2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Data Quality Characteristics for Improved Metrology in Sensor Networks

Mads Johansen[1*], Anupam Prasad Vedurmudi[2], Martha Arbayani Zaidan[3], Milos Davidovic[4], Gertjan Kok[5], Maitane Iturrate-Garcia[6], Shahin Tabandeh[7]

[1]*FORCE Technology, Hørsholm, Denmark*
[2]*Physikalisch-Technische Bundesanstalt (PTB), Braunschweig, Germany*
[3]*University of Helsinki, Helsinki, Finland*
[4]*Vinca Institute of Nuclear Sciences (VINS), Belgrade, Serbia*
[5]*Van Swinden Laboratorium (VSL), Delft, Netherlands*
[6]*Federal Institute of Metrology (METAS), Bern-Wabern, Switzerland*
[7]*Technical Research Centre of Finland (VTT), Helsinki, Finland*
[*]*Corresponding author. E-mail address: majh@forcetechnology.com*

Abstract – **Sensor networks are becoming increasingly practical to deploy in largely varying settings, which combined with the growing availability of low-cost sensors and the increasing scale of sensor networks, makes it highly challenging to ensure the trustworthiness and reliability of measurements and data. Factors such as physical inaccessibility and cost constraints make it infeasible to use established methods for calibration, further increasing the difficulty of assessing measurement uncertainty and ensuring traceability in sensor networks. In addition, the large volume of data generated makes the assessment of data quality in sensor networks infeasible without automated, efficient, and reliable methods. This paper explores how well-known data quality characteristics can be applied in a metrologically sound manner, enabling quality assessments even when reference data or traditional calibration data are unavailable.**

Keywords: Data quality, sensor network, metrology

I. INTRODUCTION

Sensor networks are a valuable asset in many settings, such as industrial and smart city applications, as they enable a data-driven approach to decision making by collecting real-time data from various entities. Sensor networks are becoming increasingly practical to deploy, in part because of the availability of low-cost sensors and open-source tools. However, low-cost sensors often have lower measurement quality. Combining this with sensor networks becoming ever-larger and the vast amounts of data they can collect, makes it particularly challenging to ensure trustworthy and reliable data. [1]

Established methods for sensor calibration are used to ensure traceability of measurements and estimate measurement uncertainty. However, these methods involve removing a sensor from its deployment location, transporting it to a laboratory for calibration and redeploying it afterwards. For many sensor networks, this is infeasible due to several factors, such as physical inaccessibility and cost constraints, further adding to the difficulty of assessing data quality. In itself, the huge amounts of data make the assessment of data quality by conventional means impractical without the availability of automated, efficient, and reliable methods, due to the significant amounts of resources required [1]. In cases where such methods are not established, discovering issues relies on the manual inspection of data, which results in substantial risks of basing conclusions and decisions on poor quality data. Thus, from a metrological perspective, new methods must be developed to measure data quality in sensor networks.

Real-world sensor networks vary greatly in many aspects, including associated legal and financial aspects [2, 3]. Some sensor networks use highly specialized and costly sensors for normative applications (e.g. network of gas flow meters in gas distribution pipelines), which will need to be calibrated regularly to avoid high financial and legal implications in case of incorrect measurements. Other sensor networks are less specialized, possibly using battery-powered sensors, and have lower associated costs (e.g. for maintenance, sensor replacement, etc.). These kinds of sensor networks are becoming quite common, for example, in smart cities and citizen science applications (e.g. for monitoring of local air quality and comparison with reference air quality networks).

This paper focuses on less specialized sensor networks, with limited legal or financial implications where data quality is somewhat unknown due to the mentioned challenges and where a strict metrological procedure is not mandatory.

The paper is structured as follows. In Section II. alternative calibration methods for sensor networks are introduced, highlighting methods that do not require additional reference sensors. Section III. highlights metadata as a tool to address different data quality characteristics. Section IV. describes different data quality characteristics relevant to sensor data quality from a metrological perspective and discusses how they can be applied in the described context. Lastly, Section V. concludes the paper.

