International Journal on Advanced Science Engineering Information Technology

Vol.14 (2024) No. 5
ISSN: 2088-5334

Multidimensional Indicator for Data Quality Assessment in Wireless Sensor Networks: Challenges and Opportunities

Nurul Aqilah Zamri [a], M. Izham Jaya [a,*], Siti Salwani Yaakob [a], Hidra Amnur [b], Shahreen Kasim [c]

[a] *Faculty of Computing, Universiti Malaysia Pahang Al-Sultan Abdullah, Pekan, Malaysia*
[b] *Department of Information Technology, Politeknik Negeri Padang, Padang, Indonesia*
[c] *Soft Computing and Data Mining Centre (SMC), Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Parit Raja, Johor, Malaysia*

*Corresponding author: *izhamjaya@umpsa.edu.my*

Abstract*—Wireless Sensor Networks (WSN) are equipped with numerous sensors that generate vast quantities of data, essential for operational efficiency and informed decision-making. However, the value of this data is contingent upon its suitability for the specific applications it serves. A significant challenge in WSNs is the selection of appropriate data quality dimensions and metrics necessary to construct robust Data Quality Indicators (DQI) and comprehensively assess data quality in various contexts. This systematic literature review seeks to identify the key data quality dimensions and the corresponding measurement metrics within WSNs, while exploring the use of multi-dimensional data quality criteria in developing DQI. A thorough search of SCOPUS and Web of Science databases yielded 475 potential research articles, from which 64 primary studies were selected for in-depth analysis. The findings highlight four key data quality dimensions in WSN: accuracy, timeliness, completeness, and consistency. However, choosing measurement metrics for each dimension requires an in-depth understanding of the data's context. Various approaches for obtaining DQI in WSN research were identified, including weighted linear average models and application-specific contextual information. Effective DQI incorporates weights to each dimension, reflecting the priorities of specific data users, and leverages contextual information pertinent to the sensors' data. It is crucial to evaluate whether the data collected by WSNs meets established quality standards, a key aspect of WSN operation. These insights will aid in developing more robust and reliable WSNs, ensuring the provision of high-quality data essential for effective operation and decision-making.

***Keywords*— Data quality; wireless sensor networks; WSN; data quality assessment; multidimensional.**

Manuscript received 5 Dec. 2023; revised 26 Jun. 2024; accepted 4 Aug. 2024. Date of publication 31 Oct. 2024.

I. INTRODUCTION

The global adoption of Wireless Sensor Networks (WSN) is transforming various sectors, including healthcare, industry, agriculture, home automation, and transportation [1]. WSNs, embedded with numerous sensors, generate vast amounts of data crucial for operations and decision-making. However, the utility of this data in supporting informed decisions for individuals and organizations relies on its fitness for the intended applications. Consequently, assessing whether the collected sensor data meets established data quality standards is imperative.

Evaluating data quality in WSN demands a thorough assessment encompassing all key dimensions relevant to the problem context, including data completeness, accuracy, timeliness, and consistency. From the assessment, data quality indicators (DQI) are derived to determine whether the sensor data is fit for its intended purpose. A significant challenge arises in selecting the appropriate metrics to construct the indicators and fully assess data quality within the context. Often, critical dimensions of data quality may be overlooked during the process, potentially compromising the robustness of the overall evaluation.

DQIs limited to completeness and accuracy must be revised to capture the multidimensional nature of sensor data in WSN [2]. Data quality dimensions, including precision, timeliness, and duplication, are crucial for establishing effective DQI, given the inherent complexity and dynamic nature of WSN data [3]. The indicators proposed in [4] adhere to the International Organization for Standardization (ISO) standards and incorporate essential data quality management procedures specific to sensor data. However, while ISO 8000-61 provides a foundational framework for data quality management, it may need to fully address the challenges posed by WSN data, such as real-time data variability and sensor reliability. This underscores the need for more tailored data quality assessments that comprehensively address the specific issues and requirements of WSN.

In developing DQI for WSN, previous research has identified data quality dimensions and corresponding assessment metrics through a comprehensive literature review, thereby establishing the theoretical frameworks and methodological foundations for DQI [2], [5]. Another approach centers on the perspectives of data users, utilizing surveys, interviews, and questionnaires to gather insights [1], [3]. Each method sparks a critical discourse on the optimal strategy for developing DQI for data generated by WSN. The nuanced consideration of these divergent methodologies prompts an exploration into which approach may prove more effective in advancing the understanding and assessment of data quality within the complex and dynamic context of WSN environments.

Wang and Strong [6] laid the foundation for data quality research by presenting a comprehensive list of twenty dimensions closely aligned with the concept of data quality. These dimensions have become a critical reference point for subsequent studies in the field. Building on this work, Wang [7] developed a methodology aimed at improving data quality, emphasizing the importance of these dimensions in ensuring data reliability and usefulness. In the context of WSN, data quality dimensions such as accuracy, completeness, consistency, and currentness have been utilized in data quality assessments [8], [9]. However, these studies overlook other data quality dimensions relevant to WSN. Notably, both studies rely exclusively on literature reviews rather than adopting a more comprehensive approach that solicits expert opinions to delineate the dimensions under scrutiny.

This systematic literature review aims to discern the key data quality dimensions and the corresponding measurement metrics in the context of WSN, as delineated in the recent literature. Additionally, this paper aims to identify the utilization of multi-dimensional data quality criteria in developing a robust DQI, thereby facilitating data quality assessment in WSN. The remainder of the paper is organized as follows: Section 2 details the methodology employed in conducting the systematic literature review. Section 3 presents the results and specifically addresses the research questions. Section 4 presents the key findings from the analysis, accompanied by a thoughtful discussion of potential future research directions. Finally, section 5 encapsulates the study's core insights and contributions, providing a comprehensive conclusion for the reader.

II. Material and Method

The systematic literature review framework provided by Kitchenham et al. [10] serves as a foundational approach, guiding this paper through the systematic processes of identifying, selecting, and critically evaluating relevant research articles, thereby contributing to the reliability of the findings. Aligned with the core objective of this paper, the initial phase focuses on identifying key data quality dimensions and measurement metrics necessary for developing a robust DQI for WSN. To achieve this objective, a comprehensive investigation is undertaken rigorously following accepted protocols for systematic literature reviews. The research methodology encompasses eight distinct phases: (1) formulation of research questions, (2) development of a search strategy with targeted keyword identification, (3) establishment of inclusion and exclusion criteria, (4) definition of quality assessment criteria, (5) selection of primary studies, (6) meticulous data extraction, (7) thorough data synthesis, and (8) rigorous quality assessment. Figure 1 visually depicts the research process, and each step is subsequently elaborated upon in further detail.

