Chapter 23
Development of a Methodological Approach for Data Quality Ontology in Diabetes Management

Alireza Rahimi
*University of New South Wales, Australia &
Isfahan University of Medical Sciences, Iran &
SWSLHD General Practice Unit, Australia*

Nandan Parameswaran
University of New South Wales, Australia

Pradeep Kumar Ray
University of New South Wales, Australia

Jane Taggart
*University of New South Wales Australia, &
SWSLHD General Practice Unit, Australia*

Hairong Yu
University of New South Wales, Australia

Siaw-Teng Liaw
*University of New South Wales, Australia &
SWSLHD General Practice Unit, Australia*

ABSTRACT

The role of ontologies in chronic disease management and associated challenges such as defining data quality (DQ) and its specification is a current topic of interest. In domains such as Diabetes Management, a robust Data Quality Ontology (DQO) is required to support the automation of data extraction semantically from Electronic Health Record (EHR) and access and manage DQ, so that the data set is fit for purpose. A five steps strategy is proposed in this paper to create the DQO which captures the semantics of clinical data. It consists of: (1) Knowledge acquisition; (2) Conceptualization; (3) Semantic modeling; (4) Knowledge representation; and (5) Validation. The DQO was applied to the identification of patients with Type 2 Diabetes Mellitus (T2DM) in EHRs, which included an assessment of the DQ of the EHR. The five steps methodology is generalizable and reusable in other domains.

1. INTRODUCTION

Improving data quality (DQ) in health organizations can improve quality of decisions and support better policy, strategies, and evidence-based patient care. DQ can be defined in terms of its *fitness for purpose* (Wang, 1998). The most frequently used DQ dimensions are *accuracy, completeness, consistency, correctness* and *timeliness* (S. T. Liaw et al., 2013). Research in DQ has tended to focus on the identification of generic quality characteristics that are applicable in a wide range of domains (Wand & Wang, 1996).

In the field of healthcare, data is collected routinely and may be used for research. It is becoming apparent that the quality of routinely collected data is not as good as it should be for many research applications. It is still not clear how DQ can be expressed in the context of fitness for purpose. Reference terminologies and ontologies have been used to specify DQ thus influencing data collection and analysis (Brown, Warmington, Laurence, & Prevost, 2003). They also act as benchmarks for assessing DQ (S. Liaw, Taggart, Dennis, & Yeo, 2011). An ontological approach can play a major role in the assessment of DQ and specification of fitness for purpose of a dataset (S. T. Liaw, et al., 2013; Rahimi, Liaw, Ray, Taggart, & Yu, 2014).

Building robust ontologies for DQ in healthcare helps automation of data extraction from the Electronic Health Records (EHRs) into clinical data warehouses; assessment and management of the quality of big data so that they are fit for purposes such as research, quality improvement, health information exchange and sharing; management of controlled vocabularies and optimizing semantic interoperability; curation of data for use by human users and applications such as electronic decision support systems; mining of data to discover relationships between the concepts; discovery of new knowledge; and finally reuse of knowledge in the management of chronic diseases (Wand & Wang, 1996).

In the biomedical informatics literature, ontologies have been described as collections of formal, machine process-able and human interpretable representation of the entities, and the relations among those entities, within a definition of the application domain (Rubin et al., 2006). Pipino (2002) proposed the most widely accepted definition, where he considers ontologies as an explicit specification of a conceptualization (Pipino, Lee, & Wang, 2002). Ontology provides a vocabulary of terms, their meanings and relationships to be used in various application contexts (Borst, 1997). This allows intelligent software agents to act more meaningfully in spite of differences in concepts and terminology.

We have previously described and discussed an ontology based approach (S. T. Liaw, et al., 2013; Rahimi, et al., 2014) to assessing the completeness, correctness and consistency (the 3Cs of DQ) of data and datasets. This approach is helpful in modeling the domain and representation of data and metadata requirements to identify diabetes on the data set from the University of NSW electronic Practice Based Research Network (ePBRN). This study used the dataset of 927 active patients from a general practice participating in the ePBRN, hereafter referred to as the General Practice Unit (GPU) dataset.

The ePBRN DQ research and development has focused on the 3Cs of DQ for ongoing ontology-based work to better define and address DQ, examine the issues and challenges for the network of data extraction and linkage, and semantic interoperability of large data sets (S. Liaw, et al., 2011). The ontology based approach can assist the terminology management and decision support to identify and classify different types of diabetes (S. Liaw, et al., 2011). This approach is also helpful in developing automated techniques and tools to extract and semantically link data elements (and concepts) in large data sets derived from multiple EHRs.

Related Content

Design Analytics of Complex Communication Systems Involving Two Different Sensory Disabilities
Gahangir Hossain (2017). *International Journal of Healthcare Information Systems and Informatics (pp. 65-80).*
www.irma-international.org/article/design-analytics-of-complex-communication-systems-involving-two-different-sensory-disabilities/178628

Evaluation of a Tool to Enhance Searching for Useful Medical Information on the Internet
David Parry (2008). *Healthcare Information Systems and Informatics: Research and Practices (pp. 330-343).*
www.irma-international.org/chapter/evaluation-tool-enhance-searching-useful/22131

A Study on Machine Learning and Supervised and Deep Learning Algorithms to Predict the Risk of Patients: Ten Year Coronary Heart Disease
Md Imtiaz Ahmedand Fatima Shefaq (2022). *International Journal of Practical Healthcare Innovation and Management Techniques (pp. 1-12).*
www.irma-international.org/article/a-study-on-machine-learning-and-supervised-and-deep-learning-algorithms-to-predict-the-risk-of-patients/305127

RMAC: Customised MAC Protocol for Roundabout Management Using VANET for Coperative Driving
Hitender Vatsand Ranjeet Singh Tomar (2021). *International Journal of E-Health and Medical Communications (pp. 77-92).*
www.irma-international.org/article/rmac/267956

Intermediation Structures in Electronic Healthcare Portals
Jonathan Warehamand Richard Klein (2002). *Effective Healthcare Information Systems (pp. 237-248).*
www.irma-international.org/chapter/intermediation-structures-electronic-healthcare-portals/9234