DATA QUALITY ANALYSIS FOR E-HEALTH MONITORING APPLICATIONS

Claudia Gutiérrez Rodríguez
I3S – CNRS - University of Nice - Sophia Antipolis
930 route des colles – BP 145, 06902 Sophia Antipolis, France
cgutierr@i3s.unice.fr

Michel Riveill
I3S – CNRS - University of Nice - Sophia Antipolis
930 route des colles – BP 145, 06902 Sophia Antipolis, France
riveill@i3s.unice.fr

ABSTRACT
Nowadays, technological improvement has been a motivation to develop new healthcare programs and approaches to better assist chronic patients. E-Health monitoring refers to a continuous observation of patient’s condition (physiological and physical) traditionally performed by one or several body sensors. Data coming from such sensors are then transferred to a server, then monitored, processed and accessed by medical experts. In this kind of applications, very important actions and decisions are based on sensed data (as remote diagnosis, consultations, emergency intervention, hospitalization, etc.) and thus a high level quality of data is essential to manage healthcare effectively. Data quality in this domain is an important issue that has to be addressed. Data Quality, being important part of information management, helps to ensure the correct processing of collected data as well as the appropriate interpretation and intervention of medical services. In this paper we explore the issues that define data quality in e-health applications.

KEYWORDS
Data quality, e-Health applications, medical monitoring, health informatics

1. INTRODUCTION
In medical domain, the improvement of ICT (Information and Communication Technologies) strongly helps to provide better quality of healthcare, using for example, high-technology sensors (i.e. pulse, body temperature, ECG, etc.), wired and wireless communications technologies, real-time data processing, interactive interfaces, etc. This improvement introduces “e-Health” which is considered actually as a primary engine to assist patients and which combines the use of electronic communication and information technology with the use of digital data transmitted, stored and retrieved electronically, both locally and at distance.

The monitoring of patients’ condition with chronic diseases is now possible by continuously recording and processing their vital signs or activity every day. Traditionally, data coming from patients sensors are transferred (at real or differed time) via wired or wireless communication to a server and being analyzed, monitored and managed by medical professionals. This kind of applications enables to capture precisely atypical symptoms or activities at any time, and providing accessibility of the healthcare independently of geographic location. Patients can be monitored at their natural environment (Varshney, 2007).

However, the management of data in this kind of systems is becoming increasingly complex. Sometimes decision makers (medical experts or professionals, users…) are confronted with inaccurate, incomplete or too much information. As a result, more and more questions concerning data quality, security and privacy in this domain arise. Ensuring the quality of data in healthcare domain remains a critical aspect, if data quality is ignored, collected information may have a considerably negative impact on the achievement of the application and decision making. For e-Health monitoring applications, the problem is particularly more
difficult to avoid due to the criticality of the domain. In this paper, we claim that data quality concerning patients’ monitoring and assistance cannot be neglected. We introduce an analysis of data quality issues in this particular domain. This analysis is associated within a project, grouping industrial and academic research teams, users and manufacturers.

2. E-HEALTH MONITORING SCENARIO

In order to better integrate healthcare systems between patients and medical services, more and more patient-centric approaches are exploiting pervasive and ubiquitous infrastructures allowing patients to be an active part of their own healthcare (i.e. Medic4you, Health Guide, Medmobile…). We base our analysis of data quality on a current research project which proposes a monitoring for patients with chronic pathologies requiring continuous surveillance and medical assistance (Figure 1). This project proposes to integrate a secure medical monitoring (using IMDs – Implant Medical Device) with wireless transmission allowing a complete mobility of a patient (i.e. using a Smartphone) and being continuously monitored by medical experts. Also it is associated to the development of a data hub (i.e. MicroSD card) and to dedicated Human Machine Interfaces (HIMs) opening new perspectives for e-Health monitoring applications.

![Figure 1. e-Health monitoring application](image)

The scenario that we explore in this paper is the monitoring of patients with cardiac problems and equipped by with a cardiac IMD. First it is necessary to establish a communication between the IMD and the external programmer at medical center (hospital or clinic) during the IMD implantation. This communication is essential in order to setup the device and prepare it to follow-up data (FU – data collection). In fact, during the implantation the parameters of the IMD are fixed allowing monitoring the device. Next, in this use case, two patients monitoring are considered: one in real-time which can be continuous, triggered or on-demand and a second one in differed-time (at FU for instance). In both cases, the sensed data can be pre-stored and pre-processed at implant side, several warnings and pre-diagnosis can be programmed at this point. Collected data is then transferred via 3G/GSM/GPRS (en real-time or a posteriori) to a back-end server for much complex analysis and processing. Moreover, some consultations over the year are scheduled with the patient. These consultations (routine, triggered by warnings or on-demand) are principally oriented to follow-up data, control the implant and verify patient’s condition. Regarding traditional healthcare applications, these consultations can be performed remotely and a constant monitoring can be also associated. Nevertheless, the introduction of IMDs, mobile devices, wireless communication and others comes with some quality, security and privacy issues. More and more medical experts are confronted to a lot of information, sometimes inaccurate or incomplete. If data quality is ignored, acquired information may have a considerably negative impact on the achievement of the application and decision making.

3. DATA QUALITY IN E-HEALTH MONITORING APPLICATIONS

In the last years, technological improvement opens new possibilities to healthcare and practice medicine, but carries some inherit risks and leaves decision makers with numerous unanswered questions about quality,
security, privacy, ethics, risk management and other important matters. Some surveys and approaches have showed the importance of data quality for end-users, in particular, in healthcare domain (Shaw et al. 2009). Many quality criteria have been proposed without a general consensus; in fact each domain has its specific vision of data quality as well as the solutions to solve the quality problems (Wang et al., 1996). Data quality is often considered as “fitness-for-use”, it is based on the specific use of data and the requirements to be satisfied.

