

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/334776380>

Data quality and the Internet of Things

Article in *Computing* · February 2020

DOI: 10.1007/s00607-019-00746-z

CITATIONS

2

READS

488

4 authors, including:



Patrick Nitschke

Universität Koblenz-Landau

2 PUBLICATIONS 9 CITATIONS

[SEE PROFILE](#)



Susan Patricia Williams

Universität Koblenz-Landau

75 PUBLICATIONS 1,239 CITATIONS

[SEE PROFILE](#)



Didar Zowghi

University of Technology Sydney

183 PUBLICATIONS 4,294 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Optimising mobile learning in science and maths [View project](#)



CEIR - Enterprise Collaboration Systems (ECS) [View project](#)



Data quality and the Internet of Things

Caihua Liu¹ · Patrick Nitschke² · Susan P. Williams² · Didar Zowghi¹

Received: 10 January 2019 / Accepted: 25 July 2019 / Published online: 30 July 2019
© Springer-Verlag GmbH Austria, part of Springer Nature 2019

Abstract

The Internet of Things (IoT) is driving technological change and the development of new products and services that rely heavily on the quality of the data collected by IoT devices. There is a large body of research on data quality management and improvement in IoT, however, to date a systematic review of data quality measurement in IoT is not available. This paper presents a systematic literature review (SLR) about data quality in IoT from the emergence of the term IoT in 1999 to 2018. We reviewed and analyzed 45 empirical studies to identify research themes on data quality in IoT. Based on this analysis we have established the links between data quality dimensions, manifestations of data quality problems, and methods utilized to measure data quality. The findings of this SLR suggest new research areas for further investigation and identify implications for practitioners in defining and measuring data quality in IoT.

Keywords Internet of Things · Data quality · Systematic review · Research themes

Mathematics Subject Classification 0002

✉ Caihua Liu
Caihua.Liu@uts.edu.au

Patrick Nitschke
nitschke@uni-koblenz.de

Susan P. Williams
williams@uni-koblenz.de

Didar Zowghi
Didar.Zowghi@uts.edu.au
<http://www.uts.edu.au/staff/didar.zowghi>

¹ Faculty of Engineering and Information Technology, University of Technology Sydney, Broadway 15, Ultimo, NSW 2007, Australia

² Faculty of Computer Science, University of Koblenz-Landau, Universitätsstraße 1, 56070 Koblenz, Germany

1 Introduction

The ongoing and seamless integration of the physical and digital worlds through the incorporation of sensors and devices into everyday objects is predicted to transform the ways we live and work in the future [1,2]. The Internet of Things (IoT) has been identified as a key driver for this technological revolution, which will foster the creation of new products and services in diverse domains, ranging from agriculture [3] to manufacturing and everything in between [4]. These new products, systems and services are expected to create an annual economic impact of \$2.7 trillion to \$6.2 trillion by 2025 [5]. Despite the diverse areas of IoT application, the majority of new systems and services rely heavily on the data collected by IoT devices and ensuring the quality of the data that provides the baseline of the IoT services is crucial and a fundamental concern in the design of IoT based products and services [6,7]. For example, in March 2019, the Tesla's Autopilot was engaged in a fatal crash of a Tesla electric vehicle, because the data coming from the vehicle's self-driving sensor (i.e. radars) did not match with actual road situations, failing to detect objects crossing the road and causing the vehicle to crash into a truck [8].

To date a growing body of research studies have investigated data quality (DQ) focusing on aspects such as: DQ dimensions, DQ problems, and techniques to improve DQ in IoT. However, this body of research and the terms used to describe DQ dimensions are fragmented and inconsistent [9]. The inconsistent use of the terms could pose challenges in (1) understanding similar or different DQ dimensions being discussed, and (2) explaining similar or different DQ problems for a certain dimension. Furthermore, prior reviews of data quality in IoT [9–11] are concerned more with the techniques used to improve DQ and limited attention is directed towards the identification of methods to measure DQ. However, DQ measurement is an important facet of managing data and understanding DQ measurement methods for IoT could (1) assist in the accurate measurement and quality assessment of the collected IoT data, and (2) provide a baseline for supporting subsequent data management and data usage activities.

Therefore, the aim of this study is to review the extant literature to identify IoT DQ dimensions, the DQ problems related to these dimensions, and the methods used to measure these dimensions. By providing a clear picture of the research themes relating to IoT DQ our goal is to map out which aspects have been studied and identify potential areas for further investigation. The findings of this review will also serve as a starting point for defining and measuring IoT DQ dimensions and identify problem areas that need to be addressed in order to improve IoT DQ. To achieve this aim the study is guided by four research questions (RQs):

RQ1: What are the key research themes in existing research related to data quality in IoT?

RQ2: What are the dimensions of data quality in IoT?

RQ3: What are the manifestations of data quality problems in IoT?

RQ4: What methods are used to measure data quality in IoT?

Following the guidelines of Wolfswinkel et al. [12] and Kitchenham et al. [13], we conducted a systematic literature review (SLR) of the empirical studies that have a focus on DQ in IoT from the emergence of the term IoT in 1999 [14] to 2018. SLR is

a methodical process of collecting and collating the published empirical studies with systematic criteria for selection to reduce bias and provide transparency to the process.

The paper is organized as follows: Sect. 2 defines the concepts used in this research and reviews related studies and Sect. 3 describes the research method. Guided by the RQs, Sect. 4 presents the results from the data extraction and analysis processes. Section 5 provides a discussion of the research findings and proposes potential areas for further investigation. Section 6 provides a summary and conclusions to the study.

2 Concepts and related studies

2.1 Concepts: data quality, dimension and manifestation

In this section, we provide the definitions for three important concepts used in our study: data quality, dimension, and manifestation. The term quality has been defined both as “fitness for use” and as “conformance to requirements” [15]. This generic understanding of quality assists in the tasks of conceptually comparing and distinguishing data quality and information quality [16]. For example, Tilly et al. [16] differentiated the two based on the definition of data and information: data is objective presenting a phenomenon in the real/physical world, while information is subjective and places data into a context using an information system that users can understand. In this study, we are interested in the quality of IoT data, and therefore, we draw on the published literature on DQ to study and understand the topic of interest.

Another two important concepts used in the study are dimension and manifestation. A dimension of DQ refers to an individual aspect (or attribute) of DQ such as completeness and accuracy [17], which can be empirically measured. Manifestation of DQ problems is defined as a symptom of, or challenge arising from, data errors or anomalies.

2.2 Related studies

This section summarizes prior literature reviews on DQ in IoT. As shown in Table 1, there is a growing body of research that provides reviews of articles that study DQ in the IoT context [9–11, 18]. For example, Cai and Zhu [18] identified the challenges of DQ in Big Data that is increasingly accumulated by advanced technologies such as IoT, and developed a framework for quality assessment on DQ dimensions. During the same period, Karkouch et al. [10] reviewed data outliers and their impacts on IoT applications and compared data cleaning techniques used to address the uncertainty in the collected data. In addition, Qin et al. [11] identified multiple DQ problems based on data streams and data storage models, while Karkouch et al. [9] defined a set of IoT-related DQ dimensions, and discussed the factors influencing DQ, manifestations of DQ problems and techniques to improve DQ.

Several observations can be made from Table 1. First, related studies were reviewed and the DQ literature was analyzed from diverse perspectives to improve our understanding of this academic niche in the realm of IoT. However, some reviews only partially studied DQ in the IoT context. Moreover, they do not focus on the methods used to measure DQ or do not completely link these methods with the DQ dimen-

Table 1 Characteristics of review studies on DQ in IoT

Characteristic	Ref. [18]	Ref. [10]	Ref. [11]	Ref. [9]
Publication year	2015	2015	2016	2016
No. of the papers included	NM	NM	NM	NM
No. of the databases used	NM	NM	NM	NM
Time frame of publications	NM	NM	NM	NM
Focus on DQ?	Yes	Yes	Partially	Yes
Focus on IoT?	Partially	Yes	Yes	Yes
Dimensions defined?	Yes	No	Yes	Yes
Measurement of DQ elaborated?	No	No	No	No
Manifestations of DQ problems identified?	Yes	Yes	Yes	Yes

NM Not Mentioned

sions. To address this limitation, our review examines methods used to measure DQ, manifestations of DQ problems and maps these to the DQ dimensions identified in this study. We identify a set of research themes related to DQ in IoT. The goal is to contribute to knowledge about the defining and measuring of DQ in IoT, and provide a means to enable practitioners to become more aware of DQ problems and to begin to address these problems. Second, few existing reviews have reported their literature analysis strategies, providing limited insight into the criteria and processes used to search and select articles for review and analysis. We therefore search, select and analyze the relevant studies, which provides greater transparency to the process and enables future researchers to review and build on the results.

