Guidelines for Health IT Addressing the Quality of Data in EHR Information Systems

Nabil Georges Badr [ID][a]
Higher Institute for Public Health, USJ, Lebanon

Keywords: Data Quality, Electronic Health Records, Electronic Health Records Information Systems, Medical Informatics.

Abstract: Quality of patient care is dependent on the quality of patient healthcare data. Electronic Health Record Information Systems (EHR-IS) capture patient health data for diagnosis, treatment, testing, medication and patient support. Issues in healthcare data quality comprise missing, incorrect, imprecise, and irrelevant data. Stakeholders of health data from practitioners, to patients, governments and lawmakers have long concerned themselves with these issues. Our paper looks at data quality in healthcare from the locus of ensuing risks, challenges and approaches in the literature. The paper proposes a reference for designing Electronic Health Record Information Systems and the evaluation of data quality in EHR-IS implementations.

1 INTRODUCTION

Electronic health record information systems are the components of IS that capture, manage and host the data in an electronic health record (EHR). Electronic health record information systems (EHR-IS) rely on datasets that include medical diagnoses, allergies, demographics and laboratory test results, computerized provider order entry for prescriptions, decision systems, rule based alerts and reminders, etc., and provide reporting and population health management through secondary use of data for research and disease control (Hoffman and Podgurski, 2008).

For four decades, quality of data in the medical record attracted medical researchers (Feinstein, 1970). The concept of a central database for health information, assigning trusted sources of data in a consolidated view of what was referred to as "*integrated clinical databases*" (Kahn, 1997).

Unchecked data sources present quality risk factors to patient care introduced by human system errors, implementation issues and lack of standards. These data sources must withstand quality maintenance approaches to maintain desired levels of data quality and standards to be able to remove the impediments and maintain a basis of quality in the data (Win et al., 2002). Adopters of EHR-IS have made it obvious that quality of patient care is dependent on the quality of healthcare data (Jones and Blavin, 2013). Jha et al 2008 explain that it is difficult for hospitals to obtain quality data in EHR-IS which are reliant on data prone to potential mistakes. Data quality concerns could be introduced by device borne issues of connectivity, synchronization, volume of data captured (Yao et al., 2011) and data formatting (Karkouch et al., 2016) in connected or wireless data capture devices (Zafar, 2017). Governments, policy makers and standards bodies have supported the development of architectures and guiding principles for semantically interoperable infrastructures, such as Health Level Seven (HL7) (Dolin and Alschuler, 2011) for the purpose of quality data exchange (Yun and Kim, 2007). Practitioners are seeking and adopting technologies to lessen the chances of errors, such as wireless handheld devices with for timely access to data entry and retrieval, calculation assistance for prescription dosage aimed at error reduction (Lu et al., 2005). Implementations of features of EHR systems such as closed loop medication administration have attracted significant attention, due to the serious looming risk of prescription errors (Singh et al., 2009). A lack of prudence of a physician might impact downstream healthcare quality; an error at this stage, may not

[a] https://orcid.org/0000−0001−7110−3718

169

show up until much later, and cause potentially a patient health risk. Patients are advised to be aware of this trend and take steps to ensure the accuracy of their medical records (Ash et al., 2004).

What is data quality in Healthcare IT? Can HealthCare IT address the quality of data in EHR information systems?

1.1 Approach

In order to help answer these questions, the paper reviews literature and practice publication in an exploratory style in an attempt to suggest potential IT architectural guiding principles for addressing issues of data quality in EHR-IS. First, we define dimensions for data quality relative to the context of Electronic Health Record systems. For that, we use the reference work on data quality by Wang and Strong (1996). These dimensions will help guide our exploration in the context of EHR-IS. Then, based on these dimensions, we conduct our literature review to explore data quality issues and ensuing risks in EHR-IS implementations. We perform our search in Google Scholar using a search filter of "data quality issues" AND "electronic health record systems" AND "risks", for publications dated since 1970. The query returned 517 results encompassing publications in the field of health informatics, namely the *Journal of the American Medical Informatics Association and the International Journal of Medical Informatics, Perspectives in Health Information Management, Health Informatics Journal, MIS quarterly, Health Information Science and Systems* and others. The search was limited to articles written in the English language. After screening, we identify the articles relevant to the study, selected based on their direct relevance to the subject of data quality in electronic health record systems, avoiding duplication in findings among publication and prioritizing literature review articles for their broad coverage of the subject. Articles are categorized under the four dimensions of Wang and Strong (1996) with a focus on data quality issues. We include literature on data quality assessment (Pipino et al., 2002; Weiskopf and Weng, 2013) and relevant practitioner journals such as Health Affairs, triangulated with federal agency bulletins concerned with the progress of innovations in health IT. For our final representation of a practical framework of guiding principles aimed at addressing issues of Data Quality in EHR-IS, we borrow from the four fundamental layers of data standards of content, structure, technology, and organization, a model introduced by Bott (2004). Our aim is not to present an exhaustive set of principles, but rather to

under-score essential higher order guiding principles. These Meta principles are proposed as unalienable fundamental guidelines for developers and implementers of EHR-IS, in instances and best practices that maintain higher levels of Data Quality.

2 BACKGROUND

Healthcare practitioners identify an essential need for "*quality*" in data collection systems to address challenges in improving quality in healthcare (Dixon-Woods et al., 2012). Data quality in EHR/EMR systems holds first place in interest, importance and relevance in electronic health care research (Coleman et al., 2015). Patient safety and quality of care are directly related to the quality of data in the healthcare ecosystem (Gallego et al., 2015). More specifically, data quality issues in healthcare were found to comprise missing, incorrect, imprecise, and irrelevant data (Mans et al., 2015). Other data quality concerns for standardized EHR-IS are validity, believability, accessibility, security, timeliness, completeness, interpretability, ease of understanding, and consistency (Orfanidis et al., 2004).

2.1 Dimensions of Data Quality

Data quality points to the fitness for the data to be used (Juran, 1988). Data quality has been given categories and dimensions (Wang and Strong 1996). At the foundation of data quality (DQ) concepts in the perspective of IS, Wang and Strong, 1996 have suggested that "*High quality data should be intrinsically good, contextually appropriate, clearly represented and accessible to data consumers*" (Wang and Strong 1996, p. 6).

