An Approach Towards Information Quality Management of Electronic Health Records

Mohammad Mahdi Mahdavi Amjad, Kamyar Rasta, Martin Gerdes and Rune Fensli
Department of Information and Communication Technology, University of Agder, Grimstad, Norway

Keywords: Electronic Health Record, Information Quality, EHR Analysis.

Abstract: Electronic Health Records (EHR) management systems, have gained attention in many countries as public organizations working in health domain (such as hospitals and municipalities) use them to lift the quality of health care. EHRs are considered as the basis for establishing Health Information Service (HIS) systems. In spite of remarkable advantages of EHRs, the lack of quality metrics can reduce the efficiency of high level systems that are based on them. Particularly, it is smart to design a mechanism to assure the quality of data generated by sensors. In this paper, we propose an architecture in which EHRs are enriched with metadata to provide the information quality. Our architecture emphasizes on the quality of EHRs since we believe that the quality aspect of health records has not been contemplated in many commercial systems. We utilize some quality dimensions to produce the total quality metric. Moreover, we show that this architecture can provide quality-based systems with more appropriate inputs.

1 INTRODUCTION

As computers are increasingly used in health care realm, more applications are developed for managing electronic health records (EHR). Utilizing the EHRs have lifted the quality of health care (Gunter and Terry, 2005)(DesRoches et al., 2008). There are numerous applications out there and each one has its own method to store and analyze EHRs. We have studied some of them e.g. OpenMRS and OpenEHR and noticed that in almost all of them there is no mechanism for handling information quality. (Kalra et al., 2005)(Seebregts et al., 2009) As EHRs are typically raw data of health sector analyzing systems, the information quality of them is extremely important to improve.

The quality of EHRs must be ensured before being used in other higher level systems. On the other hand, the use of sensors has been drastically increased in the e-Health area and it means that there can be some errors in automatically generated measurements. Hence, the necessity of normalizing the automatically generated records shows itself.

In general, in case of using sensor networks, the probability of receiving wrong data from the subject can dramatically increase, especially if the error detection and error correction mechanisms have not been implemented (Akyildiz et al., 2002). Additionally, remarkable changes of the vital signs measurements are not usual incidents and beside other parameters, can be signs of emergency condition. This significant feature, can be used in notification systems but currently implemented systems do not support it, although there exist applications that have some thresholds for data or triggered alarm systems.

In this paper, we present an architecture which enhances EHR management systems by mechanisms to determine the information quality in terms of different quality dimensions. This architecture can be used to ensure that EHR data meet some basic criteria. Moreover, our system produces some metadata for each EHR that helps analyzing systems to achieve better performances.

2 INFORMATION QUALITY OF EHRS

To the best of our knowledge, almost none of the existing EHR management systems have mechanisms to determine the information quality of health records. This can degrade the result of EHR analyses especially when machines are responsible for generating, storing, and using them.

Majority of the EHR Management (EHRM) sys-

Mahdavi Amjad M., Rasta K., Gerdes M. and Fensli R..
An Approach Towards Information Quality Management of Electronic Health Records.
DOI: 10.5220/0004752402700275
In *Proceedings of the International Conference on Health Informatics* (HEALTHINF-2014), pages 270-275
ISBN: 978-989-758-010-9

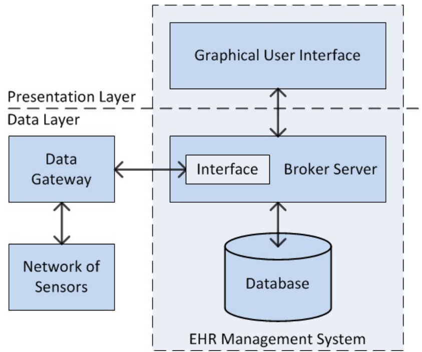

Figure 1: Conventional two-tier EHR management system.

tems use the conventional approach to manage the medical records. As Figure 1 shows, typically there is a relational database as a data storage space and a web/form application as the user interface which enables users to store and retrieve data into the storage and creates reports.