3 Research method

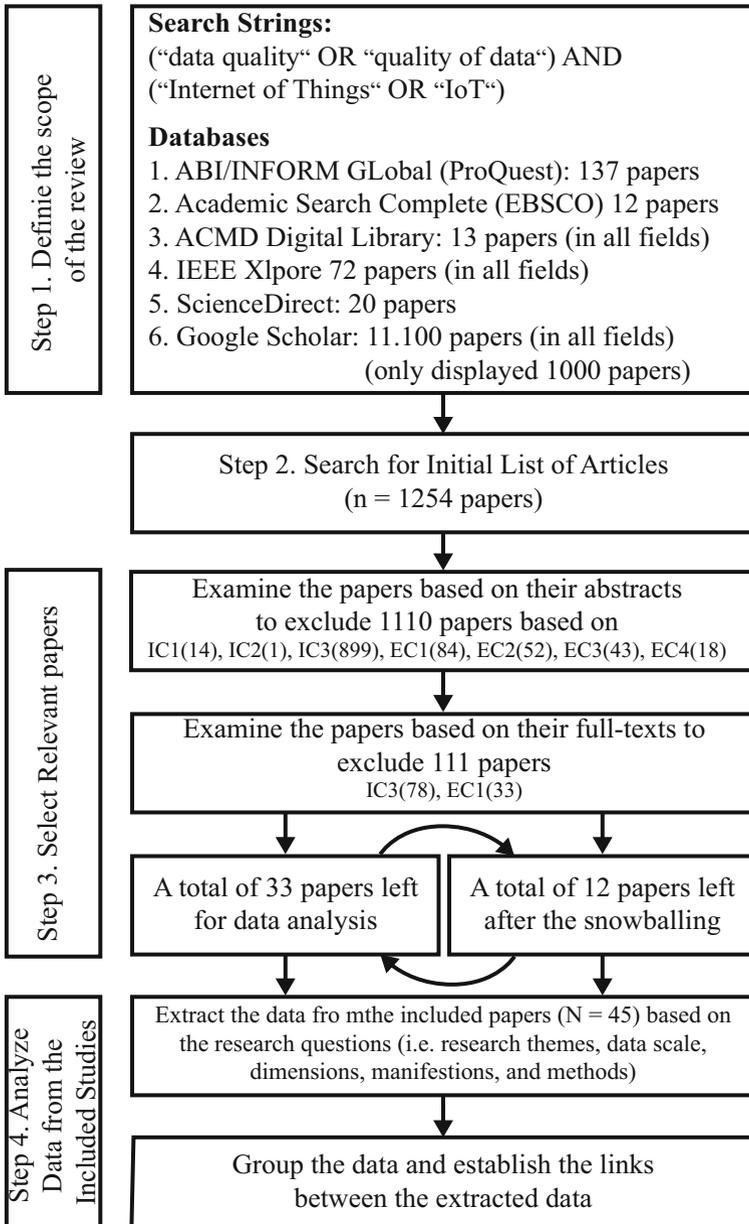
To conduct this SLR of data quality in IoT, we follow the guidelines of Wolfswinkel et al. [12] and Kitchenham et al. [13], incorporating four steps: (1) define the scope of the review; (2) search for initial list of articles; (3) select relevant papers; and (4) analyze data from the included studies. Figure 1 presents the research process employed in this review.

3.1 Define the scope of the review

Four main activities are involved in this step: establishment of inclusion and exclusion criteria of an article in the data set, identification of appropriate fields of research, selection of databases and outlets, and formulation of search terms [12,13].

3.1.1 Establishment of inclusion and exclusion criteria

The inclusion criteria adopted in this review are: (1) the articles included are published in English (IC1); (2) articles are selected with a publication date from 1999 to 2018, inclusive (IC2); and (3) the theme of the paper has a focus on DQ in the IoT context (IC3). The removal of papers is based on the following exclusion criteria: (1) the studies do not provide empirical findings themselves (EC1); (2) the articles are not



Note that the number in the brackets indicates the number of the papers that were removed based on the inclusion and exclusion criteria.

Fig. 1 Research process of this review

peer-reviewed research publications (EC2); (3) the papers are not accessible online (EC3); or (4) the papers are duplicates (EC4).

3.1.2 Identification of fields of research

As mentioned in the Introduction, IoT has widespread impacts across diverse domains and fields. Our study thus investigates a corpus of research articles that span diverse areas such as Engineering and Computer Science, covering a broad coverage of research fields to generalize our findings on DQ in IoT.

3.1.3 Selection of databases and outlets

In this review, we used six databases as the initial resources for the search as advised by [19]. By using the databases ABI/ INFORM Global (ProQuest), Academic Search Complete (EBSCO), ACM Digital Library, IEEE Xplore, ScienceDirect, and Google Scholar that contain the potential largest set of peer-reviewed and English literature on the topic of interest, an exhaustive view on the field could be established. To further enrich the quality of the sample for data analysis, backward (i.e. using the reference list to identify new papers) and forward snowballing (i.e. finding citations to the papers) approaches were used to exhaust the data [12,13].

3.1.4 Formulation of search terms

Our search keywords developed for this review contained: (1) “data quality” and (2) “Internet of Things”. Because we differentiated the definitions of data and information, we limited the theoretical lens to the studies of DQ for the search. We thus only adopted alternative term “quality of data” in the search to describe “data quality” as suggested by [20]. Furthermore, we used the alternative term “IoT” for “Internet of Things” as advised by [19]. These selected keywords enable us to draw from a broad range of domains and areas and identify a comprehensive set of publications related to data quality in. Accordingly, our search began with those keywords by using the Boolean operators as the following search strings: (“data quality” OR “quality of data”) AND (“Internet of Things” OR “IoT”) in the selected databases.

3.2 Search for initial list of articles

With the search strings, we screened the publications in Title, Abstract, and Keywords using the online databases to centralize our search. We also customized our search with search strings in the selected databases and identified the initial list of papers as shown in Fig. 1.

3.3 Select relevant papers

The purpose of this step is to filter the relevant papers for further analysis. We identified 1254 articles, of which 1110 were removed on abstract review based on our inclusion

and exclusion criteria. After full-text review, we further discarded 111 papers based on IC3 and EC1. In the snowballing process, we identified an additional 27 papers of which 12 papers then remained after the abstract and full-text review, based on inclusion and exclusion criteria. Finally, a total of 45 papers remained eligible for analysis (see Fig. 1). The full list of all the 45 papers included in the review is included in the “Appendix”. Each paper is given a unique identifier (the letter S followed by a number) so the paper can be referenced in the analysis and reporting of the findings.

3.4 Analyze data from the included studies

We read, analyzed and coded the findings and insights in the text of each selected article relevant to our RQs. However, not every article addresses all our RQs (i.e. research themes, dimensions, manifestations of DQ problems, and methods used to measure DQ dimensions). We thus extracted and recorded relevant responses from the included papers for the RQs into a table. After completing the table, we reviewed the list of research themes studied in the publications and merged the duplicates. Then, we grouped these research themes based on their similarity and gave each group an appropriate name. The lists of dimensions, manifestations of DQ problems, and methods used to measure DQ were classified in a similar way. Finally, we mapped the manifestations of DQ problems and methods to the DQ dimensions identified from the studies being analyzed.

4 Results

This section presents our findings derived from the reviewed studies based on our RQs. Section 4.1 answers RQ1 providing an overview of DQ studied in the IoT context, including: (1) definition of DQ (further discussed in Sect. 4.2 to answer RQ2), (2) DQ measurement (further discussed in Sect. 4.3 to answer RQ4), (3) analysis of manifestations of DQ problems (further discussed in Sect. 4.2 to answer RQ3), and (4) design and development of solutions to address DQ. Note that the research theme ‘design and development’ is not extended in either Sects. 4.2 or 4.3 as with the other three research themes due to the scope of our proposed RQs. Related studies [9,11] have further discussed relevant solutions designed and developed for addressing DQ in IoT, while this is not the focus of our study.

4.1 Research themes

In this review, we identified four main research themes that concern DQ in IoT, these are: (1) definition, (2) measurement, (3) analysis, and (4) design and development. Each research theme contains its own subcategories as summarized in Fig. 2.