For this paper, we use the four dimensions of Wang and Strong (Table 1) as a springboard to examine data quality issues in healthcare and provide a summary of remedies seen in the literature and in practice. Other researchers have built upon this framework owing to its high degree of inclusiveness of essential attributes of data quality that are important to data consumers in broad contexts (Pipino et al., 2002) and specifically in medical informatics (Weiskopf and Weng, 2013).

Wang and Strong characterise data as having quality in their own right referring to it as intrinsic data quality. Intrinsic data quality dimensions stress attributes of accuracy and objectivity (data is error free and represents no bias), and believability and reputation (relating to the source of data).

Wang and Strong also stipulate that data quality must be considered within the context of the task on hand. Contextual data quality attributes relate to completeness and timeliness (levels of relevancy, value-added, and amount of data). Two other dimensions (Representational and Accessibility DQ) relate to fundamental information system (IS) functions that manipulate data. These dimensions address attributes of quality in data related to a concise and consistent representation (data formatting), maintaining interpretability and understanding (meaningful data) and accessibility. IS professionals may argue that data access authorization is to be factored into data quality, however, it is increasingly evident that data access authorization (i.e. security) as a constituent of data privacy (Wickramage et al., 2017).

Table 1: Aspects of data quality that are important to data consumers (Wang and Strong 1996, p. 6).

Essential Attributes of the four Dimensions of DQ

Intrinsic DQ

(Data have quality in their own right)
- *Accuracy (also. correctness):* Data are error free
- *Objectivity:* Data represent no bias
- *Believability and reputation (also. credibility):* Data are from a trusted source

Contextual DQ

(Data quality is to be considered within the context of the task)
- *Relevancy:* Data are current and provide value
- *Completeness:* Data are in the right amount for the need
- *Timeliness:* Data are available at the right time

Representational DQ

(Emphasize the importance of the role of systems)
- *Data Concordance (Concise / consistent representation):* Data have no mismatch between sources or tables of data
- *Interpretability and understanding:* Data are meaningful, with no ambiguity

Accessibility DQ

(Emphasize the importance of the role of systems)
- *Accessibility:* Data are readily available for use by the consumer

For the context of this paper, we presume that accessibility implicitly refers to accessibility through authorised means and will focus mainly on the level of access to important data for the task.

3 FINDINGS AND DISCUSSION

3.1 Data Quality in Health Informatics

The literature on data quality in relation to electronic health record systems is not profuse. Studies represent healthcare data quality as a multidimensional construct, with the most used dimensions being completeness, accuracy, correctness, consistency and timeliness (Liaw et al., 2012). The literature review shows a great deal of variability and overlap in the terms used for quality attributes (Weiskopf and Weng, 2013). For instance, "*accuracy*" was found sometimes to be used as "*a synonym for correctness, but in other articles meant both correctness and completeness*" (Weiskopf and Weng, 2013, p.145). Nevertheless, it remains that the most frequently studied attributes of data quality in healthcare are of "*credibility* and "*accuracy*" (Leite et al., 2015). Yet, there seems to be a level disparity on what attribute is priority and on what the prevalent definitions of data quality attributes could be (AHIMA, 2013). Of these, incompleteness (missing information) and inconsistency (information mismatch between sources or tables of data) for example, which render the specific patient records unusable (Mikkelsen and Aasly, 2005), were sometimes reclassified under the attribute of "accuracy" (Gendron and D'Onofrio, 2001; Hristdis, 2009). Further, dimensions of data quality are interrelated (De Amicis et al., 2006). The analysis of interdependencies of dimensions of data quality has shown trade-offs among these dimensions. For instance, the improvement of timeliness could adversely affect the accuracy (Ballou and Pazer, 1995); various degrees of data completeness may affect consistency (Ballou and Pazer, 1995). Yet, the literature review has informed this study of interesting concepts.

The following sections provide a more succinct classification of principal data quality attributes in the context of healthcare informatics as found in the literature, organized according to the four dimensions of the framework of Wang and Strong. Suggested approaches to address data quality issues are also proposed. Section 3.6 introduces risk factors associated with Data Quality, and section 3.7 produces a framework of Guiding Principles for addressing issues of Data Quality in EHR-IS.

3.2 Intrinsic Data Quality

At a glance, in the first dimension, the *intrinsic data quality* dimension (Table 2), our review of the

literature has identified concepts of user entry of incorrect data (Wang et al., 2015; Mans et al., 2015) and errors in data transcription and data translation (Meystre et al., 2008) were identified to affect data accuracy and correctness in EHR-IS. These findings reinforce notions that data have quality in their own right and that quality could erode due to misuse and potentially improve through corrective action, often not so obvious. Objectivity, which can reflect user bias and assumptions in data entry could be corrected by comparing the data with other patients or historical values (Bayley et al., 2013).

Table 2: Challenges Affecting Intrinsic DQ.

Challenges Affecting…
Accuracy (also Correctness)
– User entry of incorrect data / Errors in data transcription and data
Objectivity
– Validity issues – Corrected by comparing the data with other patients or historical values
Believability and reputation (also Credibility)
– Undetected issues are repeated causing loss of credibility
– Believability issues (e.g. unrealistic blood pressure); often more difficult to distinguish

Believability issues in data quality sometimes show as easy to detect oddity that could be addressed (e.g. unrealistic blood pressure), others, are harder to detect and are still more difficult to correct. Data provenance information of data can improve believability (Gendron and D'Onofrio, 2001). Undetected issues can be repeated causing loss of credibility. Error reduction principles have focused on regulations concerning the effective use of technologies (Lu et al., 2005; Aimé et al., 2015). These regulations and associated best practices introduce concepts for error handling, normalization and terminology mapping included in system design (Dolin and Alschuler, 2011). Alerting functions for incorrect data entry are recommended (Moss and Berner, 2015; Qureshi et al., 2015). In order to reduce the chances of error due to language barriers (Blumenthal and Tavenner, 2010), the use of local terminologies is suggested with a lexicon built into the system to map terminologies used to standard dictionaries for interoperability (Aimé, et al., 2015). Practices involving data comparison with other patients or values the patient history, was found to improve the objectivity of the data collected (Bayley et al 2013). On the other hand, data credibility is sustained when the data collection associates data

provenance information (Gendron and D'Onofrio, 2001).

