

# Data Streams Quality Evaluation for the Generation of Alarms in Health Domain

Saúl Fagúndez, Joaquín Fleitas, and Adriana Marotta<sup>(✉)</sup>

Universidad de la República, Montevideo, Uruguay  
{sfagundez, jfleitas, amarotta}@fing.edu.uy

**Abstract.** In this paper we present a proposal for managing data streams from sensors that are installed in patients' homes in order to monitor their health. It focuses on processing the sensors data streams taking into account data quality. In order to achieve this, a data quality model for this kind of data streams and an architecture for the monitoring system are proposed. Besides, our work induces a mechanism for avoiding false alarms generated by data quality problems.

**Keywords:** Data quality · Data streaming · Smart home · Health monitoring · Sensors · Data quality model

## 1 Introduction

The use of sensors has had an enormous increment in last years, becoming a valuable tool in many different areas, such as weather forecasting, driving assistance, water level and quality monitoring, and health monitoring. Sensors produce data streams, which in general have a simple structure but are generated with a very high rate. In this kind of scenario quality of data becomes an extremely important issue, especially in cases where critical decisions must be made based on obtained data. However, not much attention has been paid to this specific topic, existing only few works that focus on it.

Health monitoring through the use of sensors is sometimes used for elderly people care. Different kinds of sensors are installed in their homes and on the patients themselves in order to monitor their behavior as well as their vital signs (blood pressure and temperature). Their behavior is important for example in the case of patients suffering from Alzheimer Disease. Data provided by the sensors are directly transmitted to the hospital, so that the patient can be monitored, avoiding his movement from one place to another. The data received at the hospital are continuously evaluated through a monitoring system, which generates alarms when suspicious data is detected.

This work is situated in the context described in the previous paragraph, and focuses on processing the sensors data streams taking into account data quality. In order to achieve this, a data quality model for health sensors data streams and an architecture for the monitoring system are proposed.

The main contribution of this work is the proposal of a data quality model specific to data streams coming from home and on-patient sensors.

Data quality (DQ) is represented by quality dimensions, each one representing a different quality aspect [1–4]. Our approach is based on a DQ meta-model that consists

in: a *quality dimension* that captures one aspect of DQ, a *quality factor* that represents a particular aspect of a DQ dimension, and a *quality metric* that defines the criteria for the measurement of the quality factor. Metrics may be applied to data objects at different granularity levels, e.g. a data item or a set of data items. The DQ dimensions that are used along this work are the following: *Accuracy*, *Completeness*, *Consistency*, and *Freshness*, which are based on wellknown concepts, about which there is general consensus in the DQ literature [1–4].

According to [5] a data streaming is a continuous and ordered sequence of elements, where elements are presented in real-time. The mechanism for applying dynamic queries to the streams of data is through data windows, which capture certain portions of data from the streams. A window may be logically defined considering the number of elements, or physically defined considering the duration, i.e. the data that arrive to the stream during certain time period [6]. A DSMS (Data Stream Management System) provides a data model for managing dynamic data streams and continuous queries, which are constant queries that process data as they arrive.

As we previously said, some work has been done in the area of sensor data streams quality. In [7] the authors state that quality restrictions in this kind of data must not be ignored and should be carefully managed so that an exhaustive evaluation can be done. This is especially important in applications that directly consume sensor data and their quality becomes a critical issue. In some other applications data from sensors are stored in a database in order to be processed later. In these cases data quality is still essential for decision making supported by data. In [8] a data streaming meta-model is proposed in order to allow the propagation of data quality information towards the corresponding business application. A quality model is presented (five data quality dimensions are managed) and the impact of data stream processing operators on data quality is analysed. The authors focus their analysis on accuracy and completeness quality dimensions. Meanwhile, in [9] the authors present a model that is based on an intuitive notion of the sensor data completeness. They measure the quantity of data that arrive to a consuming point and compare it with the maximum possible quantity of data at that point. On the other hand, in [10] some mechanisms for reducing energy consumption in sensor networks are proposed. These mechanisms assure certain level of data quality, so that they provide a balance between energy efficiency and quality of aggregate data. In [11] a probabilistic approach is used for evaluating the quality of sensor data, modelling the uncertainty in sensor readings. Finally, an event based solution for improving quality in data streams exchanged between health organizations is proposed in [12]. They focus on two quality aspects: data consistency and duplicate detection. They use alerts to notify detected problems.