Definition This research theme addresses how DQ has been defined in IoT. Generally, researchers described their DQ requirements and defined a set of DQ dimensions. The DQ dimensions identified are: (1) Accuracy, (2) Timeliness, (3) Completeness, (4)

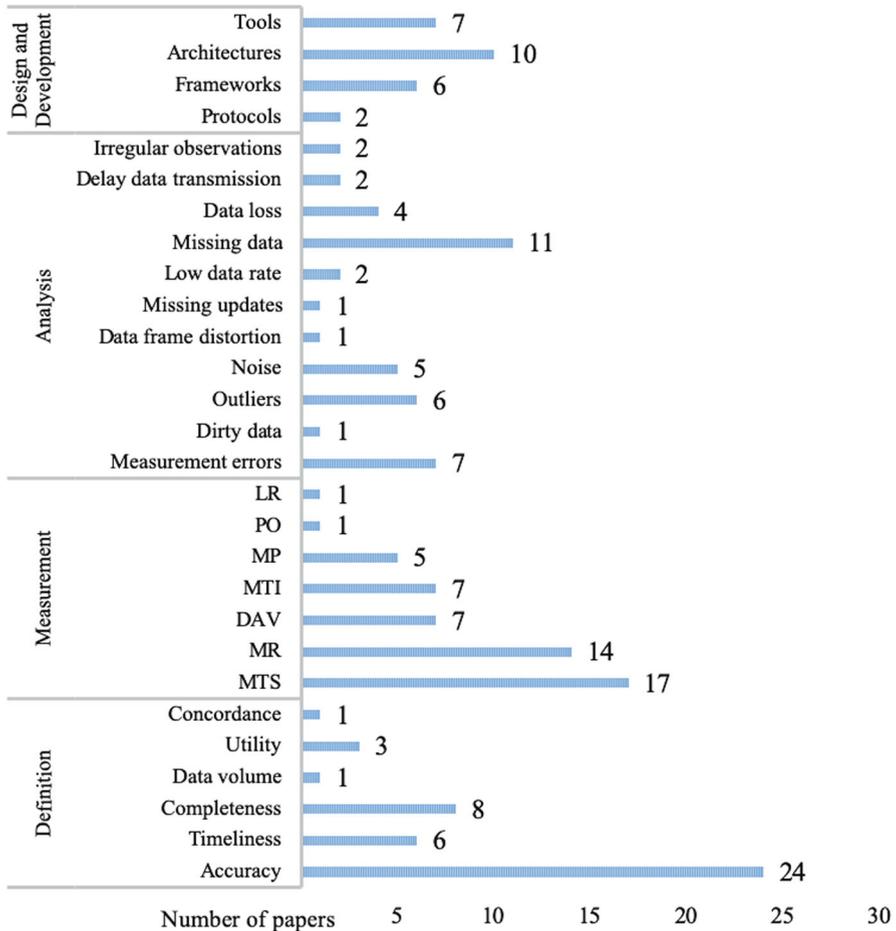


Fig. 2 Distribution of research themes on DQ in IoT in this review

Data volume, (5) Utility, and (6) Concordance, which are discussed in more detail in Sect. 4.2.

Measurement This research theme refers to the ways that the quality of data been measured in IoT. The analysis identified seven methods used to measure DQ: (1) Measurement among techniques, sources or defined attributes, (2) Measurement with a reference, (3) Devices or algorithms validation, (4) Measurement within time intervals, (5) Measurement of presence, (6) Process observation, and (7) Log files review. These methods are further described in Sect. 4.3.

Analysis This research theme concerns the analysis of DQ problems that influence the achievement of DQ in IoT. As one of the aims of this study is to identify the manifestations of DQ problems in IoT, these manifestations (further outlined in Sect. 4.2) are: measurement errors, noises, artifact error, data frame distortion, dirty data, outliers, missing data, missing updates, data loss, and delay data transmission.

Design and development This research theme describes how the quality of IoT data can be addressed and/or improved by a variety of solutions. Researchers commonly designed and developed: (1) protocols (for data transmission [S14, S25]), (2) frameworks (for storing IoT data [S9], collecting sensor data [S16, S21, S35, S40], and monitoring the delivered IoT data [S22]), (3) architectures (for monitoring DQ [S8] and filtering good data from the collected IoT data [S5, S24, S44], cleaning IoT data streams [S4, S7, S41], and providing data products [S13, S20, S30]), and (4) tools (for updating real-time data to a cloud [S6], identifying data anomalies [S18, S26], and dealing with missing data [S11, S36, S38, S39]), to address and/or improve DQ in IoT.

These findings about the research themes on DQ in IoT are consistent with the Total Data Quality Management Methodology [21]. In the included studies, firstly DQ requirements were defined and a set of DQ dimensions were developed for the measurement. Based on the results of the measurement, researchers identified relevant DQ problems and analyzed the causes of these problems, in such a way that the core areas could be identified for DQ improvement. Then, a variety of solutions were designed and developed to address and/or improve the quality of IoT data, in order to provide quality-assured IoT products and services.

4.2 Dimensions and manifestations of DQ problems

This study reviewed and analyzed DQ dimensions used in IoT and manifestations of DQ problems revealed in the included studies. These are summarized in Table 2. DQ dimensions that were only mentioned or described but not measured were not included in the table. Column two and three of Table 2 present definitions adopted from ISO25024 [22] and alternative terms that have been adopted in the literature to describe these dimensions. Column four of Table 2 summarizes the examples that have been used in the articles to define these dimensions, and delineates several instances from the included studies that have explicitly described and/or explained the manifestations of DQ problems for these dimensions, based on direct observation or actual experience of the respective authors.

4.2.1 Accuracy

Accuracy was the most frequently used dimension of DQ in IoT, being an area of focus in 53% of the included papers. Data was deemed as accurate when an observation for the object truly reflected its real-world situation [S2, S5, S6]. Li et al. [S17] argued that validity is a different notion from accuracy and correctness because validity is more subjective based on an acceptable range of a certain attribute and a specific application scenario. However, Hendrik et al. [S2] used the term validity to describe the extent to which the condition of an object is accurately represented that appears to also be relevant to the conception of accuracy. Other terms used in this review to describe this dimension include precision [S5, S29], validity [S2, S17], and correctness [S6, S23, S24, S42].

Table 2 DQ dimensions and manifestations of DQ problems in IoT identified in this review

Dimension	Definition from ISO [22]/related work	Definition (and alternative terms) form the included studies	Example of DQ problems
Accuracy	“The degree to which data has attributes that correctly represent the true value of the intended attribute of a concept or event in a specific context of use.” (Accuracy used in ISO [22])	The extent to which an observation for the object truly reflects its real-world situation [S2, S5, S6] (Precision, Validity, Correctness)	Measurement errors [S2, S5, S12, S13, S27, S37, S42], Dirty data [S13], Outliers [S7, S13, S18, S23, S26, S40], Noise [S1, S18, S43], Data frame distortion [S3]
Timeliness	“The degree to which data has attributes that are of the right age in a specific context of use.” (Currentness used in ISO [22])	The extent to which an observation for the object is updated at a desired time of interest [S5, S17] (Currency, Volatility, Latency, Freshness, Data rate, Delay, Frequency)	Missing updates [S17], Low data rate [S22]
Completeness	“The degree to which subject data associated with an entity has values for all expected attributes and related entity instances in a specific context of use.” (Completeness used in ISO [22])	The extent to which all expected data is provided by IoT services [S5, S17] (Availability, Missing data)	Missing data [S2, S5, S11, S14, S17, S28, S29, S32, S36, S38, S39]
Utility	“The degree to which data can be accessed in a specific context of use.” (Accessibility used in ISO [22])	The extent to which relevant data is accessed by data consumers from IoT datasets during a certain period of time [S9, S14, S33] (Usage, Frequency, Relevancy)	Noise [S9, S33], Data loss [S14, S33], Missing data [S14, S33]
Data volume	“The number of raw data items (values) available for use to compute a result data item (in a stream query or sub-query) [9, p. 61].”	The number of data components are transmitted from a source to a consumer for generating a data result [S3]	Data loss [S3, S34], Delay data transmission [S3, S34], Data frame distortion [S3]
Concordance	“The degree to which data has attributes that are free from contradiction and are coherent with other data in a specific context of use.” (Consistency used in ISO [22])	The extent to which the data elements from a data source are in an agreement with the data elements from further individual data sources that report correlating effects [S8]	Irregular observations [S8, S41]

Data accuracy can be impaired by measurement errors [S2, S5, S12, S13, S42], which can be caused by issues such as the wrong placement [S2, S27] or selection [S13, S37] of sensors. For example, if a temperature sensor for a product is placed outside the insulated parcel, it does not read the product temperature, but the temperature of the environment, leading to potentially wrong conclusions [S2]. Furthermore, due to limitations of the sensors, the data detected at the sensors could have uncertainty that generates inaccurate data, and high uncertainty in sensor readings could also lead to dirty data [S13].