Hence, we summarize suggested approaches to address Intrinsic DQ issues, namely, as follows:

▪ Establish regulations concerning technologies for error reduction / normalization
▪ Ensure that error reduction and terminology mapping included in system design.
▪ Implement alerting functions as a warning for incorrect data entry
▪ Use of local terminologies in standard dictionaries for interoperability
▪ Encourage practices of data comparison with patients / values from patient history (objectivity)
▪ Stress the importance of data provenance information (traceability)

3.3 Contextual Data Quality

Our review has isolated data quality challenges pertaining to the second dimension that treats the *contextual data quality* dimension (Table 3), a function on how data elements are collected, treated and manipulated. Challenges to attributes of contextual data quality relate to maximizing the use of structured data for accurate interpretation.

Table 3: Challenges Affecting Contextual DQ.

Challenges Affecting…
Relevancy
– Insufficient information content of data
– Varying levels of IT literacy among care team
Completeness
– Missing / Omitted data (lack of time for data entry)
Timeliness
– Timely collection and available data
– Delay in data entry by practitioners, nurses and labs often due to workload
– System introduced delays / synchronization with separate systems / distributed databases.
– Operational issues hinder timely data entry

The literature review shows that data quality in EHR systems may withstand irrelevant (Mans et al., 2015), insufficient and incomplete data which in some instances could be due to lack of time for data entry (Bayley et al., 2013). The use of classifications and controlled vocabularies normalize the data collected (Hennessy et al., 2013) and a versioning capable repository keeps the proper data context (Dolin and Alschuler, 2011).

Timeliness of data was indicated as an important attribute for high quality EHR systems as it directly relates to quality of patient care (Liaw et al., 2012). Timely collection and accumulation of available data is a foundation data quality for quality of care (Hopf et al., 2014). Delay in data entry by practitioners, nurses and labs due to workload could negatively affect timeliness and eventually lead to data incompleteness if never performed, forgotten or fell victim to data literacy levels of the data entry collaborator (Lluch, 2011). Further, the literature emphasizes changes to processes and that reduce workload issues to avoid delay in data entry (Hopf, 2014). Hosting systems in multiple location, with different standards for database synchronization (Srour and Badr, 2017) could introduce a negative effect on data timeliness and data quality (Willis-Shattuck et al., 2008).

Therefore, we could summarize suggested approaches to address Contextual DQ issues, namely, as follows:

- Maximize the use of structured data for accurate interpretation.
- The use of classifications and controlled vocabularies to normalize the data collected
- The use of versioning capable repository in order to keep the proper data context
- Emphasis on changes to processes that reduce workload to avoid delay in data entry

3.4 Representational Data Quality

The literature reviewed included challenges affecting *Representational data quality attributes* (Table 4), depicted as the third dimension, to emphasize the importance of the role of systems functions which are largely discussed in the context of data *consistency,* and include prescriptive guidance on the need of data accreditation standards for the removal of inconsistency and duplication. Data could be collected from multiple sources in the ecosystem, even from remote mobile sensors in the use of telemedicine (Hennessy et al., 2013). This creates opportunities for inconsistencies in data. Multiple site implementations (Liaw et al., 2012) and data gathered from different sources that may use conflicting standards for data representation (Gendron and D'Onofrio, 2001), structure (Bott, 2004) and definition (Bayley et al 2013) contribute greatly to potential of inconsistencies.

System design flaws could introduce data corruption (Hoffman and Podgurski, 2008; Phillips and Fleming, 2009) and information mismatch between sources or tables of data. Attributes of

interpretability (meaningful data with no ambiguity) refer to the need to maintain user-friendliness and proper functionality of system interfaces (Jones et al., 2011; Phillips and Fleming, 2009) and reduce imprecise or ambiguous metadata (Mans et al., 2015). Ensuring interoperability through the definition of standardized terminologies is essential to remove ambiguity and maintain interpretability (Murff et al., 2011; Bayley et al 2013). Discrepancies between data fields must be virtually eliminated in order to reduce issues with interpretability (this is improved using constructed data sets for the user to choose from and avoid the use unstructured text). Fundamentals based on architectural models for semantic interoperability, initiatives and standards are imperative to counter data quality risk factors in human and systems implementation errors. Well-defined ontological foundations address semantic interoperability, clinical decision support and complexity of information systems models (Liaw et al., 2012).

Table 4: Challenges Affecting Representational DQ.

Challenges Affecting…
Data Concordance (Concise / consistent representation)
– Lack of standardized terminologies
– Data corruption due to system bugs
– Data mismatch from different sources / multisite implementations / different standards
– Data accreditation standards needed for the removal of inconsistency and duplication
– System design flaws introduce data corruption
– Varying standards of data structure (level of structured data implementation)
Interpretability /Understanding (meaningful, with no ambiguity)
– Discrepancies between data fields
– Complicated with use of unstructured text
– Interoperability - Standardized terminologies
– System interface problems
– Lack of user-friendly functionality
– Imprecise data or ambiguous metadata

Thus, approaches to address representational DQ pivot around:

- Implementing principles of standardization with reference to best practices that address systems and implementation issues
- Addressing semantic interoperability, clinical decision support and complexity of information systems models

3.5 Accessibility Data Quality

Lastly, grouping challenges that affect the availability of data into the fourth dimension, the essential attribute of *accessibility data quality* (Table 5). This dimension mandates a requirement of easy to use interfaces (Magrabi et al., 2015), availing the user to different data formats (Häyrinen et al., 2008), through any system and from any location. A vulnerability can be observed due connectivity and system availability issues (Gendron and D'Onofrio, 2001; Bayley et al., 2013). In an era of patient centered healthcare, patients have control over who has access to their data and in what circumstances (Cimino et al., 2002). The implementation of easy to use portals and interfaces (Gendron and D'Onofrio, 2001) should maintain accessibility in different data formats, through any system and from any location with secured portals for patients to control access to their records (Häyrinen et al., 2008). Accessibility constraints should limit different levels of users to access the data based on their credentials and must maintain compliance with HIPAA (Health Insurance Portability and Accountability Act of 1996) guidelines to ensure the security and privacy of data. Limited authorizations to data access can also contribute to data quality by limiting the chances of erroneous data entry.

