IoT Data Quality Issues and Potential Solutions: A Literature Review

Taha Mansouri, School of Science, Engineering and Environment, University of Salford, Greater Manchester, UK. t.mansouri@salford.ac.uk

Mohammad Reza Sadeghi Moghadam, Department of Production and Operation Management, Faculty of Management, University of Tehran, Tehran, Iran. rezasadeghi@ut.ac.ir

Fatemeh Monshizadeh, Department of Production and Operation Management, Faculty of Management, University of Tehran, Tehran, Iran. fm.monshizadeh@ut.ac.ir

Ahad Zareravasan**,** Department of Corporate Economy, Faculty of Economics and Administration, Masaryk University, Brno, Czech Republic. Zare.ahad@mail.muni.cz

Abstract

In the Internet of Things (IoT), data gathered from dozens of devices, are the base for creating business value and developing new products and services. If data are of poor quality, decisions are likely to be non-sense. Data quality is crucial to gain business value of the IoT initiatives. This paper presents a systematic literature review (SLR) regarding IoT data quality from 2000 to 2020. We analyzed 58 articles to identify IoT data quality dimensions and issues and their categorizations. According to this analysis, we offer a classification of IoT data characterizations using the focus group method and clarify the link between dimensions and issues in each category. Manifesting a link between dimensions and issues in each category is incumbent, while this critical affair in extant categorizations is ignored. We also examine data security as an important data quality issue and suggest potential solutions to overcome IoT's security issues. The finding of this study proposes a new research discipline for additional examination for researchers and practitioners in determining data quality in the context of IoT.

Keywords: Data quality; Internet of Things (IoT); IoT Data quality dimensions; IoT Data quality issues; Literature review

1. Introduction

Internet of Things (IoT), the new form of cyber-physical infrastructure, integrates communication, analytics, and computation which is projected to change our lives [1]. IoT proposes connecting everyday items, such as watches, washing machines, vehicles, ovens, animals, plants, and the environment, to communicate with each other to ease our daily activities. The IoT paradigm began with Radio Frequency Identification (RFID). This term was mentioned by Kevin Ashton firstly in 1999 [2] and is widely utilized in diverse aspects like smart business and management [3], smart agriculture [4], smart transportation [5], smart health and medical [6], smart home [7, 8], smart environment and safety [9]. The US National Intelligence Council (NIC) estimates that by 2025, Internet connections may remain in things that we utilize every day, such as paper documents, furniture, and food package [10]. There are a variety of definitions for IoT from different points of view. It was represented as a group of smart things linked through RFID [11]. From a connection viewpoint, the IoT permits individuals to be connected anywhere and anytime with everything and everyone [12]. From a communication perspective, IoT relates to a worldwide network of objects linked to each other based on standard communication protocols [13]. From a data point of view, related smart things will become significant data producers and consumers instead of humans [14]. This aspect has numerous challenges, such as software and algorithms, performance, architecture, data quality, and hardware [15, 16]. This paper focuses on IoT's data analysis issues, and, to be more specific, data quality issues are taken into consideration.

Data is one of the worthiest assets in IoT as it is used to make new decisions, produce new goods, and enlarge markets. Many studies clarify the significance of data quality for data mining processes and the influence of low data quality upon the reliability of the outcomes [17, 18]. Data quality is defined as the adequacy of data to the purposes of the analysis [19], the degree to which a set of inherent characteristics fulfills the requirements [20], as well as how well data meets the requirements of consumers [14]. For IoT, furthermore, data quality means how appropriate the gathered data from smart things are [14]. The diversity of the sources and the volume of data result in new difficulties in the data quality field. It is important to note that practitioners and researchers aim to assess the "fitness for use" of data sets [21]. Data quality has become one of the significant aspects of IoT because poor decisions stem from a poor understanding of data quality. To do crucial tasks, each data user needs the utilized data to meet specific criteria. These Data Quality criteria are known as Data Quality Dimensions, such as accuracy, timeliness, completeness, and reliability [14]. Various literature contributions identified dimensions that affect the quality of data produced by the IoT, such as security, vulnerability, privacy preservation, accuracy, data volume, and confidence completeness. In the literature, data quality is classified into four main categories as Intrinsic, Contextual, Representational, and Accessibility [14, 22].