Note that an outlier could be defined as an observation that significantly differs from others in the sample [S13, S18]. An outlier could be a data error due to sensor faults [S7, S13, S18, S23, S26, S40]. At the same time, an outlier also could be an important event that represents a phenomenon of changing in the consistent real-world state (e.g. occurrence of fire) [S7, S13, S18]. Thus, DQ problems about outliers in this dimension refer to data errors caused by sensor faults.

Another manifestation of data inaccuracy was noise, which referred to any undesired change that deviates from the original signal [23]. This DQ problem could be caused by defective sensors [S1, S18, S43], e.g. due to exhausted batteries [24], faulty memory cells, bit error in transmission [25], or interference when multiple wireless devices transmit the data simultaneously on the same frequency bands (e.g. BLE and Wi-Fi use the same 2.4GHz bands) [26]. Data frame distortion also reflected the problem associated with data inaccuracy [S3], and meanwhile revealed DQ problems under the dimension of data volume that is detailed below.

4.2.2 Timeliness

The IoT data was considered timely when an observation for an object was updated at a desired time of interest [S5, S17]. Alternative terms adopted to describe this dimension include currency [S5, S17], volatility [S5], latency [S12], freshness [S12, S22], data rate [S22, S30], delay [S21, S23], and frequency [S21]. The manifestations of DQ problems on this dimension were missing updates [S17] and low data rate [S22]. An example of low data rates, which influence timeliness, is the deployment of devices in constrained contexts such as agriculture. In this context, devices have constrained resources such as energy and are required to communicate across large distances using technologies such as LoRaWAN or SigFox, which are prone to low data rates and high latency but require very little energy [27].

4.2.3 Completeness

Completeness was defined as whether all expected data was provided by IoT services [S5, S17]. Some studies [S5, S13, S23, S25, S37] utilized the term completeness, and others referred to data availability [S17, S19] and missing data [S14, S36, S38, S39]. Missing data can be caused by sensor inefficiencies, communication issues [S5, S14, S29, S32, S36] or by attacker's intercepting or manipulating data [28]. Furthermore, Li et al. [S17] found that the lack of data updates could affect achieving the required data, limiting data availability for users. Additionally, data owners selectively disclose

the data based on certain constraints (e.g. privacy considerations), resulting in less detailed data being available for users [S2].

4.2.4 Utility

Utility referred to frequency and relevancy of the access of data consumers (users) from the IoT dataset during a certain period of time [S9]. Alternative terms such as usage, frequency, and relevancy [S9] were used in the reviewed papers to describe this dimension. One of the main DQ problems of utility was noise [S9, S33]. As we mentioned, multiple IoT devices transmitting data simultaneously could cause noise. Thus, Liono et al. [S9] showed that there is a fixed probability for an instance of noise for each data consumer who accesses the IoT dataset. This could have an impact on the extent to which required data is accessed by different data consumers. Furthermore, research has shown that inactive sensor nodes could result in data loss in data transmission [S14, S33]. That is to say, the transmitted data could be missing when some sensor nodes fail to communicate with other connected nodes in the network. Hence, data loss or missing data could decrease the utility of IoT data.

4.2.5 Data volume

Data volume was interpreted as the number of data components transmitted from a source to a consumer for generating a result [S3] [9]. Dmitriev et al. [S3] indicated that since the physical transmission rate of a network node is fixed, data volume transmitted by a single node would be low when there is a large amount of data required, resulting in data loss and delaying data transmission that influence data volume. Furthermore, when IoT devices collect video and image data that requires data compression and subsequent sound or image recovery, data loss and delay are also accompanied in this process [S3, S34]. The reconstruction of video and image data could, thus be different from source data, presenting data frame distortion [S3].

4.2.6 Concordance

Data was considered concordant when there was an agreement between data elements from a data source and the data elements from other individual data sources that report correlating effects [S8]. The manifestation of DQ problems on this dimension was irregular readings reported by the sensors for a certain object, which could be caused by non-concurring observations from the data sources [S8] or by sensor faults [S41]. For instance, Kuemper et al. [S8] found that some vehicle speed readings and car counts were inconsistent between two different data sources for the selected traffic incidents due to different time and view of observations. While Zhang et al. [S41] indicated that malfunctioning sensors could not consistently report the readings, which will have an impact on data concordance.

We observed that the manifestations of DQ problems overlap among different DQ dimensions. For example, noise pertains to the DQ problem on data accuracy and

at the same time is related to data utility because noisy data could affect how much required data is accessed by data consumers [S9]. Data frame distortion could be a DQ problem of accuracy, presenting the reconstructed data distorted from source data [S3]. While data frame distortion also concerns the DQ problem of data volume since data distortion could influence the number of data components that are transmitted for generating a result [S3]. As to data loss, it could be a DQ problem concerning data utility [S14, S33] or data volume [S3, S34]. Furthermore, when data loss occurs in the process of transmission [S14, S33] or data is not timely updated [S3, S34], the data received could be missing that reflects a DQ problem associated with completeness or utility [S14, S33]. Accordingly, these observations imply that DQ dimensions are distinguished conceptually and defined heavily based on different purposes, but these dimensions could also interact with each other.

4.3 Methods used to measure DQ

We conducted a similar analysis process to identify the methods used to measure DQ in IoT. These methods are divided into seven categories, some of which were utilized to measure multiple dimensions of DQ, as outlined in Table 3.

Measurement between techniques, sources or defined attributes (MTS): Data elements in two or more IoT datasets that are derived from different techniques, sources or defined attributes are compared to see if there are agreements in these elements. MTS was the most frequently used method to measure DQ dimensions identified in this review. For example, researchers measured DQ among the IoT datasets collected by different experimental settings (e.g. take distances between a transmitter and a receiver into account [S3]), protocols [S25], data sources [S5, S6, S8, S9, S10, S19, S23, S30, S31, S33, S43] or algorithms [S14, S18, S29, S45].

Measurement with a reference (MR): A dataset derived from another source serves as a reference to compare with the collected IoT dataset to determine whether or not there are agreements in these elements. MR was the second frequently used method to measure DQ. For instance, Karkouch et al. [S13] measured completeness and accuracy by referring to the results with prior literature that used the same IoT dataset. Some studies adopted either actual values [S10], an applicable range of values [S6, S17, S42, S44], historical data [S1, S8, S12, S24], or spatial-temporal correlated measured values for the objects provided by the sensor and its neighbors [S7, S8, S15, S18, S45], as a reference for DQ measurement.

Devices or algorithms validation (DAV): The collected IoT dataset is examined by using well developed devices or algorithms to ascertain whether or not expected values present [S3]. DAV was the third commonly used method to measure DQ in this review. Some studies divided an IoT dataset into a training dataset and a testing dataset and then measured the accuracy by looking at the agreement between the results of the testing dataset and the expected values using the proposed approaches implemented on the training dataset [S4, S7, S8, S13, S18, S26].

Measurement between time intervals (MTI): The IoT dataset is examined during a fixed time interval to determine how good is the data collected. For instance, Liono et al. [S9] divided an IoT dataset into data slices based on a certain temporal duration,

Table 3 A summary of methods used to measure DQ dimensions in this review

Dimensions	MTS	MR	DAV	MTI	MP	PO	LR
Accuracy	[S5, S6, S10, S18, S19, S29, S30, S31, S43, S45]	[S1, S6, S7, S8, S10, S12, S13, S15, S17, S18, S24, S42, S44, S45]	[S4, S7, S8, S13, S18, S26]	[S1, S7, S31]		[S2]	[S2]
Timeliness	[S5, S29]			[S17, S21, S22]			
Completeness	[S5, S25]	[S13, S14]			[S11, S13, S17, S25, S39]		
Utility	[S14, S30]			[S9]			
Data volume	[S3]		[S3]				
Concordance	[S8]						

to determine how much data was accessed by consumers from the dataset over the time. Others reported the changes of observations for the objects [S1, S7, S17, S21, S31] or data rates generated by IoT services [S22], based on multiple updates of the observation period.

Measurement of presence (MP): The collected IoT dataset is examined to determine whether or not data elements present. In this review, most studies reported the extent to which non-missing data presented in a given IoT dataset to describe completeness [S11, S13, S17, S25, S39].

Process observation (PO): The loading process and physical association of sensors are situ monitored to ascertain whether or not the data collected makes sense. Only one study [S2] adopted this method to measure DQ in this review and it reported that a sensor placed on the outside of a parcel provides inaccurate product temperature due to external conditions, resulting in inaccurate data.