Table 5: Challenges Affecting Accessibility DQ.

Challenges Affecting…
Accessibility (Data readily available for use by consumer)
– Vulnerable to system availability issues
– Ease of use issues hindering access to data
– Different levels of authorized users access
– Different data formats, systems and locations
– Lack of standardized terminologies
Recommended Approaches…
– Conducting thorough testing of EHR-IS applications for usability
– Implementing easy to use portals / interfaces maintain accessibility in different data formats,
– Providing levels of authorized users in compliance with HIPAA privacy guidelines
– Limiting authorizations to data access, limiting chances of erroneous data entry.

Therefore, in order to address accessibility DQ issues the literature indicates:

- Conducting thorough testing of EHR-IS applications for usability
- Implementing easy to use portals / interfaces maintain accessibility in different data formats,
- Providing levels of authorized users in compliance with HIPAA privacy guidelines

- Limiting authorizations to data access, limiting chances of erroneous data entry.

3.6 EHR-IS Data Quality Risk Factors

This section looks at the literature to identify salient risk factors associated with data quality variations introduced by human or system errors in EHR-IS.

3.6.1 Human and User Errors

EHR-IS are reliant on data prone to potential mistakes emanating from accidental errors in data entry (Wang et al., 2015), in medication dosage (Kaushal et al 2003), in data transcription or even in translation, such as in transcripts of voice recognition dictation system (Meystre et al., 2008). Data errors such as data that have been compromised, partially transferred between interconnected systems, wrongfully translated, entered in error or mixed up with someone else's, etc. present an issue with data quality and a risk on patient safety (Barkhuysen et al 2014). Easy to use interfaces with predefined archetypes could alleviate impending risks of data entry. Other risks could be caused by errors of data omission (Phillips et al., 2009). This risk could manifest in the form of patient safety and quality of care (Gallego et al., 2015), costly medical malpractice liability (Mangalmurti and Mello, 2010), and or health threatening prescription errors (Singh et al., 2009), especially if multiple repeated events are incurred before such issues are detected. Secured portals have enabled patients to control who can have access to their data (Cimino et al., 2002). Patient engagement practices bring forth potentials for enhancing the quality of care. Policies and procedures related to record management are required to sustain accuracy, integrity, and quality in patient records, especially in such situations where patient data entries are permitted and incorporated into the record (Bonomi et al., 2016).

3.6.2 System and Data Errors

Analysis and design of data quality issues are an integral part of the development of an EMR system (Orfanidis et al., 2004). Problems involving human factors were found four times as likely to result in patient harm as technical problems; Nevertheless, EMR system failures such as migration of records between systems, power failures, computer viruses and messaging failures, etc. were found to account for the majority of IT related EMR events (Magrabi et al., 2015). IS practitioners are urged to address safety concerns unique to EMR technology in the contexts of EHR-enabled health care (Rea, et al 2012). At the

time of their research, (Bates, et al., 2003) found, that nearly half of serious medication errors related to the fact that clinicians had insufficient information about the patient and the medication prescriptions (Fleming, et al 2011), possibly due to varying levels of IT literacy among care team (Lluch, 2011). Ease of use is fundamental for risk reduction in EMR-IS. "*Usability errors occur as a result of system complexity, lack of user-friendly functionality (e.g., confusing user interfaces) or workflow automation incompatibility* (Phillips and Fleming, 2009). Vendors of EMR systems often add functionalities to their interface design to assist with support and documentation (Weir et al., 2003), such as copy and paste, templates, use of standard phrases and paragraphs, and automatic object insertion (e.g., clinical values brought in from other parts of the electronic record). However, when used without proper education and controls, "*these features can lead to inaccurate documentation and potentially result in medical errors or allegations of fraud*" (AHIMA (2012). On the other hand, templates can guide documentation so that elements essential to demonstrating appropriate care are not ignored. Such features could improve the efficiency of data capture, timeliness and legibility, and consistency and completeness of documentation (Reed et al., 2012). In some unfortunate cases, templates automatically fill in data elements based on certain patient characteristics or other data entries, even though this default information is not an accurate representation of that particular patient encounter (Bowman, 2013).

Embedded clinical-decision support (CDS) systems are prone to human error and cognitive constraints (Sittig and Singh, 2012). Clinical decision support systems can still give wrong clinical advice even when designed and implemented according to high-quality standards, and is working as intended (Garg et al., 2005). Electronic records replacing paper-based records have introduced what is referred to as "adjacency error," in which a provider selects an item next to the intended one in a drop-down menu, such as the wrong patient or medication (Ash et al., 2004).

Programming error that incorrectly converts from one measurement system to another (e.g., pounds to kilograms or Celsius to Fahrenheit) could occasionally introduce undetectable errors (Phillips and Fleming, 2009). The implementation of alerting mechanisms for clinical decision support tools (Moss and Berner, 2015) are incorporated into electronic prescription system for instance (Qureshi et al., 2015). Disabling functions of these alerts based on the practitioner's perception that they are distracting or disruptive (alert fatigue) could result in a critical safety feature not being deployed when needed (Wheeler, 2015).

The complexity of real life situations can disrupt proper operation of the system and render CDS recommendations unusable, especially in case of frequent use of workarounds (Ash et al., 2004). Further, atypical circumstances, such as unusual combinations of conditions or local lack of resources, are not always taken into consideration. The number of decision tree options becomes too great and the system becomes impossible to maintain and use (Sittig and Singh, 2012). Ultimately, the trustworthiness and integrity of the health record are damaged.