Various articles examined data quality dimensions, issues, and techniques to improve data quality in the context of IoT from different viewpoints [12, 14, 21, 23]. Some papers proposed a categorization scheme for IoT data quality [14, 22].

This study reviews existing contributions to recognize IoT data quality dimensions, IoT data quality issues related to dimensions, and existing categories. As our contributions, we propose a classification in which related IoT data quality dimensions and IoT data quality issues are identified. We also reviewed potential solutions for security issues as an essential issue in IoT. To reach our aims, we intend to answer the following questions:

1. What are the data quality dimensions in IoT?
2. What are the data quality issues in IoT?
3. What are the potential solutions for IoT security issues?

To answer these research questions, we review the most related research in section 2, followed by the research method in section 3. Then we summarize the review results regarding the IoT data quality dimensions, categories, and issues in section 4. We propose a categorization of IoT data quality categories and issues in section 5. Potential solutions for security issues are presented in section 6, and finally, section 7 concludes the paper.

2. Related Research

This section purposes of comparing the previous literature in the IoT data quality context. In particular, in section 2.1, we present kinds of literature that identified IoT data quality dimensions and issues. We also present IoT data quality categories and potential solutions to overcome security problems with the approach proposed in this paper in sub-section 2.2.

2.1. Contributions in the IoT data quality dimensions and issues

Various studies have investigated several different IoT data characteristics, and some have also identified their quality dimensions. A review of data quality in IoT provides a broad survey of the quality metrics. Data quality has been explained in various ways in the literature. Apart from the broadest utilized definition, "fitness for use" is defined as the data that is fit for use by data consumers. Usefulness and usability, therefore, are significant features of quality [24]. Data quality examines the essential characteristics of IoT data and gives a classification of the quality dimensions. Data quality dimensions dated back to the 1990s and were often utilized through information system specialists [25] and provide an excellent way to assess data quality. Although some authors have defined various data quality dimensions, there is no standard definition of data quality dimensions [26]. Furthermore, different domain-specific data quality dimensions have been defined for multiple particular applications [27]. To keep track of data quality and measure them efficiently, many characteristics as data quality dimensions have been addressed.

Metzger et al. [28] addressed accuracy, timeliness, and trustworthiness. The authors proposed anomaly detection techniques to remove inaccurate and noise data to enhance data quality. Another contribution claimed that validity, accuracy, and credibility are the primary data quality dimensions in IoT [29]. Sicari et al. [21] defined data quality dimensions that can be automatically assessed, including completeness, accuracy, timeliness, and source reputation. Ghallab et al. [11] focused on IoT's data analysis issues, particularly data quality issues, such as uncertainty, noise, outliers, inconsistency, and missing value. This paper proposes challenges that come from outliers. Barnaghi and Sheth [30] identified information accuracy, validity, and credibility as the key data quality dimensions to control data sources. Klein and Lehner [31] have utilized five dimensions (accuracy, confidence, completeness, data volume, and timeliness) to assess sensor data streams' quality. Qin et al. [13] have focused on uncertainty, redundancy, ambiguity, and inconsistency as the direct dimensions. Completeness, accuracy, format, and currency were considered data quality dimensions in the IoT context [12]. Togneri et al. [32] addressed data availability and veracity as data quality issues in the context of IoT. They divided data quality issues into availability (error and interruption) and veracity (unbalanced and non-correspondence of different granularity data) problems. Farooghi et al. [33] identified IoT data quality issues, including timeliness, accuracy, completeness, usability, trustworthiness, confidence, consistency, and readability. Accuracy, completeness, usability, trustworthiness, consistency, readability, accessibility, and redundancy are concerned as IoT data quality issues by authors in [34]. Barnaghi et al. [35] mentioned precision, accuracy, and granularity as issues that stem from data characteristics. Liu and colleagues [1] considered accuracy, timeliness, completeness, utility, data volume, and concordance as data quality dimensions. They also paid attention to errors, dirty data, outliers, noise as data quality problems.

2.2. Main contributions in the IoT data quality categories

Several literature contributions identified data quality dimensions, most of which are classified into four main categories. Karkouch et al. [14] addressed quality issues, RFID data, and data streams. They presented accuracy, confidence, completeness, data volume, ease of access, access security, interpretability, and timeliness as the primary data quality dimensions. They also considered additional data quality dimensions such as duplication (e-health and smart grids domain) and availability (e-health domain) as the IoT domain-specific. They identified four main categories of data quality dimensions: Intrinsic, Contextual, Representational, and Accessibility. This article reviewed data quality in IoT and introduced generic and specific data quality dimensions. Moreover, they investigated IoT-related issues affecting data quality, such as data outliers, duplication, and data leakage. Although this contribution studied IoT data quality dimensions and issues in-depth and recognized four primary categories, they did not separate related data quality dimensions and issues in the category.