Log files review (LR): Log files and/or claims investigations on the collected IoT data are reviewed to determine whether or not data errors or anomalies present. For example, Hendrik et al. [S2] reviewed the log files of errors and claims investigations to ascertain accuracy of the sensor data collected in a pharmaceutical cold supply chain.

5 Discussion

In this section, we summarize the findings of this review and the extent to which they answer our RQs. We then explore the findings and outline areas that require further exploration, and present the implications and limitations of this study.

5.1 Research themes

Findings relating to RQ1: The research themes related to DQ in IoT include: (1) Definition, (2) Measurement, (3) Analysis, and (4) Design and Development.

The focus of the empirical studies related to DQ in IoT was on DQ management and improvement from a technique perspective, which rely heavily on experiments and case studies. However, few research studies have investigated which DQ dimensions are important or should be included in the IoT context. Because the types of the data collected by IoT devices could be diverse (e.g. image and video data), this could result in different DQ requirements and DQ dimensions being defined. There is a clear lack of guidelines or checklists that suggest specific concepts of defining DQ dimensions for IoT data. Such guidelines or checklists could help users include appropriate DQ dimensions to determine how good is the data and to facilitate the consistent use of the terms that describe DQ dimensions. Thus, research is required to further identify and define the commonly used DQ dimensions for IoT data, through in-depth interviews or surveys with practitioners.

5.2 Dimension and manifestation of DQ problems

Findings relating to RQ2: The DQ dimensions used in IoT are: (1) Accuracy, (2) Timeliness, (3) Completeness, (4) Utility, (5) Data volume, and (6) Concordance.

Findings relating to RQ3: The manifestations of DQ problems identified are: measurement errors, noise, artifact error, data frame distortion, dirty data, outliers, missing data, missing updates, data loss, and delay data transmission.

The terms used to describe DQ dimensions in this review were inconsistent and a single article sometimes used different terms to describe a given dimension. In this SLR, we identified two additional dimensions “Utility” and “Concordance” that were not included in the related studies [9–11, 18]. Furthermore, Karkouch et al. [9] indicated “Access Security” as a DQ dimension for IoT, however, we found that security looks at how data sources have been encrypted and registered in IoT [S5], while DQ focuses on how good is the data collected by these sources. However, security problems could influence the quality of the collected IoT data. For example, adding noise to the data for hiding the user’s actual location could reduce data accuracy [29]. We thus argue that although related, security and DQ are two different indicators for IoT.

The DQ problems identified in this study were diverse and these problems overlapped among different DQ dimensions as presented in Sect. 4.2. A few studies suggested that DQ problems could occur in different layers of the IoT structure [9, 30]. As IoT is facilitating the development of new management models and business models based on IoT data [31, 32], this could call for higher DQ requirements. To better study and understand DQ problems and challenges in IoT, the five-layers of the IoT structure [32] that describe features and functions of the IoT, is adopted. Table 4 maps the DQ dimensions identified in this review to the layers of the IoT structure [32]. As shown in Table 4, at the device layer, sensor devices detect data that should accurately, timely, and completely represent real-world situations of an object. The detected data is then transmitted to the middleware layer via the network layer, which should deal with data loss (DQ problems on completeness and utility) and data distortion (DQ problems on accuracy and data volume) in this process. When the data arrives at the middleware layer, a number of DQ dimensions need to be considered, including accuracy, timeliness, completeness, and utility, in order to manage device services and maximize DQ. Thereafter this processed data serves as an asset for various applications, relying on users’ requirements related to accuracy, timeliness, completeness, utility, and concordance. While few studies have been conducted to address DQ at the business layer in this review. The challenges of addressing DQ for each layer of IoT structure are further discussed as below.

The *business layer* is responsible for the overall management of applications and services, enabling users to determine a future action and business strategy based on the processed data from the application layer [32]. In this review, few studies investigated DQ problems in the business layer. As IoT is facilitating the development of new management models and business models, this could facilitate the adoption of IoT in organizations to improve their competitiveness based on IoT products and services [31]. An investigation into DQ requirements at this layer and how DQ in IoT impacts

Table 4 Links between DQ dimensions and IoT structure in this review

Layer of IoT structure	DQ dimension				Concordance	Potential areas for further exploration
	Accuracy	Timeliness	Completeness	Utility		
Business layer	-	-	-	-	-	Investigating DQ requirements for the business layer; and measuring the impacts of DQ in IoT on business decision making
Application layer	[S8, S10, S13, S17, S18, S19, S23, S26, S29, S31, S43, S45]	[S17, S23, S29]	[S11, S13, S17, S23, S39]	[S33]	[S8]	Examining relationships between the factors influencing DQ in IoT applications; and measuring the impacts of DQ on IoT applications
Middleware layer	[S4, S5, S7, S12, S15, S24, S29, S30, S31, S42, S44]	[S5, S21, S22, S29]	[S5]	[S9, S33]		Maximizing DQ and minimizing energy consumption; and developing reputation evaluation mechanism for high-quality data provided by providers
Network layer	[S3]		[S25]	[S14]	[S3]	Selecting appropriate network groups to address DQ in data transmission
Device layer	[S1, S2, S6, S8, S10, S13, S17, S18, S19, S23, S26, S29, S31, S43, S45]	[S17, S23, S29]	[S2, S11, S13, S17, S23, S39]			Evaluating relationships between deployment cost and types of sensors for addressing DQ

business decision making will be in the scope of future research. This could improve the awareness of the importance of addressing DQ in IoT.

The *application layer* deals with smart management of the application based on the processed data in the middle layer [26]. The focus of the DQ mission at this layer is on providing quality-assured data (e.g. accurate data) that captures the IoT environment and individual interaction with it for smart use, such as indicating free traffic flows [S8], figuring out sensor faults and events [S13, S18, S26], and providing assistance to the disabled and elderly people in their life activities [S10, S19, S39, S43]. The reviewed studies have investigated the factors that influence DQ in IoT applications, however, none of them examined possible relationships between these factors that could explain the underlying mechanisms of achieving high-quality data, or measured the impacts of DQ on these applications that could allow users to realize the role of DQ in smart use.

The *middleware layer* addresses connection and communication between multiple devices that have the same service type, dealing with data storage and decision making on service management [32]. Because the increasing number of Internet connected things create a large amount of traffic and require much more data storage [31,32], this layer is responsible for more complex data management, covering a larger number of DQ dimensions to ensure DQ before use of the data. To address DQ in this layer, for each incoming data stream, DQ measurement has been computed into frameworks or architectures to monitor DQ and filter good data from the large amount of the collected data. Based on the methods used to measure DQ summarized in Sect. 4.3, multiple methods of measuring DQ in IoT could be computed for different purposes. Furthermore, poor-quality data identified from DQ measurement could be recovered [S11, S36, S38, S39]. Although automatic DQ measurement could preserve DQ to some degree, two challenges in achieving DQ in this layer exist as presented below.

First, because data communication requires node battery energy and network loading [S3, S22] [33], the tradeoff exists between DQ and energy consumption. The more intense the data processing (e.g. data compression for video data) [S3] and frequent updates from the node [33], the larger energy consumption of sensors and systems. However, sensors and wireless devices are battery-constrained. When consumers request a high requirement for DQ and data sharing over the network, this significantly challenges the energy consumption of the IoT devices [S12]. Thus, maximizing DQ and minimizing energy consumption will be research hot spots in IoT.

Second, in an IoT-based environment, providers can sense and share their local data. However, these activities are commonly motivated with a sufficient reward [34]. For instance, when a Wi-Fi network is not available before the deadline of uploading the data, users will decide whether to participate in these activities based on the reward. If the reward is small, providers will only upload the data when they have access to Wi-Fi networks free of charge. If the reward is large, users will upload the data through the cellular network before the deadline. The data requesters post their tasks to a platform and these tasks are further assigned to providers who provide the required data to the platform [S20]. A challenge arises in maximizing the quality of the data received by requesters with a boundary of shared budget for performing the activities of providing data. A reputation evaluation mechanism for the provider who continuously provides

good data that facilitates the high-quality data shared in IoT will be a future direction for research and practice [S20, S35, S42].

The *network layer* of the IoT structure places emphasis on data transmission from sensor devices to the information processing system [32]. Accuracy and data volume have been used to determine whether the data can be successfully transmitted to the destination via the network. Due to unpredictable node movement in networks (e.g. Mobile Wireless Sensor Network), data packets could drop resulting in data loss or delay during transmission [33]. The path selection for ensuring the link quality between each pair of nodes plays an important role in the success of packets delivery. Furthermore, different network groups (e.g. Wireless Personal Area Network) have their own limitations for data transmission [4]. An appropriate network group or the coexistence of multiple network groups will need to be selected to address DQ during data transmission for purposes.