Our work manages a broad set of quality dimensions and its model and mechanism can be naturally extended with more dimensions. It also distinguishes different quality factors, which allows a more detailed and complete study of the data quality. In addition, besides defining the data quality model, our work induces a mechanism for avoiding false alarms generated by data quality problems.

The rest of the paper is organized as follows. Section 2 presents the proposed system architecture, Sect. 3 focuses on the data quality model and its application, and Sect. 4 presents the conclusions and future work.

## 2 Health Monitoring System

We consider a smart home with three rooms; a bedroom, a kitchen and bathroom, and a person suffering from Alzheimer's disease. Each room is equipped with two ultrasound distance sensors which measure the distance of some object to the sensor. When the person is in the room, the sensors report the distance to that person. We also have two on-body sensors: one of blood pressure and a thermometer.

At the same time, there is a system which receives and manages data from the sensors in order to detect whether the person has certain variations in his behaviour or in his vital signs. It is a real time and autonomous system, which is able to analyse the data streaming coming from different sensors and to send alarms in predefined situations.

### 2.1 Proposed Architecture

The proposed architecture consists on different components that are shown in Fig. 1. The user's access point to the system is the *Monitoring* component, where he should first define the requirements, quality parameters, and alarms needed in his particular context. Then, the *Middleware* is responsible for managing the execution of distributed and dynamic queries. The *Data Quality Manager* is responsible for measuring the level of quality of data obtained from the queries and enriching data with quality values. This module interacts with the *Middleware*, so that the *Middleware* is able to return the data window enriched with quality values to the *Monitoring* component. A database containing historical blood pressure data is maintained and queried by the *Data Quality Manager* component in order to evaluate the accuracy of pressure values. The *Monitoring* component is responsible for carrying out the monitoring of the person at home. Some of its functions are to control the temperature and blood pressure of the patient and know in which room of the house he is located. It includes a system of alarms that are activated according to the parameters set and the information obtained from sensors. The *Data Processing* component has the functionality of managing information obtained from the *Middleware* and returning the result to the *Monitoring* component.

## 3 Data Quality Model and Management

Several data quality issues must be considered in the proposed scenario: (1) Possible errors locating the person in the house because of wrong sensor measurements. (2) Absence of sensor measurement during a predefined time period. This problem encompasses all types of sensors. (3) Blood pressure sensor values whose measures are higher than normal values that are expected according to historical data of person's blood pressure. This may be due to a health problem in the patient (an alarm should be sent, see Sect. 3.1) or due to a data quality problem. (4) For both blood pressure sensors and temperature sensors data, there are a maximum and a minimum valid value that should be respected. (5) Adequate sensor measurement rate. When increasing sensor measurement rate, energy cost and network traffic increase, therefore, this rate should balance the data frequency needed and the energy and traffic supported by the system.

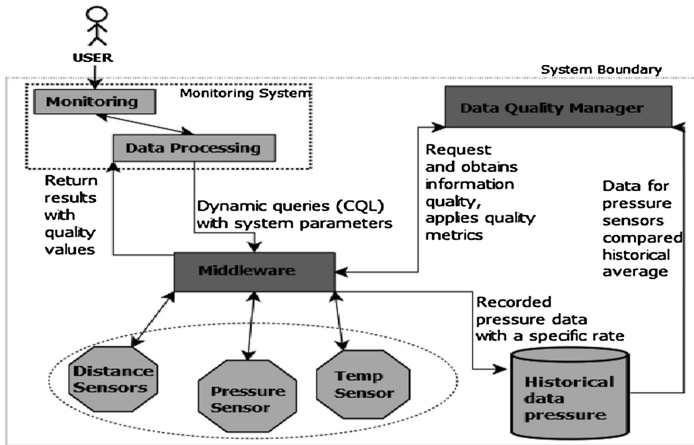


Fig. 1. Architecture

Taking into account the previously described problems and in order to manage them, we define a data quality model that specifies a set of metrics to be applied to the involved data. Table 1 shows the defined data quality model. A metric is defined for each quality factor applied to a kind of sensor, for example, *Dist-Prec* metric is defined for *Precision* factor applied to *distance* sensors data.