Authors in [22], [24], and [36] have summarized data quality dimensions into four main categories, and splitting dimensions and issues in the related category did not concern. All four categories are gathered in table 1.

Table 1: Data quality categories and dimensions

Data Quality Categories	***Data Quality Dimensions***	***References***
Intrinsic	Accuracy, Reputation	[14]
Contextual	Timelines, Completeness, Data volume	
Representational	Interpretability, Ease of understanding	
Accessibility	Accessibility, Access Security	
Intrinsic	Accuracy, Objectivity, Believability, Reputation	[22, 24]
Accessibility	Accessibility, Access security	
Contextual	Relevancy, Value-added, Timeliness, Completeness, Amount of data	
Representational	Interpretability, Ease of understanding, Concise representation, Consistent representation	
Intrinsic	Accuracy, Believability, Objectivity, Reputation	[36]
Contextual	Appropriate amount of data, Completeness, Relevancy, Timeliness, Value-added	
Representational	Concise representation, Ease of manipulation, Interpretability, Representation consistency	
Accessibility	Accessibility, Access security	

We examine existing categories of data quality dimensions in literature and their definitions.

- Intrinsic: This category has to do with the innate quality in data or data inherited, such as accuracy.
- Contextual: Dimensions in this category describe the quality of tasks utilizing data like timeliness and completeness.
- Accessibility: This addresses the accessibility of data for data users.
- Representational: Dimensions describe how data formats are comprehensible and representative, such as ease of understanding.

Analyzing mentioned studies, it can be concluded that there are three exciting gaps. First of all, existing articles do not have a broad view of IoT data quality. In fact, they did not consider all data quality dimensions and issues. They only have focused on some of them. We gathered data quality dimensions and problems with their definitions. Secondly, it should be noted that there are some IoT data quality categories in which various dimensions were identified. However, none of them separated related data quality dimensions and issues. Developing a new category into five main categories and separating related dimensions and issues in each category is another study's contribution. Finally, we analyzed problems and suggested potential solutions for security issues.

3. Research Method

Data quality is one of the critical studies in IoT and attracted much researcher attentions. An astonishing variety of articles have been published about IoT data quality. To do this survey, we carried out a systematic literature review (SLR) of the empirical studies focusing on data quality in the context of the IoT. We identified IoT data quality dimensions and issues up to 2020. IoT data quality categories were recognized as well.

We follow the guidelines of [1, 37] and conduct four steps: (1) define the scope of the review; (2) search for a preliminary list of papers; (3) select relevant papers; and (4) investigate data from the included papers. Table 2 illustrates several steps to do processes in this paper.

3.1. Define the scope of the review

Four prime activities are included in this step, such as the formation of inclusion and exclusion standards of an article in the data set, recognition of relevant research fields, selecting databases, and formulation of search terms [1, 37].

3.1.1. Formation of inclusion and exclusion

The inclusion criteria selected in this study are: (1) papers published in English (IC1); (2) papers published from 2000 to 2020 inclusive (IC2); and (3) the proposition of the paper has a focus on data quality in the IoT context (IC3). The removal of papers is according to the following exclusion criteria: (1) the papers do not come up with empirical findings (EC1); (2) papers are not available online (EC2); or (3) papers are identical (EC3).

3.1.2. Recognition of relevant research fields

As we mentioned in the introduction, IoT has numerous challenges like software, hardware, performance, architecture, and data quality. Our study hence focuses on data quality in the context of IoT.

3.1.3. Selection of databases

This review utilized Scopus, which possesses many peer-reviewed and English papers on the related topic. To further enhance the sample's quality for data analysis, backward (i.e., utilizing the reference list to recognize up-to-date articles) and forward snowballing (i.e., detecting citations to the papers) methods were employed to access the data [1].

3.1.4. Formulation of search terms

Search keywords in this study included: "Internet of Things" and "data quality." We utilized "IoT" as an alternative keyword. Using these keywords provides us a wide range of publications in the context of IoT and data quality. Consequently, our search started including mentioned keywords through utilizing the Boolean operations. The search strings are as follows: ("IoT" OR "Internet of Things") AND ("data quality") in the chosen database.