We have developed an application which can normalize EHRs in a relational database. This application has a collection of rules that can be applied to almost all types of EHR. For instance, assume that we have a system in which a sensor measures the body temperature of a subject and sends it to the recording system. The sensor stops functioning after a while and generates fake data. This can severely affect the quality of the health record and consequently the validity of analysis.

In spite of conventional approach, we propose an architecture in which a Quality-Enhanced Broker Server (Q-EBS) controls the EHR's quality. The Q-EBS is responsible for calculating and assigning the quality metric of each EHR. The quality metric is a value that represents the outcome of the quality dimension values. From information quality point of view, there are several popular quality dimensions (Wand and Wang, 1996) such as accuracy, timeliness, completeness, consistency, access security, relevancy, understandability, etc. which are used in many systems. However, in this paper we just deal with the first four and since connectivity is important in mobile networks, we consider it as another quality dimension as well. Other quality dimensions can be engaged if necessary.

2.1 Quality Dimensions

The computation of the final quality metric requires the values of the quality dimensions. In this section we briefly define our quality dimensions and the corresponding calculation equations. There is no consensus on the definitions of quality dimensions, and in most cases the definitions are customized according to the requirements of the system (Wand and Wang, 1996) (Baumgartner et al., 2010) (Strong et al., 1997).

2.1.1 Accuracy

The first quality dimension that we propose is accuracy. In our system, accuracy shows the ratio of the observed value and the real value. We name this real value as reference value. In real world there is no way to retrieve the reference (or the real) value. As any other measurement, ours is just a reflection of truth and therefore cannot be exactly true. One possible way for generating the reference value in health domain is to ask a nurse to manually produce it. Hence, we can consider this value as the reference and simply ignore the measurement errors.

The reference value can be generated by the system automatically as well. To do this, we should consider a proper time for system to reach to a reference value which can be simply the average of the values received in a past period of time. Nevertheless, the production process of a reliable reference value is out of the scope of this paper and so we do not focus on how to get the reference value. We just want to show how to calculate the accuracy of EHRs based on a given reference value.

Whether using an automatically or manually generated reference, we can compute the accuracy dimension values to ensure that incoming data is not produced by a malfunctioning sensor. Eq. (1) is used to calculate the accuracy of an observed value where r represents the reference value and o represents the observed value. Min and Max functions are utilized for calculating the minimum and maximum values of the reference and the observed values respectively.

$$Accuracy = \frac{Min(r,o)}{Max(r,o)} * 100 \quad (1)$$

Note that in reality there is no practical way to generate reference values for each EHR. Nevertheless, reference number is a concept that can be used for evaluating the quality of an observed data.

2.1.2 Completeness

The second quality dimension we use is the completeness. The completeness dimension, demonstrates the proportion of the number of received data items and the expected number of data items as a number between zero and hundred. In other words, the completeness shows how many data items are received from the sensor and how many of them we expect to receive. For example, assume that we have a device sending five values of blood oxygen saturation every one minute and a listener expecting to receive it. Using the completeness quality dimension, we can determine if the entire information package is received by the listener or not.

As our system knows about the number of data items in a package, it can evaluate the completeness of the received information and assign a quality dimension value to it. Therefore, the completeness of an information package can be calculated utilizing Eq. (2) in which $N_{received}$ shows the number of received data items by the system and N_{sent} represents the number of sent data items from the device.

$$Completeness = \frac{N_{received}}{N_{sent}} * 100 \tag{2}$$

2.1.3 Timeliness

The timeliness is the third proposed quality dimension when generating EHRs. Timeliness refers to the fact that the quality of an EHR can be sometimes strictly associated to the time it has been generated. Some EHRs must be committed to the system in a proper time to be useful. For instance, you can think of a notification system in which EHRs are used to detect the emergency condition of the subject.