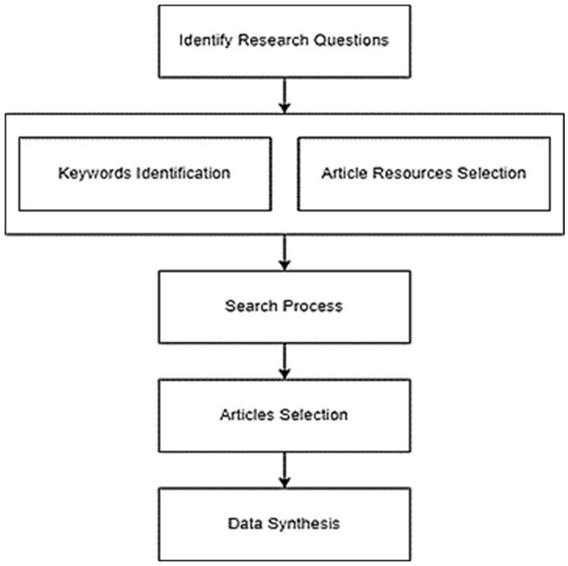

Fig.1 Systematic literature review process.

A. Research Questions

The development of a robust DQI for WSN necessitates the identification of crucial quality dimensions. Therefore, this paper proposes the following research questions to guide this endeavor:

a. RQ1: What key dimensions are essential for comprehensively assessing data quality within WSN?
b. RQ2: Which measurement metrics effectively assess key data quality dimensions in WSN?
c. RQ3: How can key data quality dimensions be effectively integrated into a robust DQI framework for WSN?

RQ1 seeks to identify the key dimensions critical to assessing data quality in WSN. This ensures that the chosen dimensions are relevant and capture the unique characteristics of WSN data. Building upon this foundation, RQ2 focuses on developing metrics to quantify these key dimensions, effectively verifying their impact on data quality. Finally, RQ3 addresses the formulation of a robust DQI that integrates these key data quality dimensions.

B. Search Strategy

This paper employed a search strategy based on the chosen keywords and deliberately selected databases to ensure comprehensive and relevant search results. Research articles were acquired through an automated search function across all database resources using the selected keywords. Keywords were chosen based on three main categories: data quality, WSN, and the Internet of Things (IoT). Each research article's title and abstract were searched for the chosen keywords. By including related keywords in the search process, this paper aimed to include all relevant research articles that discussed data quality within the context of WSN. Additionally, given the diverse applications of WSNs, this paper identified specific keywords related to areas of IoT, ensuring that the review includes all research relevant to the research objectives. Table 1 presents the specific keywords used for each category.

TABLE I
KEYWORDS BY CATEGORIES

Category	Keywords
Data Quality	'data quality', 'dq', 'quality of data', 'data quality indicators', 'dimensions'
IoT	'internet of things', 'IoT', 'sensors'
WSN	'sensor networks', 'WSN', 'WSN', 'wireless sensor network'

Distinct research databases were selected in this paper, as outlined in Table 2. The selection of databases from this list aimed at minimizing redundancy in the gathered data. Furthermore, the rationale for opting for these databases rested on their incorporation of an advanced search feature, enabling the utilization of logical operators. To enhance the temporal relevance of the retrieved literature, the search strategy was designed to include a publication date filter, thereby focusing on articles published within a specific period.

TABLE III
SELECTED RESEARCH DATABASES

Database Name	URL
Scopus	https://www.scopus.com
Web of Science	https://www.webofscience.com

C. Inclusion and Exclusion Criteria

To capture recent advancements and emerging trends in data quality research within WSN, the search targeted research publications from 2013 until 2023. A two-step selection process was implemented to refine the search results. The initial phase involved title and abstract screening of retrieved articles to identify potentially relevant studies. Duplicate entries were systematically excluded to ensure a non-redundant selection. Subsequently, a thorough full-text review of shortlisted articles was conducted to verify their alignment with pre-defined study objectives. The primary study selection was further refined by implementing additional inclusion and exclusion criteria as outlined below.

1) Inclusion Criteria:

a. Includes research articles that were published between January 2013 and December 2023.
b. Research articles that investigate the multifaceted nature of data quality within the domain of WSN.
c. Research articles that primarily investigate data quality dimensions specific to WSNs.
d. Research articles published in English for broader reach and dissemination of findings.
e. Research articles published in an indexed journal or proceedings to ensure research quality and validity.

2) Exclusion criteria:

a. Exclude editorial articles and other non-research publications.
b. Exclude studies that do not address any dimension of data quality.
c. Exclude studies with a primary focus unrelated to data quality dimensions in WSN.
d. Exclude studies that do not consider data quality at all.

The initial screening of retrieved research articles comprised two sequential iterations. The initial stage of the article search employed a two-step screening process based on pre-defined inclusion criteria. This process involved scanning both titles and abstracts of identified studies to assess their potential relevance to the research question. Subsequently, during the second iteration, shortlisted articles underwent a thorough full-text review to ensure alignment with the research objectives. Exclusion criteria were rigorously applied at this stage to eliminate non-pertinent studies. Table 3 details the search queries employed within each database, utilizing relevant search terms and logical operators through complex search functionalities.

TABLE IIIII
QUERIES USED IN THE SEARCH PROCESS

Database Name	Search Query
Scopus	("data quality" OR DQ OR "quality of data") AND (assessment OR evaluation OR measurement) AND ("internet of things" OR IoT OR "WSN" OR WSN OR "sensor networks")
Web of Science	("data quality" OR DQ OR "quality of data") AND (assessment OR evaluation OR measurement) AND ("internet of things" OR IoT OR "WSN" OR WSN OR "sensor networks")

D. Quality Criteria

This section aims to ascertain the suitability of primary research articles for addressing research inquiries by evaluating their information adequacy. Quality assessment criteria (QAC), as listed below, were used to evaluate the quality of the primary articles.

a. QAC.1: Does the article thoroughly define the key data quality dimensions particular to WSN?
b. QAC.2: Does the article propose or discuss a specific metric to measure the key data quality dimension in WSN?
c. QAC.3: Does the paper present a comparative evaluation of the proposed DQI?

E. Collection of Primary Articles

The initial search using the abovementioned queries identified 475 potential research articles. Subsequent titles and abstract screening eliminated irrelevant articles, reducing the pool to 199. The application of predetermined inclusion and exclusion criteria led to the exclusion of an additional 135 articles. Sixty-four primary articles were deemed sufficient information to address the research inquiries effectively.

F. Data Extraction

Each primary article was analyzed, and relevant data was extracted based on the categories and subcategories outlined in Table 4. The subcategories are then mapped to the specific research questions posed in this paper.

TABLE IVV
DATA EXTRACTION FORM

Category	Subcategory	Research Question (RQ)
Dimensions	Key dimensions in WSN	RQ1
Metrics	Metrics that measure key data quality dimensions in WSN	RQ2

Data extraction from the articles employed automated and manual procedures to ensure accuracy and completeness. The dimensions category played a pivotal role in identifying key dimensions crucial for a comprehensive assessment of data quality in WSN, thereby contributing to the completion of RQ1. Next, to address RQ2, a detailed list of metrics for measuring these key dimensions was compiled. This analysis aimed to identify the most prevalent metrics for evaluating data quality dimensions within the WSN domain. Further, RQ3 focused on formulating a DQI, integrating all previously identified dimensions.