3.2. Search for a preliminary list of papers

We selected the publications in Title, Keywords, and Abstract, applying the online databases to organize our search by the search strings. We also have a preliminary list of papers, as illustrated in table 2.

3.3. Sample

This step aims to filter the related papers for additional investigation. We recognized 325 articles, of which 209 were removed on abstract review according to our inclusion and exclusion standards. After a full-text review, we extra abandoned 83 papers based on EC2. In the snowballing process, we identified 28 papers additionally which 25 papers remained after the abstract and full-text review, according to inclusion and exclusion standards. Subsequently, a total of 58 papers remained acceptable for examination (see Appendix I for the list of sources and Table 2 for the research process) [1].

3.4. Investigate data from the included papers

This section manifests our findings obtained from the reviewed contributions based upon our Research Questions. Section 4 reviews data quality contributions in the context of the IoT, covering different definitions of data quality, data quality dimensions and their explanation, data quality issues, and their explanation, as well as data quality categories. Analyzing existing data quality dimensions, issues, and categories, we develop a new category in which each issue has been identified to related dimensions.

Table 2: Research process

Step 1: Define the scope of the review	**Search strings**: ("IoT" OR "Internet of Things") AND ("data quality") **Databases**: Scopus: 325
Step 2: search for a preliminary list of papers	Number of the papers: 325
Step 3: select relevant papers	Analyze papers according to their abstracts 209 papers excluded based on EC1(167), EC3(42)
	Analyze papers according to their full-text 83 papers were excluded based on EC2(83), and 25 were included after snowballing
Step 4: investigate data from the included papers	Extract IoT data quality dimensions, issues, and categories from the included papers (IC3(33) and snowballing (25), (N=58))
	Develop new category Suggest solutions to overcome security issues

4. Findings

Numerous features are usually correlated with data in the context of the IoT. Whereas some of these features might be viewed as omnipresent, other features are not general and significantly depend on the context. Below, we have gathered data quality dimensions and issues, along with their definitions. It is worth considering that there is no specific definition for each dimension and issue, and everyone has various definitions.

4.1. IoT data quality dimensions

Accuracy measures the closeness of captured values of data points with their original values; in other words, accuracy is closely connected with correctness [21, 38-40]. **Usability** regards the amount of time data can be maintained before it comes devoid of value [33]. **Privacy** assures limits on a person who is permitted to access the data [14, 22]. **Confidence** demonstrates the maximum probability of the expected statistical error to occur [14, 40]. Along with privacy and confidence, **Security** is a much broader concept meaning the attempt to protect the privacy, confidentiality, and integrity of sensor data [14]. **Currency** is the user's perception of the degree of data that is up to date [41]; in other words, concepts such as **Timeliness**, latency, currency, or volatility display freshness of captured data and their punctuality regarding the application context [21, 38-40]. **Validity** or freshness is the period in which the reading value is still valid in the original sensor during reporting [38]. **Completeness** defines the extends to which values are available in a dataset [39]. Moreover, according to the application requirements, this measure has a close relationship with non-imputed values in a data stream [40]. **Trustworthiness** is associated with the concept of source reputation and reliability and defines whether the sensor feed was collected and processed by genuine infrastructures [38]. **Availability** refers to the amount of time a sensor feed is operational and available for use [38, 42]. **Ease of access** shows how easy data retrieval is [14]; unlike availability, this dimension cops with the amount of required preprocessing. **Volume (Throughput)** refers to the amount of raw data expected to be processed to reach the target information [14, 40]. **Frequency** is a close concept to volume even though it shows the temporal resolution and the reading rate regarding a specific time, such as reading among 2 minutes [38]. Moreover, **Capacity** is related to concurrent access and shows the maximum extent of concurrency, not just the volume of incoming data [39]. **Interpretability** means data has a meaningful and easy to interpret schema [14]. The **granularity** is a spatial and temporal dimension that measures how detailed are the stored data or, in other words, the level of abstraction in the stored data. Although different applications have different requirements for the level of granularity, it directly affects some other qualitative dimensions such as timeliness and completeness. Moreover, using some interpolation models such as linear, polynomial interpolation, and Gaussian is beneficial to reducing data granularity. While more detailed data is requested, the source can be used, and while less detailed is needed, the application would access to sampled or aggregated data [35]. **Format** refers to the user's perception of how well the information is presented. We gathered all reported existing data quality dimensions in table 3.