In our systems, there is a time stamp for generation time and another one for the commitment time of each EHR. Technically, if the difference between these two times is more than the threshold, the EHR is not meaningful. The threshold, cab be defined according to the level of importance of the time in our system. For example, it is completely acceptable (but not desirable) for an SMS to be delivered twenty four hours after being sent but what about the report of a heart attack? Typically in mobile networks, frequent connections and disconnections are inevitable properties of the environment (Huang and Garcia-Molina, 2001). Therefore, the timeliness is very important when determining the quality of an EHR.

The timeliness quality dimension can have various definitions in different contexts. However, what we generally need in this paper is a number (preferably between 0 and 100) which shows the quality dimension value. All we need to compute the timeliness value, is to subtract the generation time (T generate) from the commitment time T_{commit} as it is shown in Eq. (3).

$$Timeliness = T_{commit} - T_{generate} \tag{3}$$

Table 1, explains how to convert the timeliness value to the quality dimension values between 0 and 100. With the aid of this mapping table, the timeliness becomes unified with other quality dimensions so that we can calculate the final quality metric.

Table 1: An example of timeliness mapping table.

T1	T2	T2 - T1	Timeliness
0	5	5	95%
0	10	10	90%
0	15	15	85%

Note that the timeliness dimension can be defined using other equations and mapping tables in respect to the system environment. To design a mapping table, the maximum value of the acceptable delay must be defined. This value is calculated considering the importance level of the expected data, required time for taking appropriate action, the type of network infrastructure (e.g. mobile, fixed, etc.), and/or any other influencing parameters.

2.1.4 Connectivity

Although connectivity is not among the popular quality dimensions, we believe that it can be useful in our architecture. As mentioned earlier, frequent connections and disconnections of components are inalienable properties of mobile environments. Additionally, mobile networks are very good environment for deployment of our architecture. Therefore, connectivity is very important and somehow it should be measured and stored as metadata of EHR.

To make it more clear, assume that we have some sensors measuring the heart pulse rate of a subject and a system monitoring the values they generate (i.e. the ECG curves). After a while, our monitoring system receives no signal from the sensor. This incident, can be interpreted at least in two ways: the loss of signal or the subject's heart attack. The loss of signal can be caused by sensor failure, network issues, etc. but the subject's heart attack must be confronted differently.

The connectivity quality dimension demonstrates the connection status of the network when an EHR is being generated and committed. We assign 0 to the connectivity dimension value when the sensor and/or data gateway are disconnected from quality manager and 100 when the connection is perfectly fine.

2.1.5 Consistency

The last quality dimension we use is the consistency. Consistency has a wide range of definitions in different environments, but we use it as a quality dimension which shows that whether EHR values are produced in proper range of validity. In many health records, especially those which are associated with measurements, there is a range of validity. For example, in EHR management systems there is a validity range for heart pulse rate (e.g. 0-230).

Although the probability of experiencing a heart rate out of this range is very low, it is possible to have an anomalous value due to sensor failure. In such a case, consistency is a very good dimension to determine the quality of the EHR. According to Eq. (4), the consistency is calculated by the proportion of the number of in-range data $(N_{in-range})$ and the total number of data (N_{total}).

$$Consistency = \frac{N_{in-range}}{N_{total}} * 100 \tag{4}$$

2.1.6 Quality Metric

Now that we have all required quality dimension values, we can calculate the quality metric. Quality metrics are shown as numbers between 0 and 100 so that bigger values show higher quality of the information. In fact, the quality is an outcome of different quality metric. As mentioned, several different quality dimensions can be considered. The final quality metric is computed using Eq. (5) in which N is the number of quality dimensions, (Q_i) is the quality dimension value and the (W_i) is the weight of each quality dimension.

$$QualityMetric = \frac{1}{N}\sum_{i=1}^{N} Q_i * Wi \tag{5}$$

As you can see in Eq. (5), we define a weight value (W_i) for each quality dimension. These values are used to control the effect of each quality dimension on the final quality metric. For instance, in some systems accuracy is more important than the timeliness. In such a system we assign lesser weight to the timeliness so that timeliness value can less affect the quality metric in comparison with accuracy value. This feature provides our system with more flexibility and causes more compatibility to diverse environments.