This extracted data is deemed highly valuable for several vital reasons. Firstly, it enables researchers to identify gaps and areas requiring further investigation. Secondly, it unveils potential weaknesses that may be addressed through alternative approaches. Finally, it provides valuable insights and possible solutions to explore in future research endeavors.

G. Data Synthesis

The data synthesis procedure facilitated data integration tailored to the specific requirements of each RQ. This paper utilized a narrative synthesis approach to systematically integrate data from various heterogeneous resources to address RQ1, RQ2, and RQ3.

H. Quality Assessment

Beyond applying predefined inclusion and exclusion criteria, each initial research underwent a rigorous quality assessment using established QAC questions to minimize preconceived notions and ensure a robust literature corpus. This systematic evaluation involved assigning numerical scores (0-1) to each primary research article, with one indicating a complete and satisfactory response to the QAC question, 0.5 signifying a partially addressed response, and 0 representing no engagement with the question. Overall study quality was then determined by summing individual QAC scores. As detailed in Table 5, all selected primary research articles achieved a collective score exceeding 50% for each QAC, demonstrating their substantial content about data quality, particularly crucial data quality dimensions within the WSNs context.

TABLE V
THE COLLECTIVE SCORE FOR QAC

QAC Criteria	Responding Score	Total Collective score
QAC 01	{0,0.5,1} (No, partially, yes)	1
QAC 02	{0,0.5,1} (No, partially, yes)	0.5
QAC 03	{0,0.5,1} (No, partially, yes)	0.5

III. RESULTS AND DISCUSSIONS

Guided by the procedure explicated in the preceding section, this section utilizes data extracted from the primary research articles to answer each of the established RQ, thereby presenting the findings of this systematic literature review.

A. RQ1. What are the Key Dimensions Essential for Comprehensively Assess Data Quality within WSN?

The analysis of data quality dimensions discussed in the primary research articles, as conferred in Fig. 2, has yielded a beneficial understanding of the key dimensions of data quality assessment within the WSN domain. Accuracy, timeliness, completeness, and consistency are identified as the most frequently emphasized data quality dimensions in the collection of primary articles. This finding underscores their pivotal role in guaranteeing the validity and comprehensiveness of collected sensor data.

Table 6 presents a detailed overview of the various data quality dimensions evaluated across the primary research articles, highlighting their domain, source of the problem, and related quality criteria within the context of WSN and IoT.

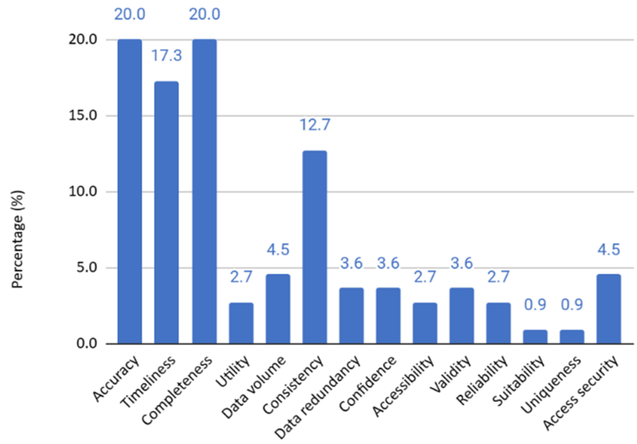

Fig.2 Percentage of data quality dimension mentioned in the primary research articles.

TABLE VI
ANALYSIS OF DATA QUALITY DIMENSIONS IN WSN

Data Quality Dimension	Dataset and Domain	Source of problem	References	Related quality criteria
Accuracy	IoT, WSN	Measurement errors dirty data, outliers, noise, data misinterpretation	[3], [4], [9], [11] - [26]	Uncertainty, bias, precision, validity, correctness
Timeliness	IoT, WSN	Not updated data, low data rate	[2], [3], [5], [9], [11] – [30]	Min. time coverage, presentiveness, currency, currentness, freshness, volatility, latency, delay, frequency
Completeness	IoT, WSN	Missing value	[1]-[5], [8], [11] - [13], [15], [16], [18] - [20], [22], [23], [25], [26], [28], [30], [31]	Low data capture, data availability, min. time coverage
Utility	IoT, WSN	Noise, data loss, missing data	[11], [5], [32]	Usage, frequency, relevancy
Data volume	IoT, WSN	Delayed data transmission, data loss, data frame distortion	[11], [12], [3], [5], [1]	Min. number of sampling points, representativeness

Data Quality Dimension	Dataset and Domain	Source of problem	References	Related quality criteria
Consistency	IoT, WSN	Irregular observations	[2]-[5], [8], [9], [11], [12], [16], [20], [23], [26], [30], [32]	Concordance, Comparability, Conformity
Data redundancy	IoT	Data duplication	[3], [15], [5], [1]	Duplicates
Confidence	IoT	Source reliability	[3], [13], [5], [1]	Uncertainty, interpretability
Accessibility	IoT	Limited node capability	[4], [5], [1]	Availability, Ease of access
Validity	IoT	Source malfunction and failure	[5], [22], [8], [29]	Plausibility, detection limit
Reliability	IoT	Nodes failure	[13], [14], [5]	Trust, source reputation
Suitability	IoT	Sensor's capabilities	[19]	-
Uniqueness	IoT	Repetitiveness	[20]	-
Access security	IoT WSN	Limited resources	[13], [20], [5], [8], [28]	artificiality

The following subsection discusses the top four most frequently addressed data quality dimensions identified in the primary research articles.

1) Accuracy:

Within the domain of WSN research, accuracy consistently emerged as the most critical data quality dimension investigated across the reviewed research articles [9], [3], [19], [2], [5], [21], [22], [23], [33], [34], [35]. Accuracy is the extent to which observed data accurately represents the true, real-world condition of the investigated event [28]. On the other hand, Hendrik et al. [33] introduce a nuanced distinction by employing the term "validity" to denote the faithfulness of data value, overlapping with the conceptual domain of accuracy. Accuracy is also defined as the proximity of a measured value to the true value [12], and it implies that repeated measurements should produce results that are both consistent and close to the true value [28]. Terms like precision [3], [5], [23], [34], [36], validity [23], [33], correctness [35], [37], [38] and uncertainty [3] are used interchangeably with accuracy.

Crucial considerations for ensuring data accuracy in WSN revolve around mitigating systematic and random errors. Systematic errors, documented in several articles [33], [34], [39], [40], often arise from improper sensor placement or selection [33], [40], [41] for example, the erroneous ambient temperature readings captured by a misplaced product sensor [33]. Additionally, sensor limitations in giving accurate values can introduce uncertainty, rendering dirty and unusable data for analysis [40].