The *device layer* of the IoT structure deals with the identification and collection of an object's information by sensor devices [32]. Accuracy, timeliness, completeness and concordance have been used as indicators to determine whether the sensor devices could provide high-quality data. If data errors or anomalies occur at this layer, this could further corrupt the quality of the data transmitted to upper layers in the IoT structure. Thus, choosing appropriate sensor devices and locations of sensors could be the focus of this layer. The current literature suggests that smartphones could provide better data than Xsens-like devices and bespoke sensor devices [S19]. Because IoT sensor devices have different capabilities for data collection, the limitations of these devices (e.g. sensing range) could affect the accuracy of the collected data [S1, S32]. However, the selection of sensor devices and their locations are limited by a deployment budget. The evaluation of relationships between deployment cost and types of sensors (and/or locations of sensors) in achieving complete and accurate data will become an important theme in IoT research.

5.3 Methods used to measure DQ in IoT

Findings relating to RQ4: The methods used to measure DQ in IoT identified are: (1) Measurement between techniques, sources or defined attributes, (2) Measurement with a reference, (3) Devices or algorithms validation, (4) Measurement within time intervals, (5) Measurement of presence, (6) Process observation, and (7) Log files review.

As shown in Table 3, data accuracy was measured by looking at agreements between elements among different datasets collected in IoT [S5, S6, S10, S18, S19, S23, S29, S30, S31, S43, S45], or comparing with a reference [S1, S6, S7, S8, S10, S12, S13, S15, S17, S18, S24, S42, S44, S45]. Some verified whether the expected values present in the collected IoT dataset [S4, S7, S8, S13, S18, S26]. A few studies measured whether IoT devices provide readings for an object during a fixed time interval [S1, S7, S31]. Only one study in this review identified data errors in situ monitoring and reviewing log files of the data detected by IoT devices [S2]. While timeliness could be measured by examining whether data elements were timely collected by different sources [S5, S23, S29] or whether data elements were collected by IoT devices within a reasonable

period of time [S17, S21, S22]. As for completeness, a number of studies measured whether or not desired data presents in the collected IoT dataset [S11, S13, S17, S25, S39]; a set of articles looked at variances among the observations for the same object from different sources [S5, S23, S25]; and others referred to prior literature to see whether or not there are agreements between data elements presented in the collected IoT dataset [S13]. In terms of utility, a few studies measured how much IoT data can be accessed by different techniques [S14] or from different data sources [S9, S33], and meantime the utility of data was checked at a desired time of interest [S9]. Data volume was measured by counting the number of data components transmitted to the destination based on defined IoT devices [S3]. While researchers checked whether or not there are agreements between data elements within the IoT datasets collected by different sources to measure the concordance [S8].

We found that most measurements of DQ were instantiated in a ratio scale (the number of actual data items which have related values was divided by the number of data items defined in the IoT context). This finding is similar with ISO 25024 DQ measurement in general [35]. Furthermore, a number of types of DQ measurement are observed in this review. To measure the quality of IoT data, some research adopted process observation and log files review, enabling the further identification of the causes of bad data [S2]. However, these methods may be time consuming. Some studies measured DQ within a desired period of time, while setting up an appropriate time interval is a difficult task as sensors may miss some important information of an object [S17, S22]. Others determined DQ by measuring the presence of data elements in the collected IoT dataset (e.g. [S13]). Although it is easy to calculate the presences, inappropriate measurement of missing data (e.g. considering zero as missing data for temperature) may underestimate DQ. A few studies conducted a comparative evaluation for measuring the quality of the IoT datasets collected by different techniques, sources or experiment settings (e.g. [S3]) to ascertain a better solution to achieve DQ. As IoT devices generate a large amount of data that could have noise and uncertainty [S9], without domain awareness of the data or any feedback from data users, the quality of the collected data is difficult to measure [S8]. To end this, the spatial-temporal correlated measured values for the objects provided by the node and its neighbors were proposed to ascertain the DQ [S7, S8, S14, S15, S18, S45]. While some researchers raised the concern that it is hard to find two data streams updated for the same or nearby real-world attribute for this checking [S15]. Some researchers used the algorithm's validation to learn the pattern of the data at hand to determine its DQ [S4, S7, S8, S13, S18, S26], while a challenge could arise in minimizing the computational complexity of the algorithms. Because DQ is defined as fitness for use [15], a certain method of quality measurement for a given IoT dataset may not be applicable in other datasets or other uses. A quality measurement framework for IoT data therefore plays an important role in suggesting a unified process of these DQ measurements. Such a framework could contribute to appropriately determining DQ in IoT and supporting the decision-making about how to use the data.

5.4 Implications

This study provides theoretical and practical implications as outlined below.

(1) *Theoretical implications* Theoretical contributions of this review include a unique study on the identification of methods used to measure DQ in IoT and the establishment of the links between dimensions, manifestations of DQ problems and methods used to measure DQ, which could further the understanding of this phenomenon. The alternative terms used to describe DQ dimensions that were identified can be added into the keywords list for bibliographic search. The ISO standardized definitions for DQ dimensions in general [22] could facilitate the consistent use of the terms used in the literature to describe DQ dimensions in the realm of IoT, while users could also benefit from the definitions of DQ dimensions summarized from the existing literature in Table 2 to define their DQ in the IoT context for different purposes. The strengths and limitations outlined for the methods used to measure DQ in Sect. 5.3 could facilitate the improvement of the methods for measuring DQ in IoT.

Furthermore, we have classified the research themes of DQ that have been studied in the IoT context into four main groups (i.e. definition, measurement, analysis, and design and development), that could help scholars ascertain research topics when studying DQ in IoT. Based on the data analysis, we also identify potential areas for further investigation: (1) developing guidelines or checklists for defining specific DQ dimensions of IoT data (see Sect. 5.1), (2) addressing DQ problems based on different layers of IoT structure as outlined in Table 4 (see Sect. 5.2), and (3) constructing DQ frameworks in the IoT context through combing literature reviews, field studies and expert reviews (see Sect. 5.3).

(2) *Practical implications* Our findings of the DQ dimensions and the methods used to measure DQ identified in this review will be of interest to IoT architects who are developing IoT systems in organizations. A good understanding of DQ dimensions contributes to developing the IoT products and services that meet the customers' requirements. The methods used to measure DQ identified in this study could offer some candidates for measuring DQ in IoT based on a given task.

For different layers of IoT structure, the IoT architects should take layers of IoT structures into account when looking at DQ problems. For example, at the device layer, due to limited memory, power, and capacity of IoT devices, deployment of these devices (e.g. selection and location of different types of sensor devices) plays an essential role in addressing DQ. At the network layer, appropriate network groups selected for addressing packets delivery could help ensure DQ in data transmission. Monitoring DQ for coming data streams and dealing with the tradeoff between DQ and energy consumption to make decisions on IoT service management should receive attention at the middleware layer, and multiple DQ measurements computed into this technical architecture would facilitate the monitoring. As for both application and business layers, the IoT data collected heavily rely on users' requirements for smart use and business decision making, and the quality of collected IoT data could be further addressed using data cleaning techniques.

Additionally, the manifestations of DQ problems that were identified is of relevance to those IoT engineers who are maintaining IoT systems and services. When the engineers observe a specific manifestation, they could trace to a certain DQ problem under one or more than one dimension, which could assist in the task of determining a core area (e.g. a specific layer of IoT structure or specific IoT components) to address the problem.

5.5 Limitations

Although we have consistently followed a search and selection strategy using the guidelines of Wolfswinkel et al. [12] and Kitchenham et al. [13] to ensure the completeness of our sample, there may still be some research studies that have not been included in our data collection. Firstly, the final review process was limited to the six specific online databases as advised by [19] and using a restricted keyword set as advised by [19,20]. There could potentially be articles, such as in different languages, not included in the databases or identified by using our keywords. However, these are the main sources for academic studies of IoT and there is a high confidence that the key literature has been identified. Secondly, we were only interested in empirical studies that investigated DQ in IoT, so we might have underestimated the current state of investigation of DQ in IoT. Thirdly, the articles screened were based on our inclusion and exclusion criteria and the data extracted from the included studies was based on our RQs. Hence, the results of articles selected together with data analysis were subjective. The original authors of the primary studies may not agree with our interpretation. However, to ensure research quality the selected studies have been coded, reviewed and evaluated by multiple researchers. For example, one co-author randomly selected and reviewed some papers from the data sample, to compare with the results of the main coder. The results showed a high level of consistency (100%) on the paper selection between the two coders. Furthermore, the data extraction table assisted in consolidating and better understanding the content of studies and to coordinate the analysis and synthesis of the required data from the literature.