II. CALIBRATION METHODS

There are a few different approaches to performing calibration in sensor networks in scenarios where no ground-truth reference sensor is available. Such calibrations are likely to be performed as online calibrations, which refer to calibrations performed during normal operation of sensors [4], i.e. without removing sensors or turning off (parts of) the sensor network. In-situ calibration is used in the case of online calibration, as it refers to the calibration of a sensor performed at the location of its deployment without having to disassemble and transport it to a calibration laboratory [5]. It is possible to perform both with and without a reference sensor. Co-calibration can be considered as a special case of in-situ calibration where nearby sensors already present in the network are used as reference devices [6]. In the case of co-calibration, the reference value itself may have to be estimated at the position of the device under test using appropriate interpolation and sensor fusion techniques. This is also related to blind calibration, which refers to calibration without controlled stimuli or ground-truth data [4].

From a sensor perspective self-calibration refers to the sensor being calibrated against its own measurements based on physical self-evident principles or similar. From a sensor network perspective it is similar to co-calibration where calibration is performed using other sensors in the network. [4]

Pre-calibration referring to calibration performed before deployment is also possible for sensors in sensor networks both in laboratory and in-situ. With this approach, the challenge lies in performing recalibration of sensors at a later stage and in potential differences between controlled and uncontrolled environments. If sensors are calibrated using pre-calibration, they can be used to calibrate other sensors through distributed calibration after deployment. [4]

Assuming pre-calibration is performed, another cost-reducing form of calibration is global calibration, where a calibration model is derived using the data from a subset of sensors [7]. The resulting model can then be transferred to similar sensors operating under comparable conditions, thereby reducing the complexity of future recalibrations.

Lastly, a sensor might have been pre-calibrated from the factory, or there might be a specification of its uncertainty in its data sheet. If no other method is possible, this data should be provided as metadata for the sensor in the sensor network system.

III. METADATA

Metadata, i.e. structured data that provides information about another piece of data, can vastly improve the trustworthiness, reliability, and (re)usability of sensor network data. In the case of sensor networks, especially sensor networks without access to reference sensor data, metadata can provide useful information about data structure, measurement units, measurement variables, placement of sensors, measurement method, uncertainty, etc. Metadata should be stored and accessible in the system. Having no metadata can complicate the analysis and meaningful use of the generated measurement data, as it will be ambiguous and difficult to use. Improving the metadata can not only help understanding and use of data, it can also help improve the quality, addressing the below-mentioned data quality characteristics. Examples of how are given in the respective sections for the data quality characteristics.

IV. DATA QUALITY CHARACTERISTICS

If none of the above calibration methods are possible or feasible for a given sensor network it is even more challenging to describe the uncertainty and traceability of measurements and to measure the data quality in general.

There exist several well-known data quality characteristics and these are being used within different domains to measure and describe data quality. The number and definitions of data quality characteristics vary from source to source. Not all are relevant from a metrological perspective and depending on the use case data quality requirements for the different characteristics can vary.

Table 1 lists data-quality characteristics drawn from two standards: ISO/IEC 5259-2 [8], which is based on ISO/IEC 25012 [9], and ISO 8000-210 [10], which is part of the ISO 8000 series on data management. Table 1 also lists the data quality characteristics selected for this paper due to their relevance from a metrological perspective.

ISO/IEC 5259-2 includes the data quality characteristics from ISO/IEC 25012 and uses the same categorization: inherent data quality and system-dependent data quality. Compared to ISO/IEC 25012, ISO/IEC 5259-2 extends with an additional category called "additional data quality characteristics". Only the inherent data quality characteristics of ISO/IEC 5259-2 are listed in Table 1. This is because the inherent characteristics are relevant for the data itself and not just for the system around it.

ISO 8000-210 focuses on data quality characteristics relevant to sensor data and describes only five characteristics.

A subset of characteristics (accuracy, completeness, consistency, precision, timeliness, traceability) was selected as most relevant from a metrological perspective and proposed as a basis for describing data quality in a metrologically sound manner. The selected characteristics are shown in Table 1. All have relatively consistent definitions,

Table 1. Data quality characteristics from two standards and selected characteristics for this paper.