Inaccurate data produces outliers, defined as data points deviating significantly from the norm [40], [42]. While sensor malfunctions can undoubtedly cause outliers [37], [40], [42], [43], [44], they can also be crucial indicators of real-world events [40], [42]. Thus, outlier analysis differentiates between erroneous data and valuable insights.

Data accuracy is impacted by noise, defined as unwanted signal distortions [38]. Defective sensors, often caused by factors like depleted batteries, faulty memory, or transmission errors [19], [42], [45], [46], constitute a significant source of noise. In WSN, sensor noise can be reduced by avoiding interference between wireless devices operating on the same frequency [47].

2) Timeliness:

Within the context of WSN, data timeliness embodies the concept encompassing the frequency and regularity with which observed sensor data aligns with critical decision-making moments and points of interest [34]. However, the notion extends beyond ordinary frequency, considering the freshness of the sensor data. Sensor data freshness is assessed by measuring the time gap between recorded data and the current system's arrival time. The metric reflects the degree to which an observation aligns with the target period of interest [12].

Data timeliness in WSN has been referred to by various terms, highlighting its multifaceted nature - currency [34], volatility [12], [34], latency [39], freshness [39], [48], data rate [48], [49], delay [44], [36], frequency [50], minimum time coverage [3], representativeness [3], and currentness [9]. Volatility characterizes the frequency of data fluctuations and is often measured by the time the data retains its accuracy and relevance [12].

Lacks in data timeliness primarily manifest in two forms: missing updates and insufficient data rates [48]. Data rates can significantly impact the overall timeliness of data acquisition and analysis. This challenge is particularly prevalent in resource-constrained settings, such as agricultural environments, where WSN devices operate with limited energy resources and often span vast communication distances. Consequently, low-power network technologies like LoRaWAN and SigFox have become attractive due to their energy efficiency. However, these technologies have inherent drawbacks – reduced data rates and increased latency [51]. While they enable energy-efficient data transmission, they compromise timeliness, potentially limiting the responsiveness and effectiveness of WSN applications in these scenarios. This presents a crucial trade-off between energy efficiency and timeliness, demanding careful consideration when designing and deploying WSNs for specific applications.

3) Completeness:

Data completeness in WSN represents the extent to which the WSN network delivers all expected data points without omissions or inconsistencies [12], [5], [34]. This fundamental concept is related to data accessibility and reaches into the absence of missing values, ensuring the integrity and reliability of collected sensor data for accurate decision-making and analysis.

Within the WSN domain, interpretations of the concept of completeness vary. While numerous articles directly employ the term "completeness" [2], [3], [5], [9], [22], [23], [34], [40], [48], [37], [52], others utilize related expressions like "data availability" [53], "missing data" [54] - [57], "minimum data capture" [3], and "minimum time coverage" [3] to emphasize specific aspects of completeness dimension.

Data completeness in WSNs is influenced by sensor inefficiencies, ranging from hardware malfunctions to limited sensing capabilities, which lead to data gaps, compromising the overall integrity of the collected data [34], [36], [54], [55]. Additionally, communication disruptions caused by network congestion, interference, or signal degradation can contribute to data loss, further hindering data completeness. Furthermore, malicious actors who intercept or manipulate sensor data can inflict severe damage, jeopardizing data integrity and data completeness in WSN [58].

Beyond technical challenges, data update frequency plays a crucial role in ensuring data completeness. Inadequate data updates can render outdated sensor data incomplete, as they fail to reflect current conditions or real-time changes. This highlights the importance of timely data collection and transmission to maintain a complete picture of the monitored environment.

Moreover, data owners might intentionally withhold specific information due to diverse constraints, such as privacy concerns or proprietary interests [33]. While such filtering may be understandable from particular perspectives, it results in less comprehensive sensor data and impacts the overall completeness of the available data for users. The relationship between technical challenges and human factors demands a holistic approach to address data completeness issues in WSNs, balancing technical solutions with ethical considerations and user needs.

4) Consistency:

Data consistency refers to the absence of conflicts within and between data sources in WSN, ensuring coherence and agreement across different contexts [2], [8], [11]. It encompasses two key aspects: internal consistency and external consistency. Internal consistency, where data elements within a single source are free from errors and inconsistencies, and external consistency, where data elements from different sources align and complement each other [8]. Additionally, consistency can be measured by the degree of adherence to standardized formats [20].

The term consistency is widely used in primary research articles [2], [3], [9], [23]. However, different terms such as concordance [3], [5], comparability [3], and conformity [20] are also used to represent data consistency. Data consistency in WSN is affected by irregularity between sensor readings from multiple sensors in the network. Specifically, sensors originating from various sources within a WSN may produce varying readings for the same observation, leading to discrepancies in the data [8], [59], [60].

Inconsistencies can stem from non-concurring observations across various data sources in WSN and faulty sensors incapable of generating consistent readings [8], [59]. In the context of traffic event analysis, Kuemper et al. [8] discovered discrepancies between vehicle speed measurements and vehicle counts collected from two independent origins. These discrepancies were attributed to differences in the timing and vantage point of the observations, highlighting the importance of concordance in ensuring data alignment and reducing ambiguity.

Significantly, data quality issues in WSNs often exhibit interdimensional overlap. Consider noise, typically associated with data accuracy concerns, also demonstrably impacting data utility. Noisy data can hinder the accessibility of desired information and limit the value derived by data consumers [32]. Similarly, data frame distortion presents challenges to data quality dimensions: accuracy and volume. As discussed in [61], distorted data deviates from its original form, compromising the fidelity of the derived results. Moreover, reconstructing distorted data can increase the volume of transmitted components, further burdening system resources. Likewise, data loss can adversely impact both data utility and data volume. When data loss occurs during transmission [54], [62] and data updates are not timely [62], [63], the resulting incomplete data in WSN raises concerns regarding data usage for user needs and the overall information quantity.

The findings in RQ 1 underscore a critical point: while data quality dimensions in WSN are conceptually distinct and often defined with specific purposes in mind, they are not entirely isolated. In practice, significant relationships and interdependencies exist between these dimensions. Addressing deficiencies in one dimension frequently requires consideration of its potential impact on others. Understanding these multidimensional relationships is crucial for implementing comprehensive data quality solutions that effectively safeguard the reliability and coherence of data within the complex and interconnected landscapes of WSN.

B. RQ2.Which Measurement Metrics Effectively Assess Key Data Quality Dimensions in WSN?

Various approaches have emerged to gauge the fitness of data for its intended purpose. DeepDQ is proposed in [20] to measure multiple data quality metrics. DeepDQ delivers detailed insights into data inconsistencies or shortcomings in the data quality metric. Mathematically defined metrics for data quality assessment are utilized by the works of [1], [3], [8], [9], [19]. These studies advocate for a more formal and universally applicable approach, grounding their methodologies in rigorous mathematical foundations. In the following subsections, this paper elucidates the mathematically defined metrics used to assess the key data quality dimensions defined in RQ 1.