*Dist-Prec* verifies if the sensor satisfies the minimal distance between the sensor and the person from which the sensor values have enough precision (issue 1).

Table 1. DQ model

Dimension	Factor	Sensor	Metric
Accuracy	Precision	Distance	<i>Dist-Prec</i> : Satisfaction of minimum distance supported by the sensor Result range: (0..1)
	Semantic Accuracy	Pressure	<i>Pres-SAcc</i> : Satisfaction of a maximum threshold calculated from historical personal pressure data Result range: (0..1)
Completeness	Density	Distance	<i>Dist-Dens</i> : Non-null values quantity Result range: (0..1)
Freshness	Currency	Distance, Temperature, Pressure	<i>Dist-Curr</i> , <i>Temp-Curr</i> , <i>Pres-Curr</i> : Time period between two consecutive data windows issued by a sensor Result range: (0..1)
Consistency	Domain Integrity	Distance, Temperature, Pressure	<i>Dist-Dom</i> , <i>Temp_Dom</i> , <i>Pres-Dom</i> : Correspondence of the sensor values to a pre-defined domain: positive integer in a pre-defined range Result range: (0..1)

*Pres-SAcc* evaluates if blood pressure values are out of the expected values, in which case there can be a data quality problem or a situation that deserves special attention (issue 3). *Dist-Dens* is applied to distance sensors because a minimum quantity of non-null sensor values are needed for calculating a person location in a room (issue 1). Each sensor should issue data with a minimum frequency that is defined in the system; the metrics for *Currency* factor are used to verify the satisfaction of this requirement (issues 2 and 5). The metrics for *Domain Integrity* factor control if the sensor values belong to certain integer ranges, which are defined in the system (issue 4).

The granularity for all defined metrics is: data window. Note that *Pres-SAcc* and *Domain Integrity* metrics calculate the result considering the quantity of the window values that satisfy the required condition.

DQ information must be attached to the data streams. Figure 2 shows the conceptual schema corresponding to the data window and the DQ information attached to it.

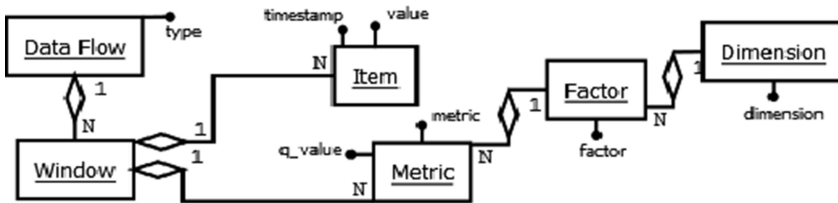


Fig. 2. Data window with DQ information

**Example:** Consider the distance sensors in the home rooms. Each data stream sent from the Middleware to the Data Processing component has the format shown in Table 2.

Table 2. Distance data stream with quality information

Timestamp	10	20	30	40	50	60	70	80	90
Distance value	2,5	2,5	2	2	0,3	0,4		2,5	2,5
Precision	1			0,3			1		
Density	1			1			0,7		
Currency	2			2			2		
Domain Integrity	1			1			1		

Quality values are calculated using the proposed quality model applying the respective quality metric for distance sensors over the data windows. In this example, in the second data window there is a problem of sensor precision, since some of its values are lower than the minimum for the sensor, so Precision value = 0,3. Meanwhile, in the third window there is a NULL value, so Density value = 0,7.

The Data Processing component integrates data from all distance sensors of the rooms, detecting where the person is located in the home and calculating associated quality information. Table 3 shows an example of the generated data stream.

We consider a range of 10 min and the rooms of the house: Bedroom “Be”, Kitchen “K” and Bathroom “B”. In the table we can see that the system returns the location of the patient, and the corresponding quality values, using a window of size 3.