ISO/IEC 5259-2	ISO 8000-210	Selected
Accuracy	Accuracy	Accuracy
Completeness	Completeness	Completeness
Consistency	Consistency	Consistency
Credibility	Precision	Precision
Currentness	Timeliness	Timeliness
Accessibility		Traceability
Compliance		
Confidentiality		
Efficiency		
Precision		
Traceability		
Understandability		

and four of them (accuracy, completeness, consistency, timeliness) are among those mentioned most frequently in literature with traceability and precision not mentioned as often [11], but they are essential to metrology.

It is necessary to make a note on the difference between timeliness and currentness. Timeliness was selected as a characteristic instead of currentness to keep more aligned with ISO 8000-210 because of its focus on sensor data and because of the umbrella aspect of timeliness, which also includes currentness. Timeliness is described in ISO/IEC 5259-2, but not included under it in Table 1, as in that standard timeliness is defined as the latency between when data is recorded and when data is available for use. This definition is in line with DAMA DMBOK [12], which gives an overview of different characteristics from different sources. It describes timeliness as "expectation of availability" and currentness as "current with the real world", but it also describes timeliness as an umbrella term of different aspects including currency and latency. It is the currentness part which is most interesting from a metrological point of view, i.e. not the latency but with emphasis on having the correct timestamp of the measurement.

ISO 8000-210 [10] includes another aspect which is interesting from a metrology point of view, where the data quality characteristics are categorized using the three categories for data quality described in ISO 8000-8 [13]: Syntactic, Semantic, and Pragmatic. Even if metrology focuses mainly on the semantic quality part, the syntactic category, encompassing the representational precision characteristic (i.e. representation of the measurement values with the correct amount of digits matching the measurement precision) is also important. The pragmatic category, which comprises timeliness, is likewise crucial for data quality because accurate timestamps strongly affect subsequent analyses. This is shown in Table 2.

Table 2. ISO 8000-210 data quality characteristics categorized according to ISO 8000-8

Category	Characteristic
Syntactic	Representational precision
Semantic	Accuracy
	Completeness
	Measurement precision
	Consistency
Pragmatic	Timeliness

A. *Data Quality Measurement*

When it comes to describing the quality of a dataset the data quality characteristics are usually applied using different data quality rules with associated metrics and measures for each characteristic.

In this paper a measure is "measuring" how many data points adhere to a given criterion, a metric is a measurement function calculating the result for a rule, and the rule itself is a threshold defining when the result is acceptable or not. The metrics are usually of the following form:

$$Metric = \frac{A}{B} \cdot 100\% \quad (1)$$

Where A is a measure, usually a count of how many data points adhere to a given criterion, and B is the total amount of data points considered for the given rule. The fraction can be multiplied by 100 % to express the result as a percentage.

In this way a rule could, for example, state that the result for completeness should be 95 % or above, i.e. the rule defines a target. This would be measured by having A be the counted number of data points and B be the expected number of data points. If the result is below the target, the data quality for that characteristic is not acceptable.

B. *Accuracy*

Accuracy is one of the most important data quality characteristics from a metrological perspective. Accuracy expresses how closely the data reflect the real world. Traditionally, accuracy is assessed with a reference sensor, which has been calibrated using reference standards that are traceable to the International System of Units (SI), that quantifies the calibrated sensor's uncertainty. This method also ensures the traceability of the sensor. Without a reference sensor, other methods must be used.

One method is to make multiple measurements over a short period of time. The difference between the measurements, which is a measure of repeatability, can be used as a lower limit for the accuracy of the sensor.

Monitoring the rate of change is another simple technique. Looking at the change in measurement value between two consecutive measurements can help indicate issues or accuracy of a sensor, e.g. if it is too large.

Another alternative method is to use sensor fusion, which can be used in two different ways. One is to have multiple sensors measuring the same variable. In that case, variations in the measurements between the different sensors can be used to indicate accuracy. Another way is to use multiple sensor signals to infer another measurement result and comparing the inferred measurement with an actual measurement of a given variable. This is a common practice in air quality applications. For example, if we have a single device including a temperature, humidity, CO_2 and pressure sensor, we could infer the data for the humidity sensor based on the temperature and pressure sensor outputs. Differences between the inferred value and the actual measurement may indicate accuracy and sensor issues.

The correlation between sensors can also be used to estimate accuracy. E.g. if sensors in a network are measuring the same variable, but at different locations, the correlation between sensors can be utilized to indicate when sensors deviate from their normal operation or when they drift.

