TOWARD A FRAMEWORK FOR DATA QUALITY IN ELECTRONIC HEALTH RECORD

Omar Almutiry, Gary Wills and Richard Crowder
School of Electronics and Computer Science, University of Southampton, Southampton, UK.
{osa1a11,gbw,rmc}@ecs.soton.ac.uk

ABSTRACT

Electronic Health Record (EHR) refers to the digital form of a patient's medical record. It is defined as a repository of patient data in digital form. This record is stored and exchanged securely and accessible by different levels of authorized users. Its key purpose is to support the continuity of care, and allow the exchange and integration of medical information for a patient. However, this would not be achieved without ensuring the quality of data populated in the EHR as the data quality can have a great impact on the overall effectiveness of EHR. The assurance of the quality of data used in healthcare systems is a pressing need to help the continuity and quality of care. Identification of data quality dimensions is a challenging issue as EHR data quality often focus only on data validation and verification, and overlook, for example, the appropriateness of use. Some research proposed frameworks of the data quality dimensions without taking into consideration the nature of e-healthcare systems. In this paper, we proposed an initial framework that fits the data quality attributes. This framework reflects the main elements of the healthcare systems and the functionality of EHR.

KEYWORDS

Electronic Health Record (EHR), Data Quality (DQ), DQ Dimensions, Information System.

1 INTRODUCTION

EHR refers to the digital form of a patient's medical record. It is defined as a repository of patient data in digital form. This record is stored and exchanged securely and accessible by different levels of authorized users (Häyrinen et al. 2008). One of the most noticeable advantages of adopting EHR systems in healthcare providers is the improved quality of the patient care as the key purpose of digital recording in healthcare services is to improve the quality of health services provided to a patient. Many studies (Thakkar & Davis 2006; Yoon-Flannery 2008) highlighted and emphasized how such systems could enhance the quality of care and support its continuity.

Data quality in information systems with its dimensions have been widely discussed by many researchers (D. P. Ballou & Pazer 1985; Tayi & D. Ballou 1998; Strong et al. 1997; R. Y. Wang et al. 1995; Fox et al. 1994; Anany Levitin & Thomas Redman 1995; Canadian Institute for Health Information 2009; Orfanidis et al. 2004). As a result, many frameworks of Data Quality (DQ) dimensions have been introduced and discussed in order to assure the quality of data populated in any information system. However, these frameworks have missed some important attributes that need to be involved to ensure, for example, the integrity and the origin of information. This is due to the fact that the frameworks are generic, and do not reflect the nature of the domain.

In the area of Health Information System, some issues and challenges have been arise that have an impact on the widespread adoption of EHRs. Data quality assurance is a challenging issue as the key barriers of optimally using data populated in EHRs is the increasing data quantity with poor quality. "Fitness for use" is one of the best definitions of the data quality. This definition takes us even further beyond the traditional concerns with accuracy of data, as it will end up with many other dimensions of data quality. So data quality is a concept with multi-dimensions. The existing frameworks of the dimensions were usually based on literature review, industrial experiences or intuitive understanding. Consequently, there is variety of DQ frameworks, and the definition of a dimension may vary from one framework to another.

We developed an initial framework that concerns DQ in the context of electronic health care systems. This framework is a result of filtering the existing data quality dimensions in many research, and checking their suitability to the nature of e-health systems.
This paper reviewed EHR systems and their functionalities, and data quality. After that, it discussed the proposed framework and its life development. The paper concludes with discussion and future work.

2 EHR AND ITS FUNCTIONALITIES

EHR refers to the digital form of a patient's medical record. It is defined as a repository of patient data in digital form. This record is stored and exchanged securely and accessible by different levels of authorized users (Häyrinen et al. 2008). Regarding its functionalities, the Institute of Medicine (IOM) Committee in the USA (Hoffman & Podgurski 2008) identified the key components of EHR systems and highlighted its functionalities. These core functionalities fall into eight categories, and are briefly discussed below.