3.6.3 Implementation Risks

Patient safety risks can certainly vary with the implementation stages of EMR in an organization (Lenert, 2002). Priorities for patient safety in the midst of an EMR rollout have been noted to differ from those of an organization that has used a fully integrated EMR system for 5 or more years (Dean et al., 2011). Issues with delays in data entry by practitioners, nurses and labs due to workload and operational issues hindering timely data entry. Insufficient training and preparation is liable to introduce data quality issues due to varying levels of IT literacy among care team (Lluch, 2011). EMR and related health information system designers and those responsible for integrated EMR implementation and management should be aware of the related types of errors and should take them into account as they build and implement such systems. These types are often due to (1) system faults, (2) metadata setup errors, (3) completeness of tests (how to backtrack data) and (4) system configuration errors (Ash et al., 2004). The risk of EHR downtime on clinical operations and patient safety increases with tightly coupled systems and widespread geographic areas (Sittig and Singh, 2012). Disparate systems and distributed data bases introduce delays in data synchronization (Srour and Badr, 2017). Patient safety could then be compromised as a result of miscommunication between the components of an EHR system causing a potentially unavoidable metadata mismatch.

3.7 Framework

In order to further the sense-making of our approach, we propose a higher order classification of the recommended approaches identified in the literature.

This classification is presented in a framework for addressing issues of Data Quality in EHR-IS (Table 6) using the suggested model of Bott, 2004 who have classified data standards under four fundamental layers: content, structure, technology, and organization (Bott, 2004). From our result in the previous sections, in the context of dimensions of data quality, potential measures to address the risks associated with implementation and use errors, it becomes clear that data quality standards for EMR-IS span all of these four layers.

Therefore, for our framework, we choose to triangulate our findings with this four-layer model in order to add rigor and significance to our framework. That said, for our final representation of the framework of guiding principles aimed at addressing issues of Data Quality in EHR-IS, we represent our findings under these four fundamental layers:

- *The Content Layer* deals with terminological issues such as classifications or controlled vocabularies.
- *The Structure Layer* defines dataset related practices improve data quality in EHR (Boyle and Cunningham, 2002) and are required for traceability (Yun and Kim, 2007).
- *The Technology Layer* contains regulations concerning technologies in healthcare aimed at error reduction (Lu et al., 2005).
- *The Organization Layer* relates to organizations to addressing organizational challenges associated with the introduction of EHR- IS into patient care practices. These challenges generally relate to structure, policies and processes.

3.7.1 Principles for Data Error Control

Tied to the **content layer**, we introduce the first guiding principle of "*data error control*" that supports the implementation of industry regulations concerning technologies aimed at error reduction through the application of concepts for error handling, normalization and terminology mapping included in solid system design. Thus, normalizing data collected from multiple sources (Hennessy et al., 2013). To that effect the Health Level Seven (HL7) standard specifies the structure and semantics of "*clinical documents architecture*" (Clinical Document Architecture standards - CDA) for the purpose of quality data exchange (Yun and Kim, 2007). User interfaces would implement alerting functions for incorrect data entry designed to reduce the potential of alert fatigue (too many system-generated alerts that tend to be ignored).

Systems must maintain data provenance information to reinforce data validity and credibility. Finally, preserving the local terminologies can lower the risk of error by keeping familiar references in the interfaces and mapping them in the backend to standards of semantic interoperability standards through the exploit of meaningful use APIs.

3.7.2 Context of Maintaining Quality Data - Structured Data Handling

Standards that define the **structure layer** concentrate on data repository structures reducing unstructured data elements such as notes and free form data that could be limited and constrained to a structure for dissemination and reporting. For this layer, a guiding principle points to the "*Context for maintaining quality data*". This principle stipulates that EHR-IS data management systems ought to be conceived in the mindful context for maintaining quality data namely in the preservation of data relevancy, completeness and timeliness. This relies on designs that maximize structured data use, classifications or controlled vocabularies for completeness check.

A structured data approach in necessary that incorporates formatting for patient data (personal record). Structured data could include vital signs, diagnosis, prescription drugs related data and event reporting (Declerck et al., 2015). Such system would implement versioning capable repositories for contextual validation and reference.

3.7.3 Principles of Systems Design

A third set of guiding principles tackles the **technology layer** and rests on fundamentals of "*system design*". One of this principle's primary edicts is to advocate lower system design complexity. On the data management backend, this principle supports quality attributes of interpretability and understanding (meaningfulness, with no ambiguity).

Thorough metadata entry validation routines must be considered to reduce metadata setup errors and use archetypes in data definition in order to reduce issues in terminologies. In order to sustain a concise and consistent representation of data, attention ought to be drawn to how to backtrack erroneous, corrupt or damaged data. Data accreditation standards could be applied for the removal of inconsistency and duplication in structured data.

From an infrastructure perspective, special architectural considerations would be associated with the reduction of risk of system failure with an emphasis on thorough testing during and after the implementation. Accessibility ought to be certified in

different data formats, through any system and from any location. Secured portals could be used for patients to control access and help routinely inspect and validate their health data for HIPAA compliance.

Establishing consistency among EHR systems, meaningful use APIs set standards for user interface applications (patient, clinician, payer, etc.), semantics and language translation, search and index functionality and how chart and record data are stored (Blumenthal and Tavenner, 2010).

3.7.4 Addressing the Organizational Context

The fourth and last set of principles brought forth in this framework, borrows its significance from organizational change management principles. Organizational changes must be built into the adoption process of an EHR-IS in order to ease adoption resistance, avoid the obstacles of IT literacy, reduce the workload of the staff and realign the processes for optimal workflow in data collection. The introduction of EHR- IS into patient care practices challenge organizations to develop changes in structure, policies, incentives and processes. The organizational layer, as introduced by the literature, treats those changes in processes, guidelines, roles and protocols required and caused by the usage of an EHR system in an organization (Poissant et al., 2005). Rearranged working relationships, schedules, authorities and prerogatives could damage the interaction among the healthcare team or improve it based on the team's readiness to face the shifting in roles and responsibilities (Willis-Shattuck et al., 2008). Process standardization contributes to data quality by readapting the data contributors to different EHR data requirements and new workflows (Hopf, 2014). Notwithstanding, the required changes in business processes, guidelines, roles and protocols are considered baseline in improving the performance of the healthcare team and the stakeholders of the EMR-IS data (Willis-Shattuck et al., 2008), and may apply as a foundation for any data quality conversation.