Table 3: IoT data quality dimensions

Dimensions	***References***
Accuracy	[1], [12], [14], [21], [22], [25], [28], [29], [33], [35], [36], [39], [40], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51], [52], [53], [54], [55], [56], [57], [58]
Precision	[1], [13], [33], [35], [43], [45], [47], [48], [55], [59], [60]
Usability	[1], [33], [45], [57],
Relevance	[1], [22], [25], [54], [56], [57], [61], [62]

Believability	[22], [56], [57], [61]
Ease of Understanding	[22], [61]
Privacy	[43], [62], [63], [64], [65], [66], [67], [68], [69], [70],
Objectivity	[56], [57],
Reputation	[22], [57], [61]
Granularity	[35]
Integrity	[54]
Currency	[1], [12], [41], [57]
Completeness	[1], [12] ,[14], [21], [22], [33], [35], [39], [40], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51], [54], [56], [57], [58], [61], [59],
Timeliness	[1], [12], [14], [21], [22], [28], [33], [35], [38], [39], [40], [43], [44], [45], [46], [47], [48], [50], [51], [52], [54], [56], [57], [58], [61], [71],
Trustworthiness	[14], [21], [28], [29], [33], [35], [38], [45], [52], [71], [72]
Availability	[1], [14], [32], [33], [38], [42], [43], [44], [52], [54], [73], [74], [75]
Security	[14], [21], [22], [55], [56], [57], [61], [63], [71], [76]
Validity (freshness)	[1], [38], [52], [54], [57], [75]
Frequency (temporal resolution)	[1], [33], [38], [45]
Confidence	[14], [33], [40], [45], [48]
Volume (Throughput)	[14], [22], [39], [40], [48], [57], [61]
Ease of access	[14], [21], [56], [57]
Interpretability	[14], [22], [56], [57], [61], [73]
Capacity	[39]
Format	[12], [58]

4.2. IoT data quality issues

In the context of the IoT, data are vulnerable to numerous issues that affect their quality. As another contribution of this research, we present IoT data quality issues, as summarized in table 4.

Redundancy or duplicate happens when several readings contain the same data point; due to either several readings of one sensor or some readings of multiple sensors [13, 73]. Data values can be **faulty** from defective sensors or real outliers. For instance, an exponential increase in a smart house's temperature can show a fire crisis [54]. To this end, we can mention a much broader concept, **Error** which occurs owing to sensor defeats generated by severe environments, age, or network data exploitation. Errors may also happen due to the sensor's faulty installation or utilization outside the various credible operations. The error indicates making any mistake in capturing or processing the stream of data, for instance, connecting data in the different year [14, 54]. Sometimes, related information might be vague to an application owing to a **lack of context**. It can be solved partially by employing semantic information about the data. Nevertheless, the contextual data is often unstructured, hence more challenging to interpret automatically [77]. Another viewpoint arises from the mismatch between the provider and user of data [78]. When the data value displays any random deviation from the expected range of the given data point, the

outlier error likely happened [14]. These kinds of issues are hard to find. Although they may be false values, they may be rare, containing invaluable information about the application [38]. **Precision** is related to any noise in converting measures to each other; for instance, the voltage to conversion to quantities such as temperature [79]. **Inconsistency** is also popular in IoT data [80], and it occurs when multiple sensors measure the same phenomena whereas they depict different results [81]. **Uncertainty** can refer to readings of the non-existing devices [73]. In some situations, the reading might be considered different measures by different applications; this requirement may lead to **ambiguity** in the interpretation of data [13]. **Response time** is the average time to process the readings entirely [73]. It is mainly caused by network latency among sensors and applications, plus the amount of processing to turn sensor readings to the appropriate format quickly interpreting by dashboards.