- **Health Information and Data**: EHR systems should hold a defined data set that includes, for example, medical and nursing diagnoses, allergies, demographics and laboratory rest results to ensure improved access for some needed information to care stakeholders.
- **Results Management**: It is a feature that manages results of all types such as laboratory test results and radiology procedure results reports. This would prevent the redundant and additional testing, thus improving efficiency of treatment and decreasing cost.
- **Order Entry/Order Management**: Computerised provider order entry (CPOE) for areas like electronic prescribing can improve workflow processes, prevent the occurrence the lost orders and overcome the problem of ambiguities caused by illegible handwriting.
- **Decision Support**: Computerised decision support systems have demonstrated the ability to enhance clinical performance for many aspects of health care through, for instance, drug alert, rule-based alerts and reminders.
- **Electronic Communication and Connectivity**: Effective communication is crucial to the provision of the quality of health care. This can be between health care team members, patients and other partners such as pharmacy, laboratory and radiology. This communication and connectivity include the integrated medical record within the same facility, different facility within the same healthcare system, and among different systems (Thakkar & Davis 2006).
- **Patient Support**: Many forms of patient support have shown significant effectiveness in healthcare in general. These forms include patient and family education and home telemonitoring.
- **Administrative processes**: Electronic scheduling systems for hospital admission, inpatient and outpatient procedures, and visits play an important role not only to enhance the efficiency of healthcare units, but also provide better service to patients.
- **Reporting and Population Health Management**: This feature makes the process of reporting less labor-intensive and time-consuming. It helps report patient safety and quality data as well as public health data.

Reviewing the content and functionalities of EHR will help us to understand the nature of such context and, thus, the data quality requirements in order to enhance the quality of care.

3 DATA QUALITY

"Fitness for use" is one of the best definitions of the data quality. This definition takes us even further beyond the traditional concerns with accuracy of data, as it will end up with many dimensions of data quality. So data quality is a concept with multi-dimensions. Data quality includes not only data validation and verification, but also the appropriateness of use (Orfanidis et al. 2004). Despite the fact that there are many frequently used dimensions such as accuracy, consistency, completeness, and timeliness, there is no consensus on rigorously defined set of data quality dimensions (Strong et al. 1997; Tayi & D. Ballou 1998; Wand & R. Wang 1996).

3.1 Data quality dimensions

As the definition of data quality stated earlier, data has been best described with multiple dimensions. However, there is no clear consensus on the data quality dimensions. Moreover, this definition implies that many other attributes of data quality including usefulness and usability are very important aspects of quality. (Strong et al. 1997) classified these dimensions into four categories; intrinsic, accessibility, contextual and representational. Table 1 below summarised some frameworks for data quality dimensions and its originality.

Table 1. Data Quality (DQ) Dimensions

Research	Data Quality Dimensions
(D. P. Ballou & Pazer 1985)	Accuracy, completeness, consistency and timeliness.
(Strong et al. 1997)	Accuracy, objectivity, believability, reputation, accessibility, access security, relevancy, value-added, timeliness, completeness, amount of data, interpretability, ease of understanding, concise representation, consistent representation.
(R. Y. Wang et al. 1995)	Accessibility, interpretability, usefulness, believability.
(Fox et al. 1994)	Accuracy, currentness, completeness, and consistency.
(Anany Levitin & Thomas Redman 1995)	Contents (relevance, unambiguous definitions, obtainability of values), scope (comprehensiveness, essentialness), level of details (attribute granularity domain precision), composition (naturalness, occurrence identifiability, homogeneity), consistency (semantic consistency, structural consistency) and reaction to change (robustness, flexibility).

3.2 Health-related data quality dimensions

Many researches have defined some data quality dimensions in the context of health. Canadian Institute for Health Information (CIHI) defined five dimensions, namely accuracy, timeliness, comparability, usability and relevance. Each of these dimensions is divided into several related characteristics, and each characteristic is further made up to some criteria. Table 2 shows some frameworks of health-related data quality dimensions. (Liaw et al. 2012) stated in their survey the most frequent DQ dimensions; accuracy, completeness, consistency, correctness and timeliness.

Table 2. Health-related Data Quality (DQ) dimensions

Research	Data Quality Dimensions
(Canadian Institute for Health Information 2009)	Accuracy, timeliness, comparability, usability and relevance.
(Orfanidis et al. 2004)	Accessibility and availability, usability, security and confidentiality, provenance, data validation, integrity, accuracy and timeliness, completeness, and consistency.
(Liaw et al. 2012)	Accuracy, completeness, consistency, correctness and timeliness.

3.3 Impact of poor data quality

One of the key challenges of healthcare services is mitigating medication errors in the medication process. The motivation of developing a framework data quality assurance is to tackle the poor quality of data that badly affects the quality of care. These errors could lead to death as, for instance, there is an estimation of 98000 deaths each year in the U.S. costing as much as $29 billion (Hoffman & Podgurski 2008). Utilisation of EHR can result in improved patient safety by reducing the medical errors in hospitals (Bates 2000; DW et al. 1998). Even more, this type of environment could help care providers to identify and notify their patients about important changes in drug therapy (Jain et al. 2005; Parker et al. 2006).