**Table 3** Data stream generated by Data Processing component

Timestamp	10	20	30	40	50	60	70	80	90
Location	Be	Be	Be	K	K	K	B	B	B
Precision	0,8			1			0,9		
Density	0,8			0,7			0,8		
Currency	1			0,8			1		
Domain Integrity	1			0,8			0,9		
Inter-sensor consistency	1			1			1		

### 3.1 Alarms Generation

The system’s main function is to monitor the person at home using the installed sensors. This is achieved through the analysis of the data streams, considering the parameters set by the user as well as the quality of the data. Depending on this analysis, different outputs will be obtained. If sensor errors are detected certain alarms will be generated, while if potential patient health problems are detected other alarms are generated. The following is an example of a possible situation that generates alarms:

In order to detect where the person is located in the house, two distance sensors are placed in a room so that the system can get the position of the person.

- If information is missing from one or both sensors (metric *Dist-Dens*) then the system returns a **DQ alarm** that indicates the DQ problem encountered.
- If the person is located in two rooms simultaneously (metric *Dist-Dom*) then the system returns a **DQ alarm** that indicates the DQ problem encountered.
- Otherwise, if two distance sensors locate the person in a room, and in a pre-established time period other two sensors detect the person in another room, and this behaviour is repeated for another pre-established time period, then this could be a risk of agitation of the person, so the system returns a **health alarm**.

## 4 Conclusions

In this paper we present a proposal for managing data streams from sensors that are installed in patients’ homes in order to monitor their health.

A set of possible problems in the sensors data are described and taking into account these problems, a DQ model is proposed. In addition, an architecture for the system that is in charge of processing sensor data is proposed. Finally, an example of the generation of alarms is presented.

The DQ model presented in this work was specifically designed for a particular context with particular kinds of sensors. However, this proposal can be seen as a step towards the definition of a general DQ model for sensor data streams.

As future work, a deeper study on general problems in sensor data streams and on the most appropriate DQ dimensions, factors and metrics for this kind of data may be carried out. Also, quality metrics implementation using the particularities of DSMSS should be deeper studied.

## References

1. Batini, C., Scannapieco, M.: *Data Quality: Concepts, Methodologies and Techniques*. Springer, Heidelberg (2006)
2. Strong, D.M., Lee, Y.W., Wang, R.Y.: Data quality in context. *Commun. ACM* **40**, 103–110 (1997)
3. Pipino, L., Lee, Y.W., Wang, R.Y.: Data quality assessment. *Commun. ACM* **45**(4), 211–218 (2002)
4. Wang, R.Y., Strong, D.M.: Beyond accuracy: what data quality means to data consumers. *J. Manage. Inf. Syst.* **12**(4), 5–33 (1996)
5. Golab, L., Tamer Özsu, M.: Issues in data stream management. *SIGMOD Rec.* **32**(2), 5–14 (2003)
6. Hebrail, G.: Data stream management and mining. *Mining Massive Data Sets For Security*, Paris, pp. 89–102 (2008)
7. Klein, A.: *Incorporating quality aspects in sensor data streams*. ACM first Ph.D. (2007)
8. Klein, A., Lehner, W.: Representing data quality in sensor data streaming environments. *J. Data Inf. Qual.* **1**(2), 10:1–10:28 (2009)
9. Bitwas, J., Naumann, F., Qiu, Q.: Assessing the completeness of sensor data. In: 11th International Conference on DASFAA, Singapore (2006)
10. Sharaf, M.A., Beaver, J., Labrinidis, A., Chrysanthis, P.K.: Balancing energy efficiency and quality of aggregate data in sensor networks. *VLDB J.* **13**(4), 384–403 (2004)
11. Kuka, C., Nicklas, D.: Quality matters: supporting quality-aware pervasive applications by probabilistic data stream management. In: *Proceedings of the 8th ACM International Conference on Distributed Event-Based Systems*, New York, NY, USA, pp. 1–12 (2014)
12. Berry, A., Milosevic, Z.: Real-time analytics for legacy data streams in health: monitoring health data quality. In: *17th IEEE International Enterprise Distributed Object Computing Conference (EDOC)*, pp. 91–100 (2013)