1) Accuracy:

Data accuracy is mathematically specified within the range of 0 until 1 of the maximal absolute systematic error α as defined in Equation (1) [1], [3], [5]. The proximity of measured values to the true value is further clarified in [3], [5], where α denotes the difference between the measured value, $\check{v}$ and the accepted true values, v. The definition emphasizes the proximity of measured values to the true value. Implicitly, the smaller α value is, signifies higher data accuracy.

$$\text{Data Accuracy} = \max(0, 1 - \alpha) \quad (1)$$

The alternative mathematical definition of accuracy is presented in [19], where data measurements' precision is considered. As defined in Equation (2), data accuracy is quantified by the moving standard deviation concerning a given data value, ν compared to its mean of standard errors, $\overline{SE}$. The variation of individual data points around the $\overline{SE}$ reflecting the consistency of measurements.

$$A = (1 - \overline{SE}) \quad (2)$$

2) Timeliness:

The dimension of timeliness in WSN encompasses the terminology of volatility, currentness, and freshness [11], [3]. Timeliness is mathematically represented within the range of 0 until 1 in Equation (3) [3], [5], where $volatility$ is defined as the period during which data remains valid and $currency$ is the timestamp of the updated data.

$$Data\ timeliness = \max(0, currency/volatility) \quad (3)$$

Alternatively, [9] defined Currentness based on the previous timestamps of data updates, t_{i-1}, update intervals, and expected timestamps of the next update, t_{i+1}, as expressed in Equation (4). Equation (5) denotes the currentness of a data source by dividing the Currentness with the number of data updates from a similar data source. Thus, high $\text{Currentness}_{datasource}$ value indicates that the data source always updates current data.

$$Currentness = \begin{Bmatrix} 1 \\ 0 \end{Bmatrix} - \frac{|(t_{i-1}+1)-t_i|}{I} \Big\} \, if\, t_i \in \,]t_{i-1}, -t_{1+1}[\quad (4)$$

Otherwise,

$$Currentness_{datasource} = \frac{\sum_{i=1}^{n} Currentness_{event}}{n} \quad (5)$$

3) Completeness:

Data completeness in WSN is defined as the proportion of absence data points relative to the total number of data points available [1]. Similarly, [3] and [5] present a metric that compares the actual amount of data collected within a timeframe to the expected total, expressed as a ratio. This metric captures the overall data availability and highlights potential deficiencies in data acquisition as defined in Equation (6).

$$\text{Data completeness} = \frac{Number\ of\ Collected\ Values}{Number\ of\ Expected\ Values} \quad (6)$$

A sensor-specific metric for data completeness is defined in [19] where completeness is measured as the absence of missing data point within a given window, divided by the expected number of data points with the actual number collected as defined in Equation (7).

$$\text{Data completeness} = 1 - \frac{Expected\ data\ points - Collected\ data\ points}{Expected\ data\ points} \quad (7)$$

4) Consistency:

Data consistency in WSN can be assessed by comparing data values across multiple data sources using correlation analysis, considering the share of observed values from different sensors, weighted by the sensor proximity as defined in Equation (8) [8].

$$q_{con}(x_0) = \sum_{i=1}^{n} \lambda_i(x_0) \cdot c(x_0, x_i) \quad (8)$$

Expanding on this approach, the Pearson correlation coefficient value is subsequently employed to penalize correlations with distant data variables using a weighted function [3], [5]. Conversely, a data consistency metric, grounded in the agreement of data sources within specific tolerance thresholds, such as time and measurement, is proposed in [9]. Data consistency is quantified by the ratio of events that conform to the tolerance threshold to the total number of relevant events.

C. RQ3.How Can Key Data Quality Dimensions be Effectively Integrated into a Robust DQI Framework for WSN?

Cheng et al. [12] proposes a multidimensional model for DQI based on a weighted linear average model. The Pairwise Comparison Matrix (PCM) technique is employed to determine the weights, which are vital for assessing the relative significance of each dimension. This approach ensures that expert preferences and domain knowledge are factored into the final data quality indicator, providing a holistic and objective measure of data quality. The accumulated data quality score for the proposed model ranges from 0 to 1.

Alternatively, the priority-based distribution approach is employed to determine the significance of each data quality dimension in measuring DQI [3]. The proposed approach prioritizes data quality dimensions based on their correlation with the overall data quality rating by assigning higher weights to dimensions with stronger correlations. The study demonstrates that the priority-based distribution approach outperforms weighting methods, including balanced, correlation-based, and hybrid distributions, and techniques relying solely on sensor and modeling accuracy. Moreover, the priority-based distribution approach exhibits the strongest correlation with subjective data quality assessments, solidifying its effectiveness.

Weights are also utilized in [9], where each dimension receives a weight reflecting its relative importance, pre-defined by stakeholders. The minimum acceptable DQI is defined by data quality thresholds established by the same stakeholders. Dimension-specific measurement methods then calculate the individual quality indexes. For instance, accuracy is determined by comparing received measurements with established specifications, completeness by verifying the capture of all expected events within the update window, consistency by examining data coherence across sensors monitoring the same phenomenon, and currentness by evaluating event reception timeliness against the agreed update interval.

A two-pronged approach centered on accuracy and completeness is proposed in [19]. The data quality level is evaluated by comparing the expected number of data points and the corresponding value with the actual data record while considering the specific application context. Unlike other methods that depend on multiple sensors or external data sources, the proposed approach assesses sensor data quality based on internal contextual information. The resulting quality metric directly integrates into the application's decision-making and actuation processes.

D. Discussion

This section presents the findings and offers recommendations for future research directions concerning data quality in WSN.

A. Findings

The complex relationship between data quality dimensions, context, and metrics underscores the need for a flexible and domain-specific approach to DQI evaluation in WSN. Acknowledging these relationships enables data quality practitioners to determine key dimensions and metrics, thereby ensuring the reliability of sensor data for generating valuable insights within WSN.

The analysis of RQ1 reveals a spectrum of dimensions discussed in the previous research articles: accuracy, timeliness, completeness, data volume, consistency, data redundancy, uncertainty accessibility, validity, trust, suitability, uniqueness, and access security. Notably, accuracy, timeliness, completeness, and consistency are the most frequently mentioned and are considered the fundamental pillars of data quality within the WSN. These key data quality dimensions are the core for evaluating the "fitness-for-use" of sensor data in WSN applications.

However, a more complex checklist approach is needed when assessing the key data quality dimensions. In response to RQ2, selecting the appropriate metrics for each dimension necessitates a deep understanding of the data's context, as more than a one-size-fits-all approach is required. RQ2 emphasizes that diverse metrics might be suitable even within the same problem domain depending on the specific context and the data user's requirements. Furthermore, in RQ3, various methodologies exist for obtaining DQIs, including weighted linear average models [3], [64], weighted approaches [9], and application-specific contextual information [19], [65].

