project on building a third generation primary care CIS research network [3].

Results

We have formalized a taxonomy of five types of DQ probes:

Table 1 – Taxonomy of CIS Data Quality Concerns

	DQ Type		Clinical Example
0	Intrinsic	Conformance	Is diabetes on the problem list coded or uncoded?
1	Intrinsic-meta	Provenance	Recency of last A1c test in a diabetic
2	Extrinsic-internal	Internal concordance	Do patients on insulin have Diabetes on their problem list?
3	Extrinsic-external	External concordance, currency	Are the same allergies present in the hospital and primary care information systems? Which were updated most recently?
4	Statistical	Plausibility	Is the expected prevalence of diabetes in a practice (as determined in their EMR) for males 55-70 similar to published regional, jurisdictional, or national averages?

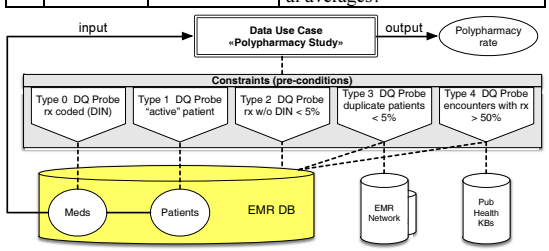

Figure 1 – DQ Probes in pre-conditions of Data Use Cases

We developed an architectural view point for designing CIS system use cases with explicit deployment of DQ probes. Our theory has been validated in a real-world project [3].

References

[1] Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J. Am. Med. Inform. Assoc., 2013; 20(1): 144-151.
[2] Mettler T, Rohner P, Baacke L. Improving data quality of health information systems: a holistic design-oriented approach. Proceedings of the 16th European Conference on Information Systems; Galway, Ireland. 2008: pp. 1883-1893.
[3] Price M, Weber JH, McCallum G. SCOOP–The social collaboratory for outcome oriented primary care. *IEEE ICHI* 2014: pp. 210–215.