6 Conclusion

This paper presents a systematic review of empirical studies of DQ in IoT. Our SLR answers four specific RQs based on the data extracted from 45 research publications from the emergence of the term IoT in 1999 [14] to 2018. The study provides a novel synthesis leading to a categorization of research themes (Sect. 4.1), dimensions and manifestations of DQ problems (Sect. 4.2), and methods used to measure DQ (Sect. 4.3). The findings suggest future directions for research and practice such as developing guidelines for defining specific DQ dimensions and DQ measurement frameworks for IoT data, as well as addressing DQ problems based on different layers of IoT structure (Sect. 5.4).

Appendix

List of articles included in the SLR:

- S1 Borges Neto JB, Silva TH, Assunção RM, Mini RA, Loureiro AA (2015) Sensing in the collaborative Internet of Things. *Sensors*, 15(3): 6607–6632
- S2 Hendrik Haan G, Hillegersberg JV, De Jong E, Sikkel K (2013) Adoption of wireless sensors in supply chains: A process view analysis of a pharmaceutical cold chain. *J Theory Appl Electron Commer Res*, 8(2): 138–154
- S3 Dmitriev A, Efremova E, Gerasimov MY (2015) Multimedia sensor networks based on ultrawideband chaotic radio pulses. *J Commun Technol Electron*, 60(4): 393–401
- S4 Gill S, Lee B, Neto E (2015) Context aware model-based cleaning of data streams. In: *Proc the 26th Irish Signals Syst Conf*, IEEE, pp 1–6
- S5 Sicari S, Rizzardi A, Miorandi D, Cappiello C, Coen-Portisini A (2016) A secure and quality-aware prototypical architecture for the Internet of Things. *Inform Syst*, 58: 43–55
- S6 Pravato L, Doyle TE (2017) IoT for remote wireless electrophysiological monitoring: Proof of concept. In: *Proc the 27th Annu Int Conf Compt Sci Softw Eng*, pp 254–258
- S7 Javed N, Wolf T (2012) Automated sensor verification using outlier detection in the Internet of Things. In: *Proc the 32nd Int Conf Distrib Comput Syst Workshop*, IEEE, pp 291–296
- S8 Kuemper D, Iggena T, Toenjes R, Pulvermueller E (2018) Valid. IoT: A framework for sensor data quality analysis and interpolation. In: *Proc the 9th Multimedia Syst Conf*, ACM, pp 294–303
- S9 Liono J, Jayaraman PP, Qin A, Nguyen T, Salim FD (2018) QDaS: Quality driven data summarisation for effective storage management in Internet of Things. *J Parallel Distrib Comput*. DOI: <https://doi.org/10.1016/j.jpdc.2018.03.013>.
- S10 Tariq M, Majeed H, Beg MO, Khan FA, Derhab A (2018) Accurate detection of sitting posture activities in a secure IoT based assisted living environment. *Future Gener Comp Syst*. DOI: <https://doi.org/10.1016/j.future.2018.02.013>.
- S11 Turabieh H, Salem AA, Abu-El-Rub N (2018) Dynamic L-RNN recovery of missing data in IoMT applications. *Future Gener Comp Syst*, 89: 575–583
- S12 Siegel JE, Kumar S, Sarma SE (2018) The future internet of things: Secure, efficient, and model-based. *IEEE Internet Things J*, 5(4): 2386–2398
- S13 Karkouch A, Mousannif H, Al Moatassime H, Noel T (2016) A model-driven architecture-based data quality management framework for the Internet of Things. In: *Proc the 2nd Int Conf Cloud Comput Technol Appl*, pp 252–259
- S14 Bijarbooneh FH, Du W, Ngai ECH, Fu X, Liu J (2016) Cloud-assisted data fusion and sensor selection for Internet of Things. *IEEE Internet Things J*, 3(3): 257–268
- S15 Sotres P, Santana JR, Sánchez L, Lanza J, Muñoz L (2017) Practical lessons from the deployment and management of a smart city Internet-of-Things infrastructure: The SmartSantander testbed case. *IEEE Access*, 5: 14309–14322
- S16 Guo Y, Fang L, Geng K, Yin L, Li F, Chen L (2018) Real-time data incentives for IoT searches. In: *Proc 2018 Int Conf Comm*, IEEE, pp 1–6

- S17 Li F, Nastic S, Dustdar S (2012) Data quality observation in pervasive environments. In: Proc the 15th Int Conf Comput Sci Eng, IEEE, pp 602–609
- S18 Nesa N, Ghosh T, Banerjee I (2018) Outlier detection in sensed data using statistical learning models for IoT. In: Proc 2018 Wireless Commun Netw Conf, IEEE, pp 1–6
- S19 Gupta M, Holloway C, Heravi BM, Hailes S (2015) A comparison between smartphone sensors and bespoke sensor devices for wheelchair accessibility studies. In: Proc the 10th Int Conf Intelligent Sensors, Sensor Netw Inform Process, IEEE, pp 1–6
- S20 Tao X, Song W (2018) Location-Dependent Task Allocation for Mobile Crowdsensing with Clustering Effect. IEEE Internet Things J. <https://doi.org/10.1109/JIOT.2018.2866973>.
- S21 Kothari A, Boddula V, Ramaswamy L, Abolhassani N (2014) DQS-cloud: A data quality-aware autonomic cloud for sensor services. In: Proc the 10th Int Conf Collaborative Compt: Netw Appl Worksharing, IEEE, pp 295–303
- S22 Candra ZM, Truong HL, Dustdar S (2016) On monitoring cyber-physical-social systems. In: Proc 2016 World Congr Serv, IEEE, pp 56–63
- S23 Leonardi A, Ziekow H, Strohbach M, Kikiras P (2016) Dealing with data quality in smart home environments—Lessons learned from a smart grid pilot. J Senser Actuator Netw, 5(1): 5
- S24 Chacko V, Bharati V (2017) Data validation and sensor life prediction layer on cloud for IoT. In: Proc 2017 Int Conf Internet Things, IEEE, pp 906–909
- S25 Atmoko R, Riantini R, Hasin M (2017) IoT real time data acquisition using MQTT protocol. J Phys: Conf Ser, 853(1): 012003
- S26 Alduais N, Abdullah J, Jamil A, Audah L, Alias R (2017) Sensor node data validation techniques for realtime IoT/WSN application. In: Proc 14th Int Multi-Conf Syst, Signals and Devices, pp 760–765
- S27 Balestrini M, Diez T, Marshall P, Gluhak A, Rogers Y (2015) IoT community technologies: Leaving users to their own devices or orchestration of engagement? EAI Endorsed Trans Internet Things, 1(1): e7
- S28 Ma Y, Jin J, Huang Q, Dan F (2018) Data preprocessing of agricultural IoT based on time series analysis. In: Proc Int Conf Intelligent Comput, pp 219–230
- S29 Bharti M, Saxena S, Kumar R (2017) Intelligent resource acquisition framework on Internet-of-Things. Compt Elect Eng, 58: 265–281
- S30 Jang B, Park S, Lee J, Han SG (2018) Three hierarchical levels of Big-Data market model over multiple data sources for Internet of Things. IEEE Access, 6: 31269–31280
- S31 Moon A, Kim J, Zhang J, Son SW (2018) Evaluating fidelity of lossy compression on spatiotemporal data from an IoT enabled smart farm. Comput Elect Agriculture, 154: 304–313
- S32 Gorenflo C, Golab L, Keshav S (2017) Managing Sensor Data Streams: Lessons Learned from the WeBike Project. In: Proc the 29th Int Conf Sci Statistical Database Manage, pp 1–11
- S33 Dong R, Ratliff LJ, Cárdenas AA, Ohlsson H, Sastry S (2018) Quantifying the utility–privacy tradeoff in the Internet of Things. ACM Trans Cyber-Physical Syst, 2(2): 1–28