B. Future Research Directions

1) Standardizing Terminology for Data Quality in WSN:

A lack of standardized terminology and metrics for data quality assessment currently hinders WSN. The disparity, where previous research articles utilize diverse definitions and metrics based on individual WSN contexts, presents a significant obstacle to the comprehensive advancement and understanding of data quality as a distinct discipline in WSN. Addressing this gap holds immense potential to unlock further advancements in the field. Establishing a specialized foundation for data quality in WSN, characterized by standardized terminology and adapted metrics, can fortify the quality, reliability, and interpretability of sensor-generated data. This, in turn, will pave the way for more robust and impactful applications across diverse domains.

2) Exploring the Utility of DQI in WSN:

DQI is crucial for evaluating data quality across diverse data sources in WSN. The potential advantages of using a DQI for WSN remain unclear, warranting further investigation. While previous research underscores DQI's capability to facilitate cross-source assessment, a more thorough investigation is required to understand its impact on large-scale WSN deployments and its effectiveness in guiding stakeholder decision-making.

IV. CONCLUSION

Recognizing the critical role of data quality in WSNs, this paper undertakes a systematic literature review to illuminate its multidimensional implications in assessing data quality. The research is guided by three key questions: RQ1, RQ2, and RQ3. RQ1 identifies key dimensions of data quality in WSN: This paper analyzes existing research articles to uncover the most prevalent dimensions used to assess data quality in WSN. The findings reveal four key dimensions: accuracy, timeliness, completeness, and consistency. These dimensions form the cornerstone for evaluating the reliability and utility of sensor data across diverse WSN applications. RQ2 investigates the mathematical equations used to measure the key data quality dimensions in WSN. The analysis indicates that existing data quality metrics are primarily formulated based on definitions and critical aspects specific to the WSN landscape. This highlights the need for context-aware metrics that adapt to the unique requirements of various contexts. RQ3 addresses the establishment of a robust DQI in previous research. This paper examines current approaches to constructing DQIs that aggregate individual dimension measurements into a holistic data quality assessment. It is essential for DQIs to incorporate weights based on specific data users' priorities and to leverage contextual information of the sensor data.

Research should focus on standardizing terminology for defining and measuring data quality in WSNs. Additionally, it is crucial to validate proposed metrics through real-world deployments and explore how DQIs influence data-driven decision-making processes in WSN. These advancements will contribute to developing more robust and dependable WSNs, ensuring the delivery of high-quality data.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support from the International Matching Grant with Project ID UIC241510 from the Universiti Malaysia Pahang Al-Sultan Abdullah (RDU242708). This support is gratefully acknowledged.

REFERENCES

[1] A. Karkouch, H. Mousannif, H. Al Moatassime, and T. Noel, "Data quality in internet of things: A state-of-the-art survey," *Journal of Network and Computer Applications*, vol. 73, pp. 57–81, Sep. 2016, doi: 10.1016/j.jnca.2016.08.002.

[2] H. Y. Teh, A. W. Kempa-Liehr, and K. I.-K. Wang, "Sensor data quality: a systematic review," *Journal of Big Data*, vol. 7, no. 1, Feb. 2020, doi: 10.1186/s40537-020-0285-1.

[3] J. H. Buelvas, D. Múnera, and N. Gaviria, "DQ-MAN: A tool for multi-dimensional data quality analysis in IoT-based air quality monitoring systems," *Internet of Things*, vol. 22, p. 100769, Jul. 2023, doi: 10.1016/j.iot.2023.100769.

[4] R. Perez-Castillo et al., "DAQUA-MASS: An ISO 8000-61 Based Data Quality Management Methodology for Sensor Data," *Sensors*, vol. 18, no. 9, p. 3105, Sep. 2018, doi: 10.3390/s18093105.

[5] J. Buelvas, D. Múnera, D. P. Tobón V., J. Aguirre, and N. Gaviria, "Data Quality in IoT-Based Air Quality Monitoring Systems: a Systematic Mapping Study," *Water, Air, & Soil Pollution*, vol. 234, no. 4, Apr. 2023, doi: 10.1007/s11270-023-06127-9.

[6] R. Y. Wang and D. M. Strong, "Beyond Accuracy: What Data Quality Means to Data Consumers," *Journal of Management Information Systems*, vol. 12, no. 4, pp. 5–33, Mar. 1996, doi:10.1080/07421222.1996.11518099.

[7] R. Y. Wang, "A product perspective on total data quality management", *Commun. ACM*, vol. 41, no. 2, pp. 58-66, 1998.

[8] D. Kuemper, T. Iggena, R. Toenjes, and E. Pulvermueller, "Valid.IoT," *Proceedings of the 9th ACM Multimedia Systems Conference*, Jun. 2018, doi: 10.1145/3204949.3204972.

[9] M. Ahmed, C. Taconet, M. Ould, S. Chabridon, and A. Bouzeghoub, "IoT Data Qualification for a Logistic Chain Traceability Smart Contract," *Sensors*, vol. 21, no. 6, p. 2239, Mar. 2021, doi:10.3390/s21062239.

[10] B. Kitchenham, ''Procedures for performing systematic reviews,' Softw. Eng. Group; Nat. ICT Aust., Keele; Eversleigh, Keele Univ., Keele, U.K., Tech. Rep. TR/SE-0401; NICTA Tech. Rep. 0400011T.1, 2004.

[11] C. Liu, P. Nitschke, S. P. Williams, and D. Zowghi, "Data quality and the Internet of Things," *Computing*, vol. 102, no. 2, pp. 573–599, Jul. 2019, doi: 10.1007/s00607-019-00746-z.

[12] H. Cheng, D. Feng, X. Shi, and C. Chen, "Data quality analysis and cleaning strategy for wireless sensor networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, no. 1, Mar. 2018, doi: 10.1186/s13638-018-1069-6.

[13] S. Sicari, C. Cappiello, F. De Pellegrini, D. Miorandi, and A. Coen-Porisini, “A security-and quality-aware system architecture for Internet of Things,” *Information Systems Frontiers*, vol. 18, no. 4, pp. 665–677, Nov. 2014, doi: 10.1007/s10796-014-9538-x.

[14] J. Byabazaire, G. O’Hare, and D. Delaney, “Data Quality and Trust : A Perception from Shared Data in IoT,” *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, Jun. 2020, doi: 10.1109/iccworkshops49005.2020.9145071.

[15] J. Guo and F. Liu, “Automatic Data Quality Control of Observations in Wireless Sensor Network,” *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 4, pp. 716–720, Apr. 2015, doi:10.1109/lgrs.2014.2359685.

[16] L. Zhang, D. Jeong, and S. Lee, “Data Quality Management in the Internet of Things,” *Sensors*, vol. 21, no. 17, p. 5834, Aug. 2021, doi:10.3390/s21175834.

[17] G. D’Aniello, M. Gaeta, and T.-P. Hong, “Effective Quality-Aware Sensor Data Management,” *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 1, pp. 65–77, Feb. 2018, doi:10.1109/tetci.2017.2782800.