4 CONCLUSION

The paper builds upon existing academic and practitioner work to consolidate principal data quality attributes in the context of healthcare informatics. We perform an in depth exploratory literature review to develop a broad overview of electronic health records data quality risk factors, expose challenges in assessing data quality in electronic health record

information systems (Tables 2, 3, 4 and 5) and explore approaches in addressing issues of data quality in electronic health record information systems (Sections 3.1 to 3.5). Then, in order to extract relevant architectural guidance, we consolidate the information into guiding principles for addressing issues of DQ in EHR-IS using categorization four fundamental layers of data standards of content, structure, technology, and organization (Bott, 2004).

As a final step, based on these fundamental layers of data standards, we present the framework with 4 Meta Principles that categorize a set of Guiding Principles for addressing issues of Data Quality in EHR-IS (Table 6):

- *P1: Principles of error control and minimization;*
- *P2: Principles for maintaining data quality through rigorous data structure and versioning;*
- *P3: Key system design principles for data quality assurance; and*
- *P4: Providing the organizational context for "fit for use" data quality sustainability.*

4.1 Contribution and Further Research

Software developers in health information systems can exploit the guidelines in table 6 in order to improve the quality of data in their design and implementation of their product. Without necessarily introducing new concepts, the principal contribution of the paper aims at raising the awareness of developers and users of EHR-IS platforms and components regarding the importance, essential dimensions of data quality. As a focusing lens, the framework provides a reference for designing EHR-IS offering a guideline for implementing measures for data quality.

Approaches in addressing issues of data quality (DQ) in EHR-IS have limitations. Measuring data quality is a complex process requiring a systemic approach to data quality assessment. The level of use of structured data that may not be sufficient for the assessment of data quality in an EHR setting (Weiskopf and Weng, 2013), narrowing the scope to data verification and validation (Sachdeva and Bhalla, 2012). On the other hand, the use of gold standards for assessing data quality can be hindered by multisite systems and databases (Bae et al., 2015). Failure to extract data from all locations and to transform into a common format would result in incomplete data. Assessing completeness and concordance of the data set may prove more successful.

Hence, further research could be useful in evaluating approaches to measure data quality in Healthcare and relate it to out framework to close the feedback loop into the success of its implementation.

4.2 Limitations

This paper succeeded in connecting academic knowledge with practitioner and legislative approaches to achieve data quality. Nevertheless, in

order to manage the risks associated with the quality of data, assessment must be improved.

We recognize also that this paper has limitations in that it looked at only English publications in peer-reviewed journals and renowned practitioner publications.

Another limitation could also be in researcher bias in the classification of certain concepts. Though this was a successful approach to build a thorough research product, improvements are always possible.

Table 6: Proposed Framework for Approaches Addressing Issues of Data Quality in EHR-IS.

Meta Principles	Guiding Principles *(Related DQ Dimension)*
P1. Principles of Error Control and Minimization (Content Layer)	– Establish regulations concerning technologies for error reduction / normalization - *(Intrinsic DQ)* – Implement alerting functions for incorrect data entry - *(Intrinsic DQ)* – Use of local terminologies in standard dictionaries for interoperability - *(Intrinsic DQ)* – Stress the importance of data provenance information (traceability) - *(Intrinsic DQ)*
P2. Principles for Maintaining Data Quality (Structure Layer)	– Maximize the use of structured data for accurate interpretation - *(Contextual DQ)* – Use classifications and controlled vocabularies to normalize the data collected - *(Contextual DQ)* – Use of versioning capable repository in order to keep the proper data context - *(Contextual DQ)*
P3. Key System Design Principles for Data quality Assurance (Technology Layer)	– Ensure that error reduction and terminology mapping included in system design *(Intrinsic DQ)* – Focus of features to reduce Metadata Errors - *(Intrinsic DQ)* – Implement semantic interoperability, clinical decision support / Error reduction - *(Contextual DQ)* – Reduce complexity of information systems models - *(Representational DQ)* – Implement standards with reference to best practices to reduce systems and implementation issues and reduce the risk of system failure - *(Representational DQ)* – Provide levels of authorized users in compliance with HIPAA privacy guidelines (secured portals for patients to control access) - *(Accessibility DQ)* – Limit data access for reduced chances of erroneous data entry - *(Accessibility DQ)* – Lower complexity of system design - *(Accessibility DQ)* – Conduct thorough testing for usability- including how to backtrack data) - *(Accessibility DQ)* – Implementing easy to use portals / interfaces maintain accessibility in different data formats, through any system and from any location - *(Accessibility DQ)*
P4. Providing the Organizational Context for Data Quality Sustainability (Organization Layer)	– Identify and implement required process changes for data entry accuracy - *(Intrinsic DQ)* – Encourage practices of data comparison with patients / values from patient history - *(Intrinsic DQ)* – Emphasize process changes to that reduce workload / avoid delay in data entry - *(Contextual DQ)* – Address workload issues induced by the insertion of the EHR-IS, possibly through quick reference guides, online help and easy navigation - *(Accessibility DQ)* – Address IT literacy issues through training and features for adoption support - *(Accessibility DQ)*

4.3 Compliance with Ethical Standards

This research represents no conflict of interest and did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

AHIMA 2012. "Electronic Documentation Templates Support ICD-10-CM/PCS Implementation." *AHIMA Practice Brief, 2012*.

AHIMA 2013. "Integrity of the Healthcare Record: Best Practices for EHR Documentation." *Journal of AHIMA 84, no.8 (August 2013): 58-62*.

Aimé, X, et al 2015. Semantic interoperability platform for Healthcare Information Exchange. *IRBM, 36(2), 62-69*.

Ash, J. S., Berg, M., & Coiera, E. 2004. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. *JAMIA, 11(2), 104-112*.

Ash, J. S., Berg, M., & Coiera, E. 2004. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. *JAMIA 11(2), 104-112*.

Bae, M. D., Charles, J., Griffith PhD, S., Fan PhD, Y., Dunphy, R. N., Thompson, M. S., ... & Irene, L. 2015.