Table 4: IoT data quality issues

Issues	***References***
Duplicate	[14], [51], [57], [77], [78]
Noisy data	[1], [11], [80]
Errors	[1], [32], [42], [52], [55], [61], [59], [79]
Lake of Context	[80]
Outliers	[1], [11], [14], [53], [55], [82], [83], [84]
Response time	[39]
Inconsistency	[11], [13], [33], [39], [43], [44], [51], [54], [59], [57], [73]
Redundancy	[13], [14], [33], [43], [51], [59], [73]
Uncertainty	[11], [13], [21], [59], [73], [85]
Ambiguity	[13], [43], [73]

5. Proposed category

As mentioned in section 2.2, a few data quality categories have been developed in the context of IoT. In section four, we reviewed existing IoT data quality dimensions and issues. Concerning analysis, we found some valuable contributions. Karkouch et al. [14] introduced a classification into four categories: intrinsic, contextual, representation, and accessibility. They have focused on nine data quality dimensions. Lee et al. [22] developed a classification that includes four categories. Although there is a similarity between Lee's category and Karkouch's category regarding the number and kind of classifications, Lee's category includes 15 data quality dimensions. His category entails new dimensions compared with Karkouch's category. Liu and colleagues [1] have gathered IoT data quality dimensions and related problems in forming six categories (Table 5).

Table 5: Data quality dimensions and problems [1]

IoT Data quality categories	***IoT Data quality dimensions***	***IoT Data quality problems***
Accuracy	Precision, Validity, Correctness	Measurement errors, Dirty data, Outliers, Noise, Data frame distortion
Timeliness	Currency, Volatility, Latency, Freshness, Data rate, Delay, Frequency	Missing updates, Low data rate
Completeness	Availability, missing data	Missing data
Utility	Usage, Frequency, Relevancy	Noise, Data loss, missing data
Data Volume	------	Data loss, Delay data transmission, Data frame distortion
Concordance	----------	Irregular observation

With regard to existing categories, it can be concluded that data quality dimensions and issues are not separated in existing categories. Moreover, all of them have focused on a few dimensions and issues, while various data quality dimensions and issues would be valuable to pay attention to them. Moreover, there is not a single and principal category. Every category has been developed from various viewpoints, which is included different dimensions and problems.

Regarding all dimensions, issues, and categories in IoT, we concluded that developing new classification and identifying related dimensions and issues in each category is beneficial. Hence, we expanded existing categories and proposed a new classification into five categories. To do so, we used the focus group method. A Focus group is a data collection method that cooperates within a group to extract valuable experiential data. To be more specific, the focus group is six to 12 individuals who are alike in some way and come together to debate an issue [86]. The standard of using the focus group method is based upon expert knowledge. Six IoT experts were involved in designing a classification based on their knowledge and experiences. First of all, after analyzing IoT data quality dimensions, issues, and existing categories, experts concluded that developing a classification of five categories was adequate. They also recognized accuracy, confidence, trustworthiness, timeliness, and completeness as principal IoT data quality categories. Then, they identified related dimensions and issues in each category. Separating dimensions and issues from each other is vital while this important affair in existing categories is ignored. It is essential to note that the positioning of every dimension and issue were changed several times by experts. Finally, they found an appropriate classification to the best of their knowledge (table 6). The proposed category includes 19 IoT data quality dimensions and 13 IoT data quality issues.

Table 6: Proposed categorization of IoT data quality categories, dimensions, and issues

IoT Data quality categories	***IoT Data quality dimensions***	***IoT Data quality issues***
Accuracy	Objectivity, Precision	Noisy, Error, Outlier
Confidence	Relevance, Ease of understanding, Interpretability, Format, Granularity	Uncertainty, Ambiguity, Lack of context
Trustworthiness	Reputation, Privacy, Security, Integrity	Insecurity, Source ambiguity, Inconsistency
Timeliness	Validity, Currency	High response time
Completeness	Usability, Ease of access, Availability, Throughput, Capacity, Frequency	Duplicate, Redundancy, Incompleteness

6. Potential solutions to IoT data quality issues

IoT relevant issues need to be addressed from the data quality aspect. Some problems arise from IoT infrastructure, and they are created due to the essence of data. As our final contribution, we identified some potential solutions for security issues from our review. We confine our proposed solutions to Artificial Intelligence (AI) and Blockchain, as we observed from the literature review that there are some potentials in these technologies to overcome security issues.

6.1. Using Artificial Intelligence to overcome security issues

Security and privacy issues are grown due to a considerable number of devices and a shortage of standardized surveys concerning IoT security. Thus, these kinds of studies should be concerned. The privacy and security issues exist in IoT infrastructure, which must be addressed to build trust between users. Addressing some issues such as privacy and security in IoT is already addressed in some prior research [87, 88]. AI plays a crucial role in IoT, since it can crunch data successfully to empower us to collect invaluable insights. Machine Learning is a sub-branch of AI and has enormous potential to distinguish the irregularities and patterns in smart sensors' data.