To tackle this issue, based on existing quality modeling approaches (Wang et al. 1996), (Naumann et al., 2000), (AHIMA, 1998), we propose to analyze the e-health monitoring systems according to appropriate quality criteria. Such strategy aims to manage data quality at every point along its transition minimizing poor data quality spread.

3.1. Impacting data quality

To correctly identify data quality issues, we have firstly to identify the source of the quality problems, analyze its impact and, where possible, propose a solution. In our particular application domain, there are many contextual reasons why it is difficult to maintain a good quality of data. Some difficulties are related to technology (i.e. equipment, QoS (Quality of Service), to human intervention (input errors, misunderstanding…), or to process of data transformation (i.e. optimal analysis and processing). Consequently, we study the characteristics of e-Health monitoring applications according to data flow (from data source until destination) and we try to determine where, when and how an impact over data quality happen. For this, we define three main levels of data management and processing over the system (Figure 2), defined as: Data collection (sensor data collection, pre-processing and transfer), Data processing (data processing, storage and delivery) and Data discovery (data access, enhancement and discovery).

![Figure 2. Data management and processing levels](image)

We believe that the context in which data are collected (Data Collection level) is a crucial aspect to be considered for data quality assessment. At this level, the quality of data can be impacted by the rate of data collection (too fast, too slow…), the correct performance of body sensors (battery, sensors life time, setting…), the quantity of data to be pre-processed and transferred (i.e. respecting data quote) as well as the quality of communication (broadband, frequency…). We claim that at this level it is necessary to implement a quality procedure in order to validate or qualify data and its context, before they arrive to back-end servers or before they are discovered by the users. For example, if several warnings are triggered from a critical pre-diagnosis at patients’ side, we have to ensure as accurate as possible the data transmitted to medical services or the patient itself. In this kind of applications, patient can be also allowed to monitor himself in real-time and thus any information with poor quality can impact its behavior.

Data processing level refers to all the processes responsible to transform and store sensed data. Thus, at this level more analysis and data enhancement can be performed. As we show in figure 2, the processes are executed at back-end server which is also considered as a data repository. Sensed data are then integrated with more heterogeneous data as EPRs (Electronic Patient Records), medical images, videos, etc. which are normally provided by external sources as medical services or others. In such a case, we are confronted with more information often provided manually and with inaccurate or incomplete information. It is also important to guarantee the data accessibility and the respect of privacy constraints. Data has to be as available and freshness as possible in order to provide a performing monitoring, also only allowed persons has totally or partially access to certain information. We think that current quality approaches at the domain of DIS
(Data Integration Systems) or DW (DataWarehousing) can be adapted to evaluate and control data quality at this level. Finally, at data discovery level, the system has to guarantee a good decision making based on reliable and secure data. We estimate important to control the quantity of data communicated to the users as well as the quality of the representation (i.e. consistency, understanding, etc.). Users (not always experts) can be confused with too much information and by the way as information is represented and communicated.

3.2. Towards data quality criteria

We observe that in medical domain the big picture of data quality is generally illustrated by the accuracy and reliability of data. The more accurate and reliable data is, more confident and relevant decisions will be taken by the actors (patients, medical experts, medical services). However, we estimate that on our context other complementary perceptions of quality are also necessary. Data availability, completeness, freshness and others are also important to certify the accuracy and reliability of data. Our perception of data quality is related to a set of quality dimensions referring the characteristics of data processing and management levels over the system, associated to several quality criteria and to a quality evaluation procedure. Thus, in order to define the optimal quality criteria, we decide to analyze, at first place, the pertinence and usefulness (or applicability) of the basic and most used quality criteria (i.e. (AHIMA, 1998)) as: Accuracy, Precision, Accessibility, Currency or Freshness, Consistency, Relevancy, Comprehensiveness. Comparing the specificities of e-Health monitoring applications and the goal of each quality criteria, we found some important correlations. In this paper, we focus specially on underlying the quality criteria principles over data collection.

Considering accuracy for example, it is necessary to specify how valid and error free are data coming from body sensors, particularly to ensure integrity, validity and reliability of all the collected data. Together, precision contributes to complete data validation. Data can be not accurate but precise enough to ensure data reliability. Thus, some ranges and categories describing data precision have to be defined. Besides, these collected data (after a pre-processing) must be available as necessary. Thus, accessibility must ensure to provide legal access as well as the required amount of data according to users description, goals, etc. Currency guarantees data up-dating. Definitions for currency or freshness for each type of obtained data must be determined (i.e. data are up-date within 2 seconds, 2 hours, 2 days). Also, regarding the characteristics of sensed data (raw and pre-processed), we have to ensure their consistency to data specifications and goals.

4. CONCLUSION

This paper describes an ongoing research dedicated to analyze data quality issues in a critical domain as e-Health monitoring. We note that this is a first attempt to analyze and define data quality in this kind of applications, and naturally this aspect requires further investigation from the research community. For example, modeling users and system requirements and associate them to quality evaluation methods, algorithms and procedures. Also, we note that the quality criteria presented previously are the core of data quality approaches but they are not exhaustive. Since the most part of quality criteria depends on the specificities of the environment and the user requirements, we plan to adequate other perspectives of quality as Quality of Service (QoS), especially on data collection level.

REFERENCES