3.4 Research Gap

The existing frameworks of the dimensions were usually based on literature review, industrial experiences or intuitive understanding. The definition of a dimension may vary from one framework to another. For instance, different definitions are given to Completeness in these researches (Wand & R. Wang 1996; Naumann et al. 1999), and accuracy are differently defined in different approaches, see the example given by (Wand & R. Wang 1996). The conception of data quality depends on the actual use of data. Thus, this depends on the application, and what may be considered as good quality of data in an application may not be sufficient in another application (Wand & R. Wang 1996). So Wand and Wang (1996) emphasized the importance of providing a design-oriented definition of data quality that will reflect the nature of information systems. The other issue of the existing approaches is the fact that they are too generic to adopt, as some attributes may be EHR-irrelative. The next section addresses these issues by proposing EHR-specific data quality framework.

4 THE PROPOSED FRAMEWORK

This framework should address the issues identified earlier. By reviewing the contents and functionalities of EHR systems, and discussing the DQ dimensions within the context of EHR system, this would lead to clear definitions of the attributes as well as EHR-relevant dimensions.

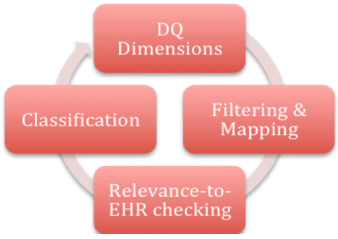

Figure 1. The process of the framework development

Figure 1 illustrates the process of developing the proposed framework. It started with gathering data quality dimensions in organizations and healthcare systems. These dimensions were filtered and mapped to each other, if needed, to eliminate the redundancies. In this step, literature review and dictionaries were used as a guideline to avoid having two attributes with the same implication. The next step is to check whether the dimension is relevant to the EHR functionalities and content. After that, the resulting dimensions were grouped and classified into three categories, which form the elements of our framework. These are information, communication and security, which are considered the main elements of e-healthcare systems (Shoniregun et al. 2010).

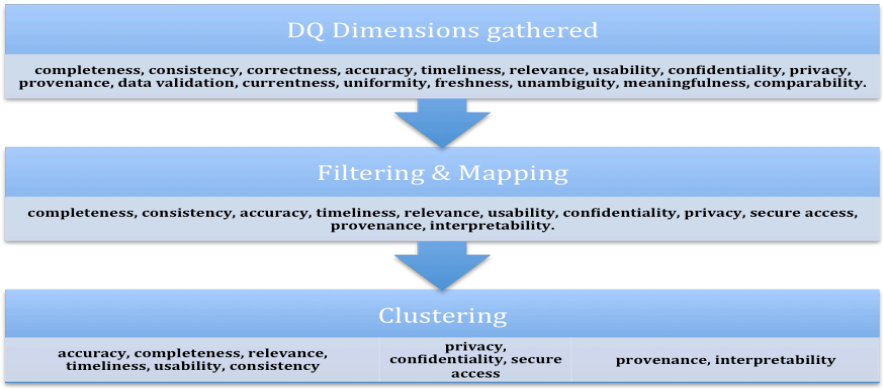

Figure 2. The flow output during the development

Figure 2 shows the flow of the dimensions output in each stage. In the last stage, the dimensions are classified into three categories. This classification fits into our framework, and this would cover all aspect of EHR systems as it balances between the comprehensiveness of DQ dimensions and the nature of EHR systems. Fitting the output into the proposed framework, we believe, would give a clear definition of each dimension, and help assure about what to measure and how.

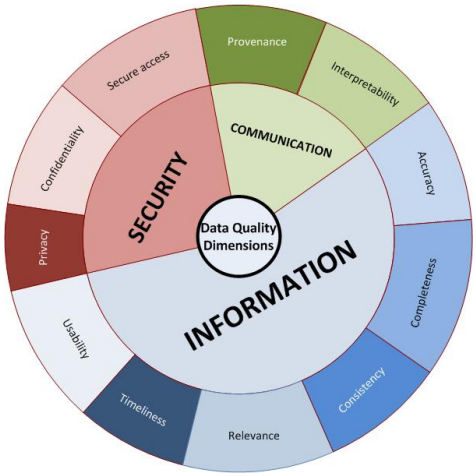

Figure 3. The proposed framework of the data quality in EHR

The characteristics of high-quality data consist of three categories: information, communication, and security aspects. As can be seen from figures 3, there are 11 data quality dimensions fitted in a framework of three categories.

4.1.1 Information

As mentioned earlier, information is one of the main elements that shape the e-healthcare systems. This category holds all dimensions that associate with data holdings. They are as follows:

- **Accuracy**: The Extent to which registered data is in conformity with the actual value.
- **Completeness**: The state in which information is not missing and sufficient for the task at hand. Linkage between data could promote the existence of further data.
- **Consistency**: The representation of data values remain the same and consistent between items of multiple data from multiple sources.
- **Relevance**: The extent to which information is appropriate and useful for the intended task.
- **Timeliness**: The state in which data is up-to-date, and its availability is on time.
- **Usability**: It reflects the ease with which data can be accessed, used, updated, understood, maintained and managed.