- S34 Huang Z, Xie T, Zhu T, Wang J, Zhang Q (2016) Application-driven sensing data reconstruction and selection based on correlation mining and dynamic feedback. In Proc 2016 Int Conf Big Data, IEEE, pp 1322–1327
- S35 Gao Y, Li X, Li J, Gao Y (2017) A dynamic-trust-based recruitment framework for mobile crowd sensing. In: Proc 2017 Int Conf Commun, IEEE, pp 1–6
- S36 Fekade B, Maksymyuk T, Kyryk M, Jo M (2018) Probabilistic recovery of incomplete sensed data in IoT. *IEEE Internet Things J*, 5(4): 2282–2292
- S37 Sta HB (2017) Quality and the efficiency of data in “Smart-Cities”. *Future Gener Comp Syst*, 74: 409–416
- S38 Yan X, Xiong W, Hu L, Wang F, Zhao K (2015) Missing value imputation based on gaussian mixture model for the Internet of Things. *Mathematical Problems Eng*, 2015: 1–8
- S39 Mary IPS, Arockiam L (2017) Imputing the missing data in IoT based on the spatial and temporal correlation. In: Proc 2017 Int Conf Current Trends Advanced Compt, IEEE, pp 1–4
- S40 Gill S, Lee B (2015) A framework for distributed cleaning of data streams. *Procedia Compt Sci*, 52: 1186–1191
- S41 Zhang Y, Szabo C, Sheng QZ (2014) Cleaning environmental sensing data streams based on individual sensor reliability. In: Proc Int Conf Web Inform Syst Eng, pp 405–414
- S42 Pouryazdan M, Kantarci B, Soyata T, Foschini L, Song H (2017) Quantifying user reputation scores, data trustworthiness, and user incentives in mobile crowd-sensing. *IEEE Access*, 5: 1382–1397
- S43 Kos A, Tomažič S, Umek A (2016) Evaluation of smartphone inertial sensor performance for cross-platform mobile applications. *Sensors*, 16(4): 477–493
- S44 Casado-Vara R, de la Prieta F, Prieto J, Corchado JM (2018) Blockchain framework for IoT data quality via edge computing. In: Proc 1st Workshop Blockchain-enabled Netw Sensor Syst, pp 19–24
- S45 Ukil A, Bandyopadhyay S, Pal A (2015) IoT data compression: Sensor-agnostic approach. In: Proc 2015 Data Compression Conf, pp 303–312

References

1. Georgakopoulos D, Jayaraman PP (2016) Internet of Things: from internet scale sensing to smart services. *Computing* 98(10):1041–1058
2. Ramirez ARG, González-Carrasco I, Jasper GH, Lopez AL, Lopez-Cuadrado JL, García-Crespo A (2017) Towards human smart cities: Internet of Things for sensory impaired individuals. *Computing* 99(1):107–126
3. Elijah O, Rahman TA, Orikumhi I, Leow CY, Hindia MN (2018) An overview of Internet of Things (IoT) and data analytics in agriculture: benefits and challenges. *IEEE Internet Things J* 5(5):3758–3773
4. Liao Y, Loures EDFR, Deschamps F (2018) Industrial Internet of Things: a systematic literature review and insights. *IEEE Internet Things J* (to be published). <https://doi.org/10.1109/JIOT.2018.2834151>
5. Manyika J, Chui M, Bughin J, Dobbs R, Bisson P, Marrs A (2013) Disruptive technologies: advances that will transform life, business, and the global economy. McKinsey Global Institute, San Francisco
6. Minteer A (2017) Analytics for the Internet of Things (IoT): intelligent analytics for your intelligent devices. Packt Publishing, Birmingham

7. Fruehe J (2015) The Internet of Things is about data, not things. *Forbes*. <https://www.forbes.com/sites/moorinsights/2015/07/30/the-internet-of-things-is-about-data-not-things/3178bdd827cf>. Accessed 12 Nov 2018
8. Villasanta A (2019) Tesla model 3 autopilot feature to blame for death of driver in crash. *Int Bus Times*. <https://www.ibtimes.com/tesla-model-3-autopilot-feature-blame-death-driver-crash-2792690>. Accessed 3 June 2019
9. Karkouch A, Mousannif H, Moatassime HAI, Noel T (2016) Data quality in Internet of Things: a state-of-the-art survey. *J Netw Comput Appl* 73:57–81
10. Karkouch A, Moatassime HAI, Mousannif H, Noel T (2015) Data quality enhancement in Internet of Things environment. In: *Proceedings of the 12th international conference on computer systems applications*. ACS/IEEE, pp 1–8
11. Qin Y, Sheng QZ, Falkner NJ, Dustdar S, Wang H, Vasilakos AV (2016) When things matter: a survey on data-centric Internet of Things. *J Netw Comput Appl* 64:137–153
12. Wolfswinkel JF, Furtmueller E, Wilderom CP (2013) Using grounded theory as a method for rigorously reviewing literature. *Eur J Inform Syst* 22(1):45–55
13. Kitchenham BA, Budgen D, Brereton P (2015) *Evidence-based software engineering and systematic reviews*. CRC Press, Boca Raton
14. Ashton K (2009) That 'Internet of Things' thing in the real world, things matter more than ideas. *RFID J*. <https://www.rfidjournal.com/articles/view?4986>. Accessed 12 Nov 2018
15. De Feo JA, Juran JM (2017) *Juran's quality handbook: the complete guide to performance excellence*, 7th edn. McGraw-Hill Education, New York
16. Tilly R, Posegga O, Fischbach K, Schoder D (2017) Towards a conceptualization of data and information quality in social information systems. *Bus Inf Syst Eng* 59(1):3–21
17. Wang RY, Strong DM (1996) Beyond accuracy: what data quality means to data consumers. *J Manag Inf Syst* 12(4):5–33
18. Cai L, Zhu Y (2015) The challenges of data quality and data quality assessment in the big data era. *Data Sci J* 14:2. <https://doi.org/10.5334/dsj-2015-002>
19. Whitmore A, Agarwal A, Xu L (2015) The Internet of Things—a survey of topics and trends. *Inf Syst Front* 17(2):261–274
20. Sadiq S, Yeganeh NK, Indulska M (2011) 20 years of data quality research: themes, trends and synergies. In: *Proceedings of the 22nd Australasian database conference*, pp 153–162
21. Wang RY (1998) A product perspective on total data quality management. *Commun ACM* 41(2):58–65
22. ISO 25000 Portal (2019) ISO/IEC 25012. <https://iso25000.com/index.php/en/iso-25000-standards/iso-25012>. ISO 25000 Standards. Accessed 27 May 2019
23. De Faria MLL, Cugnasca CE, Amazonas JRA (2018) Insights into IoT data and an innovative DWT-based technique to denoise sensor signals. *IEEE Sens J* 18(1):237–247
24. Subramaniam S, Palpanas T, Papadopoulos D, Kalogerakiand V, Gunopulos D (2006) Online outlier detection in sensor data using nonparametric models. In: *Proceedings of the 32nd international conference on very large data bases*, pp 187–198
25. Yoon I, Joung H, Lee J (2016) Zynq-based reconfigurable system for real-time edge detection of noisy video sequences. *J Sens*. <https://doi.org/10.1155/2016/2654059>
26. Spachos P, Song L, Plataniotis KN (2017) Wireless noise prevention for mobile agents in smart home. In: *Proceedings of the 2017 international conference on communities*. IEEE, pp 1–6
27. Adelantado F, Vilakaosana X, Tuset-Peiro P, Martinez B, MeliàSeguí J, Watteyne T (2017) Understanding the limits of LoRaWAN. *IEEE Commun Mag* 55(9):34–40
28. Priller P, Aldrian A, Ebner T (2014) Case study: from legacy to connectivity migrating industrial devices into the world of smart services. In: *Proceedings of the 2014 IEEE emerging technology and factory automation (ETFA)*, pp 1–8
29. Olteanu AM, Huguenin K, Shokri R, Humbert M, Hubaux JP (2017) Quantifying interdependent privacy risks with location data. *IEEE Trans Mobile Comput* 16(3):829–842
30. Williams SP, Hardy CA, Nitschke P (2018) Configuring the Internet of Things (IoT): a review and implications for Big Data Analytics. In: *Proceedings of the 52nd Hawaii international conference on system sciences* (to be published)
31. Wu M, Lu TJ, Ling FY, Sun J, Du HY (2010) Research on the architecture of Internet of Things. In: *Proceedings of the 3rd international conference on advanced computer theory and engineering*, vol 5, pp 484–487

32. Khan R, Khan SU, Zaheer R, Khan S (2012) Future internet: the Internet of Things architecture, possible applications and key challenges. In: Proceedings of the 10th international conference on frontiers of information technology. IEEE, pp 257–260
33. Farzana AF, Neduncheliyan S (2017) Ant-based routing and QoS-effective data collection for mobile wireless sensor network. *Wirel Netw* 23(6):1697–1707
34. Cheung MH, Hou F, Huang J (2018) Delay-sensitive mobile crowdsensing: algorithm design and economics. *IEEE Trans Mobile Comput* 17(12):2761–2774
35. ISO (2019) ISO/IEC 25024:2015 Systems and software engineering—systems and software quality requirements and evaluation (SQuaRE)—measurement of data quality. <https://www.iso.org/standard/35749.html>. International Organization for Standardization. Accessed 31 May 2019

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.