3 QUALITY-ENHANCED EHR APPROACH 1

Once we have calculated our quality metrics, we can concentrate on our architecture. As shown in Figure 2, we propose an architecture in which a quality manager is responsible for computing and assigning the quality metric to the health record. Typically, a standard EHR consists some information about the patient, measured values, date and time, but we need an EHR containing quality data to calculate the quality dimensions and quality metric. If a new system is going to be developed, then these quality data can be easily considered in the system. But if we want to add quality to an existing system, then extra information must be added to the original EHR and it means that some modification in the data gateway is necessary. Additionally, we need to modify the existing storage space of EHRs.

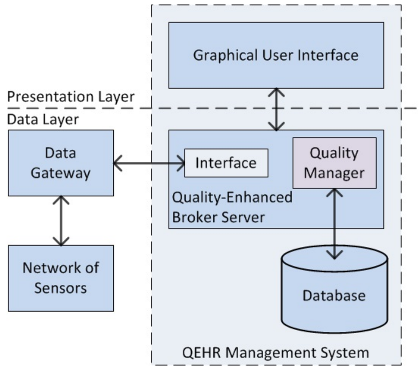

Figure 2: Quality-enhanced EHR management system.

As mentioned earlier, the majority of EHRM systems, use relational databases. Therefore, when adding the quality metadata to the original EHR, modification of involved tables, views, stored procedures and functions and other database objects must be performed. Moreover, web/form interfaces which are used for data entry should be updated as well.

When the data gateway sends quality-enhanced data to the Q-EBS, the quality manager extracts the quality data, calculates the quality dimensions and the quality metric and stores them to the database. This approach enables us to have the quality information right beside the original EHR. Therefore, there is no need to add any extra component to the current system except the quality manager. The quality manager, can be implemented as a single (while big) stored procedure or function inside the database or an application manipulating it.

4 QUALITY-ENHANCED EHR APPROACH 2

As a major possible modification of the first approach, we can design another independent component on the sensor side which manages the generation of information quality metadata. We name this component as quality agent. The quality agent can accept various roles in our architecture. For instance, it can just retrieve quality data from the data gateway and send it to the quality manager in order to be processed. The quality agent can also calculate the quality metric and only send the results.

Figure 3 shows a possible scenario of using this component in our architecture. The principles of this scenario are the same as what is shown in Figure 2, except that the quality agent component has been added to it.

5 DISCUSSION

It is very important to decide on the kind of information that should be sent to the quality manager since this can affect the performance of the whole system. Moreover, network traffic, process distribution on the smartphone and the EHRM server and the complexity of implementation in term of updates propagation are directly associated with this decision. Table 2, demonstrates the effects of distribution strategies on these aspects of the system, where traffic means the required bandwidth for communications between quality agent and the quality manager, process means the smartphone process load, and the number of update refers to the number of necessary updates in case of modification of quality computation rules.

Table 2: The effects of sending quality data and quality metric on various aspects of system performance.

Sent Item	Traffic	Process	No. of Updates
Q Data	↑	↓	↓
Q Metric	↓	↑	↑

There are some disadvantages with the quality agent approach. However, it should be considered as a possible option in mobile environments. Let us briefly describe some of its advantages and disadvantages. First of all, we must consider that having a quality agent on sensor side (i.e. smartphone) means that we have to completely split our architecture into two sections: EHR management and information quality management. It can result in a lot of overheads in terms of process, communication, storage, security, etc. On the other hand this approach enables us to manage the information quality in an independent way so that there will be no need to modify the inner structure of existing systems. The only manipulation needed, is to store the metadata (i.e. quality dimensions and quality metric) in another database with a link to corresponding EHR.