The Challenges of Data Quality Evaluation in a Joint Data Warehouse. *eGEMs, 3(1), 12.*

Ballou D. P., Pazer H., 1995 "Designing Information Systems to Optimize the Accuracy-timeliness Tradeoff." *IS Research, March 1995. pp. 51-72.*

Barkhuysen, P., de Grauw, W., Akkermans, R., Donkers, J., Schers, H., Biermans, M., ... & Peters, H. (2013). Is the quality of data in an electronic medical record sufficient for assessing the quality of primary care. *JAMIA, 21(4), 692-698.*

Bayley, K. B., Belnap, T., Savitz, L., Masica, A. L., Shah, N., & Fleming, N. S. 2013. Challenges in using electronic health record data for CER: experience of 4 learning organizations and solutions applied. *Medical Care, 51, S80-S86.*

Blumenthal, D., & Tavenner, M. 2010. The "meaningful use" regulation for electronic health records. *NEJM, 363(6), 501-504.*

Bonomi, S., Badr, N. G., Zardini, A., & Rossignoli, C. 2016 (May). Improving the Introduction of Electronic Health Record: Lessons from European and North American Countries. *In International Conference on Exploring Services Science (pp. 635-648). Springer, Cham.*

Bott, O. J. 2004. The electronic health record: Standardization and implementation. In *Proceedings of the 2nd Open ECG Workshop.*

Bowman, S. 2013. Impact of electronic health record systems on information integrity: quality and safety implications. *Perspectives in Health Information Management, 10(Fall).*

Boyle, D. I. R., & Cunningham, S. G. 2002. Resolving fundamental quality issues in linked datasets for clinical care. *Health Informatics Journal, 8(2), 73-77.*

Cimino JJ, Patel VL, Kushniruk 2002 AW. The patient clinical information system (PatCIS): technical solutions for and experience with giving patients access to their electronic medical records. *Int. J. Med. Inform. 2002; 68(1-3):113-27.*

Coleman, N., Halas, G., Peeler, W., Casaclang, N., Williamson, T., & Katz, A. 2015. From patient care to research: a validation study examining the factors contributing to data quality in a primary care electronic medical record database. *BMC family practice, 16(1), 11.*

De Amicis, F., Barone, D., & Batini, C. 2006, (November). An Analytical Framework to Analyze Dependencies Among Data Quality Dimensions. *In ICIQ (pp. 369-383).*

Dean F. Sittig, Ph.D., & Hardeep Singh, M.D., M.P.H. 2012 Electronic Health Records and National Patient-Safety Goals *NEJM 2012; 367:1854-1860 November 8, 2012*

Declerck, G., Hussain, S., Daniel, C., Yuksel, M., Laleci, G. B., Twagirumukiza, M., & Jaulent, M. C. 2015. Bridging Data Models and Terminologies to Support Adverse Drug Event Reporting Using EHR Data. *Methods Inf Med, 54(1), 24-31.*

Dixon-Woods, M., McNicol, S., & Martin, G. 2012. Ten challenges in improving quality in healthcare: lessons from the Health Foundation's programme evaluations and relevant literature. *BMJ Qual Saf, bmjqs-2011.*

Dolin, R. H., & Alschuler, L. 2011. Approaching semantic interoperability in Health Level Seven. *JAMIA, 18(1), 99-103.*

Feinstein, A. 1970. Quality of data in the medical record. *Computers and Biomedical Research, 3(5), 426-435.*

Gallego, B., Magrabi, F., Concha, O. P., Wang, Y., & Coiera, E. 2015. Insights into temporal patterns of hospital patient safety from routinely collected electronic data. *Health Information Science and Systems, (Suppl 1), S2.*

Garg, A. X., Adhikari, N. K., McDonald, H., Rosas-Arellano, M. P., Devereaux, P. J., Beyene, J., ... and Haynes, R. B. 2005. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *Jama, 293(10), 1223-1238.*

Gendron, M. S. & D'onofrio, M. J. 2001. Data quality in the healthcare industry. *Data Quality Journal 7, 1.*

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

Hennessy, M. Oentojo, C. & Ray, S. 2013 "A framework and ontology for mobile sensor platforms in home health management," in *Engineering of Mobile-Enabled Systems (MOBS)*, 2013 Workshop 2013, pp. 31-35

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

Hopf Y M, et al 2014 Views of healthcare professionals to linkage of routinely collected healthcare data: a systematic literature review; *JAMIA 1 February 2014*

Hristdis, V. 2009. "Data quality and integration issues in EHRs". In *Information Discovery on electronic Health Records*, Chapman and Hall, Ch. 4.

Jha A. K. et al 2009. "Use of Electronic Health Records in U.S. Hospitals," *NEJM, vol. 360, 16, (1628 – 1638)*

Jones, E., & Blavin, F. 2013. Lessons from the Literature on EHR Implementation (www.healthit.gov)

Jones, S. S., Koppel, R., Ridgely, M. S., Palen, T. E., Wu, S. Y., & Harrison, M. I. 2011. Guide to reducing unintended consequences of electronic health records. (www.rand.org)

Juran, J.M., Gryna, F.M. 1988 *Juran's Quality Control Handbook. 4th edn New York: McGraw-Hill, 1988*

Kahn, M. G. 1997. Three perspectives on integrated clinical databases. *Academic Medicine, 72(4), 281-6.*

Karkouch, A., Mousannif, H., Al Moatassime, H., & Noel, T. 2016. Data quality in internet of things: A state-of-the-art survey. *JNCA, 73, 57-81.*

Kaushal R, Shojania KG, Bates DW. 2003 Effects of computerized physician order entry and clinical decision support systems on medication safety: a systematic review. *Arch Intern Med 2003;163(12):1409-16.*

Leite, P., Gonçalves, J., Teixeira, P., & Rocha, Á. 2015. A model for the evaluation of data quality in health unit websites. *Health informatics journal, 1460458214567003.*

Lenert LA, Burstin H, Connell L, Gosbee J, Phillips G. 2002. "Federal patient safety initiatives panel summary". *JAMIA 9 (6 Suppl): S8–10. PMC 419408. PMID 12386172.*

Liaw, S. T., Rahimi, A., Ray, P., Taggart, J., Dennis, S., de Lusignan, S., ... & Talaei-Khoei, A. 2013. Towards an ontology for data quality in integrated chronic disease management: a realist review of the literature. *Int. J. Med. Inform. 82(1), 10-24.*

Lluch, M. 2011. Healthcare professionals' organisational barriers to health information technologies- A literature review. *Int J Med Inform, 80(12), 849-862.*

Lluch, M. 2011. Healthcare professionals' organisational barriers to health information technologies- A literature review. *Int. J. Med. Inform., 80(12), 849-862.*

Lu, Y. C., Xiao, Y., Sears, A., & Jacko, J. A. 2005. A review and a framework of handheld computer adoption in healthcare. *Int. J. Med. Inform., 74(5), 409-422.*

Magrabi, F., Baker, M., Sinha, I., Ong, M. S., Harrison, S., Kidd, M. R., ... & Coiera, E. (2015). Clinical safety of England's national programme for IT: A retrospective analysis of all reported safety events 2005 to 2011. *Int. J. Med. Inform.*, 84(3), 198-206.