[18] J. Byabazaire, G. M. P. O’Hare, R. Collier, and D. Delaney, “IoT Data Quality Assessment Framework Using Adaptive Weighted Estimation Fusion,” *Sensors*, vol. 23, no. 13, p. 5993, Jun. 2023, doi:10.3390/s23135993.

[19] K. Fizza, P. P. Jayaraman, A. Banerjee, D. Georgakopoulos, and R. Ranjan, “Evaluating Sensor Data Quality in Internet of Things Smart Agriculture Applications,” *IEEE Micro*, vol. 42, no. 1, pp. 51–60, Jan. 2022, doi: 10.1109/mm.2021.3137401.

[20] G.-O. Meritxell, B. Sierra, and S. Ferreiro, “On the Evaluation, Management and Improvement of Data Quality in Streaming Time Series,” *IEEE Access*, vol. 10, pp. 81458–81475, 2022, doi:10.1109/access.2022.3195338.

[21] L. Ehrlinger and W. Wöß, “A Survey of Data Quality Measurement and Monitoring Tools,” *Frontiers in Big Data*, vol. 5, Mar. 2022, doi:10.3389/fdata.2022.850611.

[22] A. Goknil et al., “A Systematic Review of Data Quality in CPS and IoT for Industry 4.0,” *ACM Computing Surveys*, vol. 55, no. 14s, pp. 1–38, Jul. 2023, doi: 10.1145/3593043.

[23] A. Sgora and P. Chatzimisios, “Defining and Assessing Quality in IoT Environments: A Survey,” *IoT*, vol. 3, no. 4, pp. 493–506, Dec. 2022, doi: 10.3390/iot3040026.

[24] K. Kolomvatsos, “A distributed, proactive intelligent scheme for securing quality in large scale data processing,” *Computing*, vol. 101, no. 11, pp. 1687–1710, Dec. 2018, doi: 10.1007/s00607-018-0683-9.

[25] M. Kara, O. Lamouchi, and A. Ramdane-Cherif, “A Quality Model for the Evaluation AAL Systems,” *Procedia Computer Science*, vol. 113, pp. 392–399, 2017, doi: 10.1016/j.procs.2017.08.354.

[26] S. Fagúndez, J. Fleitas, and A. Marotta, “Data Stream Quality Evaluation for the Generation of Alarms in the Health Domain,” *Journal of Intelligent Systems*, vol. 24, no. 3, pp. 361–369, Aug. 2015, doi: 10.1515/jisys-2014-0166.

[27] R. Abo and A. Even, “Sampling density and frequency as data quality determinants in smart grids,” *2017 Smart City Symposium Prague (SCSP)*, vol. 62, pp. 1–6, May 2017, doi: 10.1109/scsp.2017.7973349.

[28] S. Sicari, A. Rizzardi, C. Cappiello, D. Miorandi, and A. Coen-Porisini, “Toward Data Governance in the Internet of Things,” *New Advances in the Internet of Things*, pp. 59–74, Jun. 2017, doi: 10.1007/978-3-319-58190-3_4.

[29] C. C. G. Rodríguez and S. Servigne, “Managing Sensor Data Uncertainty,” *International Journal of Agricultural and Environmental Information Systems*, vol. 4, no. 1, pp. 35–54, Jan. 2013, doi: 10.4018/jaeis.2013010103.

[30] L. Erazo-Garzon, J. Erraez, L. Illescas-Peña, and P. Cedillo, “A Data Quality Model for AAL Systems,” *Information and Communication Technologies of Ecuador (TIC.EC)*, pp. 137–152, Nov. 2019, doi:10.1007/978-3-030-35740-5_10.

[31] C. C. Castello, J. Sanyal, J. Rossiter, Z. Hensley and J. R. New, "Sensor data management validation correction and provenance for building technologies", *ASHRAE Transactions*, vol. 120, pp. 370-382, 2014.

[32] J. Liono, P. P. Jayaraman, A. K. Qin, T. Nguyen, and F. D. Salim, “QDaS: Quality driven data summarisation for effective storage management in Internet of Things,” *Journal of Parallel and Distributed Computing*, vol. 127, pp. 196–208, May 2019, doi:10.1016/j.jpdc.2018.03.013.

[33] G. Hendrik Haan, J. van Hillegersberg, E. de Jong, and K. Sikkel, “Adoption of Wireless Sensors in Supply Chains: A Process View Analysis of a Pharmaceutical Cold Chain,” *Journal of theoretical and applied electronic commerce research*, vol. 8, no. 2, pp. 21–22, 2013, doi: 10.4067/s0718-18762013000200011.

[34] S. Sicari, A. Rizzardi, D. Miorandi, C. Cappiello, and A. Coen-Porisini, “A secure and quality-aware prototypical architecture for the Internet of Things,” *Information Systems*, vol. 58, pp. 43–55, Jun. 2016, doi:10.1016/j.is.2016.02.003.

[35] L. Pravato and T. E. Doyle, "Iot for remote wireless electrophysiological monitoring: proof of concept", *Conference of the Centre for Advanced Studies on Collaborative Research*, 2017.

[36] M. Bharti, S. Saxena, and R. Kumar, “Intelligent Resource Inquisition Framework on Internet-of-Things,” *Computers & Electrical Engineering*, vol. 58, pp. 265–281, Feb. 2017, doi:10.1016/j.compeleceng.2016.12.023.

[37] A. Leonardi, H. Ziekow, M. Strohbach, and P. Kikiras, “Dealing with Data Quality in Smart Home Environments—Lessons Learned from a Smart Grid Pilot,” *Journal of Sensor and Actuator Networks*, vol. 5, no. 1, p. 5, Mar. 2016, doi: 10.3390/jsan5010005.

[38] M. L. Lopes De Faria, C. E. Cugnasca, and J. R. A. Amazonas, “Insights Into IoT Data and an Innovative DWT-Based Technique to Denoise Sensor Signals,” *IEEE Sensors Journal*, vol. 18, no. 1, pp. 237–247, Jan. 2018, doi: 10.1109/jsen.2017.2767383.

[39] J. E. Siegel, S. Kumar, and S. E. Sarma, “The Future Internet of Things: Secure, Efficient, and Model-Based,” *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2386–2398, Aug. 2018, doi:10.1109/jiot.2017.2755620.

[40] A. Karkouch, H. Mousannif, H. A. Moatassime, and T. Noel, “A model-driven architecture-based data quality management framework for the internet of Things,” *2016 2nd International Conference on Cloud Computing Technologies and Applications (CloudTech)*, vol. 3, pp. 252–259, May 2016, doi: 10.1109/cloudtech.2016.7847707.

[41] H. B. Sta, “Quality and the efficiency of data in ‘Smart-Cities,’” *Future Generation Computer Systems*, vol. 74, pp. 409–416, Sep. 2017, doi: 10.1016/j.future.2016.12.021.

[42] N. Nesa, T. Ghosh, and I. Banerjee, “Outlier detection in sensed data using statistical learning models for IoT,” *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, vol. 29, pp. 1–6, Apr. 2018, doi: 10.1109/wcnc.2018.8376988.