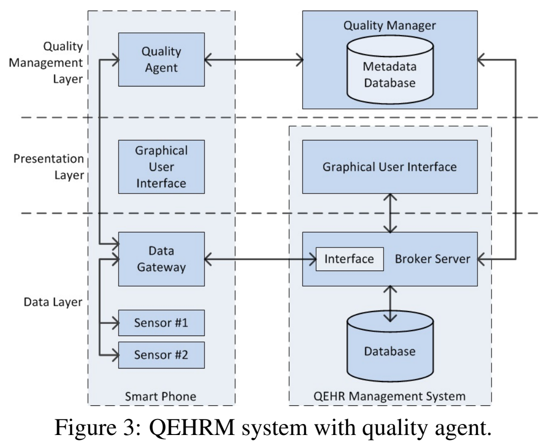

Figure 3: QEHRM system with quality agent.

Secondly, computation of quality metric and then sending it, may impose more process on the smartphone since the quality agent can be implemented inside a smartphone. On the contrary, if the quality agent calculates and sends the quality metric, then we can decentralize the process and gain better network efficiency since smaller data packets are transferred through the network.

6 CONCLUSIONS

In this paper we proposed an architecture for assuring the information quality of EHRs. Generally, our architecture is based on five quality dimensions that can be calculated using some metadata which are added to the original EHR. We engaged four simple equations for calculation of EHR quality dimensions and another one for computation of quality metric. Additionally, we proposed two different implementation alternatives for diverse environments. According the section 5, the first architecture is more suitable for fixed networks where the network bandwidth is not our most important concern. Instead, we probably prefer to distribute the process load of the centralized server to the clients.

The second approach is more appropriate for mobile networks where we want to use our valuable bandwidth in the most efficient way. Nowadays, smartphones with high performance processors can be found almost everywhere, so putting the burden of some computation on smartphones seems trivial.

6.1 Future Work

As semantic web grows, more semantic-enhanced applications are developed and migrating from relational databases to ontologies becomes a must since new customers have requirement that make conventional approaches to face serious challenges (Berners-Lee et al., 2001). Our proposed system uses some statistical techniques to ensure the quality of information. A magnificent improvement of our system can be the enhancement with semantic data and functions. This can lead us to a system that can more precisely detect the anomalous information and treat them in a better way (Chandola et al., 2009).

Another component that can be used in our system, is an intelligent EHR controller which is able to improve itself to make better decision in case of facing abnormal records.

REFERENCES

Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., and Cayirci, E. (2002). A survey on sensor networks. *Communications magazine, IEEE*, 40(8):102–114.

Baumgartner, N., Gottesheim, W., Mitsch, S., Retschitzegger, W., and Schwinger, W. (2010). Improving situation awareness in traffic management. In *Proc. Intl. Conf. on Very Large Data Bases*.

Berners-Lee, T., Hendler, J., Lassila, O., et al. (2001). The semantic web. *Scientific american*, 284(5):28–37.

Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3):15.

DesRoches, C. M., Campbell, E. G., Rao, S. R., Donelan, K., Ferris, T. G., Jha, A., Kaushal, R., Levy, D. E., Rosenbaum, S., Shields, A. E., et al. (2008). Electronic health records in ambulatory carea national survey of physicians. *New England Journal of Medicine*, 359(1):50–60.

Gunter, T. D. and Terry, N. P. (2005). The emergence of national electronic health record architectures in the united states and australia: models, costs, and questions. *Journal of Medical Internet Research*, 7(1).

Huang, Y. and Garcia-Molina, H. (2001). Publish/subscribe in a mobile enviroment. In *Proceedings of the 2nd ACM international workshop on Data engineering for wireless and mobile access*, pages 27–34. ACM.

Kalra, D., Beale, T., and Heard, S. (2005). The openehr foundation. *Studies in health technology and informatics*, 115:153–173.

Seebregts, C. J., Mamlin, B. W., Biondich, P. G., Fraser, H. S., Wolfe, B. A., Jazayeri, D., Allen, C., Miranda, J., Baker, E., Musinguzi, N., et al. (2009). The openmrs implementers network. *International journal of medical informatics*, 78(11):711–720.

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

Wand, Y. and Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, 39(11):86–95.