All methods mentioned above can provide further insight into the measurement accuracy of sensors. Although they do not address measurement uncertainty in a metrologically traceable manner, they can still be used to assess the trustworthiness of sensor network data from a metrological perspective.

Metadata regarding accuracy can be provided by including sensor specifications such as the uncertainty or measurement range from the data sheet of a sensor. Furthermore, a description of measurement principle and data processing methods can provide insight into potential quality issues.

C. *Traceability*

Metrological traceability [14] of measurements is essential from a metrological standpoint and impossible to provide without reference sensors. The calibration methods mentioned above can provide some form of traceability, in having a well-defined approach for estimating a substitute for the full measurement uncertainty of the sensors. But they are only "traced" back to other sensors in the network, or models of the network, and does not provide the same traceability as the traditional approach.

In cases without reference sensors, metadata can mitigate the lack of traceability by providing sensor-specific information about sensor type, uncertainty, measurement principle, measurement range, etc.

Another aspect of traceability is data provenance, which is concerned with providing metadata about the origin and modifications of data as it moves through different processes and transformations. This can be important when correcting data issues, such as outliers and imputation of data where data are otherwise missing, because changes to the data are "traced" as they occur.

D. *Precision*

Precision is split in two as shown in Table 2 above; measurement precision, which is similar to the definition in VIM [15], and representational precision, which refers to the number of decimals or significant digits.

It is important to be aware of both. Measurement precision is directly related to the precision of the sensor and, therefore, the reliability of the data. Representational precision is important because different representations might have different functionalities or applications associated with them. Rounding errors or loss of information can occur if values are rounded in different parts of a system. Furthermore, floating point numbers in computers have limited precision which in some cases can lead to errors.

Metadata defining the number of decimals for sensor measurements and for stored measurement data will directly address the precision characteristic of data quality.

E. *Timeliness*

Timeliness is a measure of the availability of up-to-date data for a given application and depends on the correct timestamping of data to make sure it is known when a measurement was made. Incorrect timestamps render the data non-current; this, in turn, reduces usability because the true measurement time is unknown, especially where analyses are based on measurement times, such as time of day.

It is also useful to store metadata on the frequency of sensor measurements and the frequency of sending data to the cloud. This will make it easier to determine if data is being collected at the correct frequency.

F. *Completeness*

Completeness has a high impact on quality since missing data severely reduces the usefulness of data. As an example, sensor networks produce time series data where development of forecasting models will be greatly impacted by large gaps in the data. Missing data imputation is a common practice where gaps are filled using different methods. These can range from simple methods such as last observation carried forward (LOCF) , mean imputation, or linear interpolation, to more sophisticated methods, for example using machine learning [1, 16]. Poor completeness severely affects reliability and applicability as data can either be deemed unacceptable or artificial data is needed to fill in the gaps. In the latter case, data provenance is important to keep track of the inserted data.

Metadata can help improve completeness by providing information about placement of sensors, applied communication technologies, etc. which can be used to investigate the source of potential issues.

G. Consistency

DAMA DMBoK [12] defines consistency as the absence of difference, when comparing two or more representations of a thing against a definition. This definition fits well with a definition from a metrological perspective, where consistency can relate to sensor measurements being consistent over time, i.e. under the same conditions the sensor will output the same measurement result. Another aspect is sensors being consistent with each other, i.e. sensors are consistent with each other if they, under the same conditions, produce the same outputs.

When considering sensor data, one important challenge is the mitigation of sensor drift, where a sensor over time will produce different measurement results under the same conditions.

Another challenge related to consistency is detection of a stuck sensor, which repeatedly outputs an identical value. Depending on the number of significant digits of the sensor output this might be the result of a sensor issue.

Metadata can address some issues regarding consistency. It is preferable for later data analysis that format, unit, measurement method, etc. are consistent throughout a sensor network. This improves understandability and (re)usability of data. Consistent use of units also limits the number of conversions, each of which can impact the accuracy and precision of measurements.

V. CONCLUSION

For cases where a traditional metrological approach to sensor calibration cannot be used, an alternative approach to data quality within metrology has been presented. A selection of data quality characteristics was selected from literature, representing the primary focus areas of metrology. It is proposed to actively use these to address data quality challenges in a metrologically sound manner, because while different alternative calibration methods exist, if no reference sensor is available, there is no traceability, and data quality needs to be addressed through other means.