Mangalmurti SS & Murtagh L, Mello. 2010. Medical malpractice liability in the age of electronic health records. NEJM. 2010 Nov 18; 363 (21):2060-7.

Mans, R. S., van der Aalst, W. M., & Vanwersch, R. J. 2015. Data Quality Issues. In Process Mining in Healthcare (pp. 79-88). *Springer Int. Publishing.*

Meystre S. M., et al, 2008 Extracting Information from Textual Documents in the Electronic Health Record: A Review of Recent Research, *IMIA Yearbook of Medical Informatics 2008*

Mikkelsen, G. & Aasly, J. 2005. Consequences of impaired data quality on information retrieval in electronic patient records. *Int. J. Med. Inf. 74, 5, 387–394.*

Moss, J., & Berner, E. S. (2015). Evaluating clinical decision support tools for medication administration safety in a simulated environment *Int. J. Med. Inform.* 84(5), 308-318.

Murff, H. J., FitzHenry, F., Matheny, M. E., Gentry, N., Kotter, K. L., Crimin, K., ... & Speroff, T. 2011. Automated identification of postoperative complications within an electronic medical record using natural language processing. *JAMA, 306(8), 848-855.*

Orfanidis, L., Bamidis, P. D., & Eaglestone, B. 2004 Data quality issues in electronic health records: an adaptation framework for the Greek health system. *Health informatics journal, 10(1), 23-36.*

Phillips, W., & Fleming, D. 2009. Ethical concerns in the use of electronic medical records. *Missouri medicine, 106(5), 328-333.*

Pipino, L. L., Lee, Y. W., & Wang, R. Y. 2002. Data quality assessment. *Communications of the ACM, 45(4), 211-218.*

Poissant L, Pereira J, Tamblyn R, Kawasumi Y. 2005 The Impact of Electronic Health Records on Time Efficiency of Physicians and Nurses: A Systematic Review. *JAMIA 2005; 12: 505-516.*

Qureshi, N.A. et al 2015 Electronic Health Records, Electronic Prescribing and Medication Errors: A Systematic Review of Literature, 2000-2014; *BJMMR, 5(5): 672-704, 2015; Article no.BJMMR.2015.072*

Reed, et al. 2012 Copy Functionality Toolkit: A Practical Guide: Information Management and Governance of Copy Functions in Electronic Health Record Systems. *Chicago, IL: AHIMA, 2012.*

Sachdeva, S., & Bhalla, S. 2012. Semantic interoperability in standardized electronic health record databases. *JDIQ, 3(1), 1.*

Singh H, Mani S, Espadas D, Petersen N, Franklin V, Petersen LA. 2009 Prescription errors and outcomes related to inconsistent information transmitted through computerized order entry: a prospective study. *Arch Intern Med. 2009 May 25; 169 (10):982-9.*

Sittig, D., and Singh, H. 2012 Electronic Health Records and National Patient-Safety NEJM 2012; 367:1854-1860 November 8, 2012

Srour, F. Jordan, & Badr, Nabil Georges. 2017 "Time Synchronization in Emergency Response Time Measurement." *In HEALTHINF, pp. 199-204. 2017.*

Srour, F. Jordan, & Badr, Nabil Georges. 2017 "Time Synchronization in Emergency Response Time Measurement." *In HEALTHINF, pp. 199-204. 2017.*

Wang, R. Y. & Strong, D. M. 1996. Beyond accuracy: What data quality means to data consumers? J. *Manage. Inf. Syst. 12, 4, 5–33.*

Wang, Y., Kung, L., Ting, CC & Byrd, TA 2015. "Beyond a Technical Perspective: Understanding Big Data Capabilities in Health Care," *Proceedings of 48th Annual HICSS, Kauai, Hawaii.*

Weir, C. R., Hurdle, J. F., Felgar, M. A., Hoffman, J. M., Roth, B., & Nebeker, J. R. 2003. Direct text entry in electronic progress notes. *Methods Inf Med*, 42(1), 61-67.

Weiskopf, N. G., & Weng, C. 2013. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *JAMIA, 20(1), 144-151.*

Wheeler, A. 2015. Designing a usable healthcare Information system. Mastering Informatics: *A Heatlhcare Handbook for Success*, 41.

Wickramage, C., Fidge, C., Sahama, T., Daly, A., & Wong, R. 2017. Preserving privacy through log analysis in health information systems.

Willis-Shattuck M, et al, 2008. Motivation and retention of health workers in developing countries: a systematic review. 4 Dec. 2008 *BMC Health Services Research 2008, 8:247*

Win, K. T., Croll, P., & Cooper, J. (2002). Setting a Safety Standard for Electronic Medical Records. *HIC 2002: Proceedings: Improving Quality by Lowering Barriers, 337.*

Yao, W., Chu, C. H., & Li, Z. 2011. Leveraging complex event processing for smart hospitals using RFID. *JNCA, 34(3), 799-810.*

Yun, J.H. & Kim, L.K. 2007. Processing HL7-CDA entry for semantic interoperability, *International Conference*

on Convergence Information Technology, pp. 1939–1944

Zafar, F., Khan, A., Suhail, S., Ahmed, I., Hameed, K., Khan, H. M., & Anjum, A. 2017. Trustworthy data: A survey, taxonomy and future trends of secure provenance schemes. *JNCA, 94, 50-68.*