4.1.2 Communication

Communication is the second part of the initial framework. It concerns about the correspondences between different care units. As a result of this communication, EHR systems have items of multiple data from multiple sources.

- **Provenance**: The source of data should be shown and linked to metadata about provenance of data.
- **Interpretability**: The degree to which data can be understood.

4.1.3 Security

Security will prevent personal data from being corrupted, and control the access to ensure privacy and confidentiality.

To sum up with, the proposed framework has three categories in which the data quality dimensions in the context of healthcare fit into. These categories represent the main elements of e-health systems. The development of this framework went through many stages, and reflects the nature of EHR systems.

5 CONCLUSION AND DISCUSSION

As defined earlier, data quality is “fitness for use” and a concept with multi-dimensions. Yet there is no consensus on rigorously defined set of data quality dimensions.

EHR is seen as promising solution to problems in health information management despite the threats posed during data storage and transmission. However, one of the key barriers of optimally using the routinely collected data is the increasing data quantity with poor quality. This would raise the need of automating the mechanism of data quality measurement and semantic interoperability (Liaw et al. 2012).

Existing research focuses on DQ in generic information systems. DQ in these researches is addressed in many aspects align with data consumers. So that, we developed a framework that concerns DQ in the context of electronic health care systems. This framework is a result of filtering the existing data quality dimensions in many research, and checking their suitability to the nature of e-health systems.

The next step will be examining and evaluating the proposed framework by conducting semi-structured interviews with EHR stakeholders in order to confirm the structure of the framework.

REFERENCES

Ballou, D.P. & Pazer, H.L., 1985. Modeling data and process quality in multi-input, multi-output information systems. *Management science*, 31(2), pp.150–162.

Bates, D.W., 2000. Using information technology to reduce rates of medication errors in hospitals. *BMJ*, 320(7237), pp.788–791.

Canadian Institute for Health Information, 2009. *The CIHI Data Quality Framework*,

DW, B. et al., 1998. EFfect of computerized physician order entry and a team intervention on prevention of serious medication errors. *JAMA: The Journal of the American Medical Association*, 280(15), pp.1311–1316.

Fox, C., Levitin, A & Redman, T, 1994. THE NOTION OF DATA AND ITS quality dimensions. *Information Processing & Management*, 30(I), pp.9–19.

Hoffman, S. & Podgurski, A., 2008. Finding a Cure: The Case for Regulation and Oversight of Electronic Health Record Systems. *Harv. JL & Tech.*, 22, p.103.

Häyrinen, K., Saranto, K. & Nykänen, P., 2008. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *International Journal of Medical Informatics*, 77(5), pp.291–304.

Jain, A., Atreja, A. & Harris, C., 2005. Responding to the Rofecoxib Withdrawal Crisis : A New Model for Notifying Patients at Risk and Their Health Care Providers. *Annals of internal …*, pp.182–187.

Levitin, Anany & Redman, Thomas, 1995. Quality dimensions of a conceptual view. *Information Processing & Management*, 31(1), pp.81–88.

Liaw, S.T. et al., 2012. Towards an ontology for data quality in integrated chronic disease management: A realist review of the literature. *International journal of medical informatics*, pp.1–15.

Naumann, F. et al., 1999. Quality-driven integration of heterogeneous information systems. In *PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES*. pp. 447–458.

Orfanidis, L., Bamidis, P. & Eaglestone, B., 2004. Data Quality Issues in Electronic Health Records: An Adaptation Framework for the Greek Health System. *Health Informatics Journal*.

Parker, M. et al., 2006. Data quality: How the flow of data influences data quality in a small to medium medical practice.

Shoniregun, C.A., Dube, K. & Mtenzi, F., 2010. *Electronic healthcare information security*, Springer.

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

Tayi, G. & Ballou, D., 1998. Examining Data quality. *Communications of the ACM*.

Thakkar, M. & Davis, D.C., 2006. Risks, barriers, and benefits of EHR systems: a comparative study based on size of hospital. *Perspectives in Health Information Management/AHIMA, American Health Information Management Association*, 3.

Wand, Y. & Wang, R., 1996. Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*.

Wang, R.Y., Reddy, M.P. & Kon, H.B., 1995. Toward quality data: An attribute-based approach. *Decision Support Systems*, 13(3), pp.349–372.

Yoon-Flannery, K., 2008. A qualitative analysis of an electronic health record (EHR) implementation in an academic ambulatory setting. *... in primary care*, pp.277–284.