Metadata plays a big role in addressing data quality issues in sensor networks from a metrological point of view, as this can help inform how the system works and where to improve it.

Guidelines could be made to describe what information needs to be available in a system to make quality assessment manageable from a metrological perspective. Guidelines could define minimum requirements for metadata, alternative calibration methods where applicable, or other processes to make sure the system and its application is metrologically acceptable and fit-for-purpose in view of the goal pursued by the sensor network.

VI. ACKNOWLEDGMENTS

The project (22DIT02 FunSNM) has received funding from the European Partnership on Metrology, co-financed from the European Union's Horizon Europe Research and Innovation Programme and by the Participating States.

REFERENCES

[1] H. Teh et al. "Sensor data quality: a systematic review". In: *Journal of Big Data* 7 (Feb. 2020). DOI: 10.1186/s40537-020-0285-1.

[2] S. Tabandeh et al. "Sensor network metrology: Current state and future directions". In: *Measurement: Sensors* (2025), p. 101798. ISSN: 2665-9174. DOI: 10.1016/j.measen.2024.101798.

[3] J. Pearce et al. "End-user case studies in sensor network metrology". In: *Measurement: Sensors* (2025), p. 101819. ISSN: 2665-9174. DOI: 10.1016/j.measen.2025.101819.

[4] J. M. Barcelo-Ordinas et al. "Self-calibration methods for uncontrolled environments in sensor networks: A reference survey". In: *Ad Hoc Networks* 88 (May 2019), 142–159. ISSN: 1570-8705. DOI: 10.1016/j.adhoc.2019.01.008.

[5] F. Delaine et al. "In Situ Calibration Algorithms for Environmental Sensor Networks: A Review". In: *IEEE Sensors Journal* 19.15 (2019), pp. 5968–5978. DOI: 10.1109/JSEN.2019.2910317.

[6] M. Gruber et al. "A3.2 - Co-Calibration in Distributed Homogeneous Sensor Networks". In: *SMSI 2023*. Jan. 2023, pp. 47–48. DOI: 10.5162/SMSI2023/A3.2.

[7] S. De Vito et al. "A global multiunit calibration as a method for large-scale IoT particulate matter monitoring systems deployments". In: *IEEE Transactions on Instrumentation and Measurement* 73 (2023), pp. 1–16. DOI: 10.1109/TIM.2023.3331428.

[8] ISO/IEC. *Artificial intelligence – Data quality for analytics and machine learning (ML) – Part 2: Data quality measures*. Standard 5259-2:2024. Geneva, Switzerland, 2024.

[9] ISO/IEC. *Software engineering – Software product Quality Requirements and Evaluation (SQuaRE) – Data quality model*. Standard 25012:2008. Geneva, Switzerland, 2008.

[10] ISO. *Data quality – Part 210: Sensor data: Data quality characteristics*. Standard 8000-210:2024. Geneva, Switzerland, 2024.

[11] L. Zhang et al. "Data Quality Management in the Internet of Things". In: *Sensors* 21.17 (2021). ISSN: 1424-8220. DOI: 10.3390/s21175834.

[12] D. International. *DAMA-DMBOK: Data Management Body of Knowledge (2nd Edition)*. Denville, NJ, USA: Technics Publications, LLC, 2017. ISBN: 1634622340.

[13] ISO. *Data quality – Part 8: Information and data quality: Concepts and measuring*. Standard 8000-8:2015. Geneva, Switzerland, 2015.

[14] P. J. Brewer et al. "SI traceability and scales for underpinning atmospheric monitoring of greenhouse gases". In: *Metrologia* 55.5 (2018), S174. DOI: 10.1088/1681-7575/aad830.

[15] BIPM et al. *International Vocabulary of Metrology — Basic and general concepts and associated terms (VIM)*. Joint Committee for Guides in Metrology, JCGM 200:2012. (3rd edition). DOI: 10.59161/JCGM200-2012.

[16] T. Emmanuel et al. "A survey on missing data in machine learning". In: *Journal of Big Data* 8 (Oct. 2021). DOI: 10.1186/s40537-021-00516-9.