[43] N. A. M. Alduais, J. Abdullah, A. Jamil, L. Audah, and R. Alias, “Sensor node data validation techniques for realtime IoT/WSN application,” *2017 14th International Multi-Conference on Systems, Signals & Devices (SSD)*, vol. 47, pp. 760–765, Mar. 2017, doi:10.1109/ssd.2017.8166984.

[44] S. Gill and B. Lee, “A Framework for Distributed Cleaning of Data Streams,” *Procedia Computer Science*, vol. 52, pp. 1186–1191, 2015, doi: 10.1016/j.procs.2015.05.156.

[45] J. Borges Neto, T. Silva, R. Assunção, R. Mini, and A. Loureiro, “Sensing in the Collaborative Internet of Things,” *Sensors*, vol. 15, no. 3, pp. 6607–6632, Mar. 2015, doi: 10.3390/s150306607.

[46] A. Kos, S. Tomažič, and A. Umek, “Evaluation of Smartphone Inertial Sensor Performance for Cross-Platform Mobile Applications,” *Sensors*, vol. 16, no. 4, p. 477, Apr. 2016, doi: 10.3390/s16040477.

[47] P. Spachos, L. Song, and K. N. Plataniotis, “Wireless noise prevention for mobile agents in smart home,” *2017 IEEE International Conference on Communications (ICC)*, pp. 1–6, May 2017, doi:10.1109/icc.2017.7996993.

[48] Z. C. M. Candra, H.-L. Truong, and S. Dustdar, “On Monitoring Cyber-Physical-Social Systems,” *2016 IEEE World Congress on Services (SERVICES)*, pp. 56–63, Jun. 2016, doi:10.1109/services.2016.14.

[49] B. Jang, S. Park, J. Lee, and S.-G. Hahn, “Three Hierarchical Levels of Big-Data Market Model Over Multiple Data Sources for Internet of Things,” *IEEE Access*, vol. 6, pp. 31269–31280, 2018, doi:10.1109/access.2018.2845105.

[50] A. Kothari, V. Boddula, L. Ramaswamy, and N. Abolhassani, “DQS-Cloud: A Data Quality-Aware Autonomic Cloud for Sensor Services,” *Proceedings of the 10th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing*, 2014, doi: 10.4108/icst.collaboratecom.2014.257475.

[51] F. Adelantado, X. Vilajosana, P. Tuset-Peiro, B. Martinez, J. Melia-Segui, and T. Watteyne, “Understanding the Limits of LoRaWAN,” *IEEE Communications Magazine*, vol. 55, no. 9, pp. 34–40, 2017, doi:10.1109/mcom.2017.1600613.

[52] R. A. Atmoko, R. Riantini, and M. K. Hasin, “IoT real time data acquisition using MQTT protocol,” *Journal of Physics: Conference Series*, vol. 853, p. 012003, May 2017, doi: 10.1088/1742-6596/853/1/012003.

[53] M. Gupta, C. Holloway, B. M. Heravi, and S. Hailes, “A comparison between smartphone sensors and bespoke sensor devices for wheelchair accessibility studies,” *2015 IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, vol. 65, pp. 1–6, Apr. 2015, doi:10.1109/issnip.2015.7106900.

[54] F. H. Bijarbooneh, W. Du, E. C.-H. Ngai, X. Fu, and J. Liu, “Cloud-Assisted Data Fusion and Sensor Selection for Internet of Things,” *IEEE Internet of Things Journal*, vol. 3, no. 3, pp. 257–268, Jun. 2016, doi: 10.1109/jiot.2015.2502182.

[55] B. Fekade, T. Maksymyuk, M. Kyryk, and M. Jo, “Probabilistic Recovery of Incomplete Sensed Data in IoT,” *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2282–2292, Aug. 2018, doi:10.1109/jiot.2017.2730360.

[56] X. Yan, W. Xiong, L. Hu, F. Wang, and K. Zhao, “Missing Value Imputation Based on Gaussian Mixture Model for the Internet of Things,” *Mathematical Problems in Engineering*, vol. 2015, pp. 1–8, 2015, doi: 10.1155/2015/548605.

[57] I. P. S. Mary and L. Arockiam, “Imputing the missing data in IoT based on the spatial and temporal correlation,” *2017 IEEE International Conference on Current Trends in Advanced Computing (ICCTAC)*, pp. 1–4, Mar. 2017, doi: 10.1109/icctac.2017.8249990.

[58] P. Priller, A. Aldrian, and T. Ebner, “Case study: From legacy to connectivity migrating industrial devices into the world of smart services,” *Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA)*, vol. 1, pp. 1–8, Sep. 2014, doi:10.1109/etfa.2014.7005136.

[59] Y. Zhang, C. Szabo, and Q. Z. Sheng, “Cleaning Environmental Sensing Data Streams Based on Individual Sensor Reliability,” *Web Information Systems Engineering – WISE 2014*, pp. 405–414, 2014, doi: 10.1007/978-3-319-11746-1_29.

[60] Y. Ma, J. Jin, Q. Huang, and F. Dan, “Data Preprocessing of Agricultural IoT Based on Time Series Analysis,” *Intelligent Computing Theories and Application*, pp. 219–230, 2018, doi:10.1007/978-3-319-95930-6_21.

[61] A. S. Dmitriev, E. V. Efremova, and M. Yu. Gerasimov, “Multimedia sensor networks based on ultrawideband chaotic radio pulses,” *Journal of Communications Technology and Electronics*, vol. 60, no. 4, pp. 393–401, Apr. 2015, doi: 10.1134/s1064226915040051.

[62] R. Dong, L. J. Ratliff, A. A. Cárdenas, H. Ohlsson, and S. S. Sastry, “Quantifying the Utility--Privacy Tradeoff in the Internet of Things,” *ACM Transactions on Cyber-Physical Systems*, vol. 2, no. 2, pp. 1–28, Apr. 2018, doi: 10.1145/3185511.

[63] Z. Huang, T. Xie, T. Zhu, J. Wang, and Q. Zhang, “Application-driven sensing data reconstruction and selection based on correlation mining and dynamic feedback,” *2016 IEEE International Conference on Big Data (Big Data)*, pp. 1322–1327, Dec. 2016, doi:10.1109/bigdata.2016.7840737.

[64] D. Pal, V. Vanijja, C. Arpnikanondt, X. Zhang, and B. Papasratorn, “A Quantitative Approach for Evaluating the Quality of Experience of Smart-Wearables From the Quality of Data and Quality of Information: An End User Perspective,” *IEEE Access*, vol. 7, pp. 64266–64278, 2019, doi: 10.1109/access.2019.2917061.

[65] M. Kohbalan, M. HasanAli, M. A. Ismail, C. W. Howe, M. S. Mohamad, S.Deris, “An Evaluation of Machine Learning Algorithms for Missing Values Imputation,” *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, pp. 415-420 2019.