Some of the IoT applications require a decision that should be taken before the real event happens. For instance, anticipating the fire in a kitchen and alarm the sound to halt the fire. A practical framework is needed to process and measure tremendous data utilizing machine learning techniques [87]. Hossain et al. [89] reviewed security issues in terms of applying machine learning in a smart application. Moreover, the volume of the data generated by devices is vast because the number of devices connected to the network is enormous. Processing and performing computation create difficulties in an IoT environment. Therefore, artificial intelligence comes as a release with other rising technologies to solve IoT security issues. In fact, IoT and AI can link to improve system analysis, improve operational effectiveness, and improve the precision rate. Ghosh et al. [90] described that AI could help IoT tremendous volume, heterogeneous data, and unstructured data to calculate real-time and make the system realistic. Data outliers are also critical indications of data uncertainty. Outliers relate to the class of unreliable readings or those out of bounds without any particular reason. Outliers or anomaly detection as a class of machine learning techniques improve data sets' quality by making them more consistent. Furthermore, the results' accuracy and reliability are increased because outlier detection represents the first state for unreliable reading [87].

6.2. Using blockchain to overcome security issue

Blockchain was introduced with Bitcoin [60] to solve the double-spending problem. A blockchain is built utilizing cryptography. Its cryptographic hash identifies each block, and each block indicates the hash of the previous block. This creates a link among the blocks, forming a Blockchain.

Blockchain technology is a distributed/decentralized network where each of them is connected to others. A block comprises various genuine transactions and their associated qualities [60]. A new set of applications appear because blockchain proposes a safe value exchange among entities in the network. To empower IoT devices to conserve security and privacy, the authors [87] have suggested a Blockchain-connected Gateway for Bluetooth Low Energy (BLE). Personal user privacy is preserved as the gateway restricts users' sensitive data from being reached without their permission. Hosseini Bamakan et al. [91] also developed an integrated Blockchain-IoT-Big data enabled framework for evaluating service supply chain performance, tacking the issues of security, privacy and trust.

The second generation of Blockchain-enabled smart contracts has introduced new opportunities for data quality improvements. Concerning data quality in the IoT device layer, Cha et al. [92] proposed that since theoretically, the Blockchain data is an exact representation of the events that occurred in the real world, the data integrity and quality increase with the adoption of blockchain technology. In fact, the adoption of blockchain offers an automated means for creating, processing, storing, and sharing information in machine-to-machine communication. Utterly, smart contracts can reduce human errors and improve data accuracy, completeness, and accessibility through automated data creation and storage [92, 93]. On top of this, blockchain stored data in blocks are cryptographically sealed.

Meanwhile, stored data is irreversible and cannot be altered arbitrarily, denoting a high data security degree. In this line, Azaria et al. [94] proposed a Blockchain-based medical record management system that improved data quality. Casado et al. [93] proposed a Blockchain-based system to assess and enhance sensor data quality in an IoT platform. DoS attacks are also not a severe concern in Blockchain-based IoT platforms due to the distributed structure of blockchain. The blockchain has no single point of failure problem in this case [95] in the application layer; there is a threat of identity forgery attacks. An external attacker may try to use a fabricated signature of a node to forge its identity. Similarly, due to the use of a secure digital signature algorithm of the blockchain, such an attack will not work. Therefore, it can be argued that blockchain has enhanced different aspects of data quality in general and IoT in particular [92].

7. Conclusions

Nowadays, IoT is one of the most swiftly utilized technologies in multiple applications. Data is inspiring agriculture, healthy, manufacturing, and diverse business decisions. It is crucial to consider and evaluate data quality since poor choices are rooted in low data quality.

Various studies have been conducted regarding data quality in the context of the IoT. They examined data quality from different viewpoints, identified data quality dimensions and issues, and some of them have developed a category for data quality. Nevertheless, none of them separated data quality dimensions and issues into related categories. Moreover, positioning related dimensions and issues in each category are ignored in existing categories.

This study reviewed data quality in the context of IoT. We reviewed data quality dimensions (section 3) and data quality issues (section 4) in general and specific domains. Moreover, we reviewed existing data quality categories. Concerning existing IoT data quality dimensions, issues, and categories, we proposed a new classification of five categories and identified related dimensions (19 dimensions) and issues (13 issues) using the focus group method (section 5). This category would enormously aid the practitioners and researchers to recognize related dimensions and issues in each category to enhance the quality of data and mitigate issues in the context of IoT.

Data are vulnerable to manifold problems. Some studies have examined major problems about IoT data quality, such as data outliers [14], while security issues in IoT data quality are ignored. In response to another gap in the IoT literature in data quality, we discuss security as an essential issue and recognize their potential solutions (section 6). We found AI and Blockchain technology as potential solutions to overcome security and privacy issues.

The most critical limitation we encountered was the low number of IoT data quality categories. More importantly, there is no single and principal category with specified dimensions and issues, each categorization has been developed from a different point of view. Moreover, there are no distinct differences between dimensions and problems in the literature, and we try to develop a new category based on IoT experts' knowledge. However, there are some differences between our proposed category and existing categories. Lack of enough experience and implementation leads to not enough publication in several domains of IoT; therefore, we need to admit that IoT, as well as its applications and issues, data quality included, is still on its way to evolution; so, issues are raising whereas for some of them have not been able to come up with a specific solution stem from the literature. Accordingly, it is valuable for future researchers to tackle different aspects of data quality issues and relate them to IoT architectures and particular applications such as smart cities, manufacturing, and agriculture.

References

1. Liu, C., et al., *Data quality and the Internet of Things.* Computing, 2020. **102**(2): p. 573-599.
2. Ashton, K., *That 'internet of things' thing.* RFID journal, 2009. **22**(7): p. 97-114.
3. Delgosha, M.S., N. Hajiheydari, and M. Talafidaryani, *Discovering IoT implications in business and management: A computational thematic analysis.* Technovation, 2021: p. 102236.
4. Virk, A.L., et al., *Smart farming: an overview.* Smart Village Technology, 2020: p. 191-201.
5. Rathore, M.M., et al. *Efficient graph-oriented smart transportation using internet of things generated big data*. in *2015 11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*. 2015. IEEE.
6. Hajiheydari, N., M.S. Delgosha, and H. Olya, *Scepticism and resistance to IoMT in healthcare: Application of behavioural reasoning theory with configurational perspective.* Technological Forecasting and Social Change, 2021. **169**: p. 120807.

7. Dimakis, N., A. Filippoupolitis, and E. Gelenbe, *Distributed building evacuation simulator for smart emergency management.* The Computer Journal, 2010. **53**(9): p. 1384-1400.
8. Sovacool, B.K. and D.D.F. Del Rio, *Smart home technologies in Europe: a critical review of concepts, benefits, risks and policies.* Renewable and sustainable energy reviews, 2020. **120**: p. 109663.
9. Saeed, F., et al., *IoT-based intelligent modeling of smart home environment for fire prevention and safety.* Journal of Sensor and Actuator Networks, 2018. **7**(1): p. 11.
10. Atzori, L., A. Iera, and G. Morabito, *The internet of things: A survey.* Computer networks, 2010. **54**(15): p. 2787-2805.
11. Ghallab, H., H. Fahmy, and M. Nasr, *Detection outliers on internet of things using big data technology.* Egyptian Informatics Journal, 2020. **21**(3): p. 131-138.
12. Côrte-Real, N., P. Ruivo, and T. Oliveira, *Leveraging internet of things and big data analytics initiatives in European and American firms: Is data quality a way to extract business value?* Information & Management, 2020. **57**(1): p. 103141.
13. Qin, Y., et al., *When things matter: A survey on data-centric internet of things.* Journal of Network and Computer Applications, 2016. **64**: p. 137-153.
14. Karkouch, A., et al., *Data quality in internet of things: A state-of-the-art survey.* Journal of Network and Computer Applications, 2016. **73**: p. 57-81.
15. Kiruthika, J. and S. Khaddaj. *Software quality issues and challenges of internet of things*. in *2015 14th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)*. 2015. IEEE.
16. Shah, S.H. and I. Yaqoob. *A survey: Internet of Things (IOT) technologies, applications and challenges*. in *2016 IEEE Smart Energy Grid Engineering (SEGE)*. 2016. IEEE.
17. Berti-Equille, L., *Measuring and modelling data quality for quality-awareness in data mining*, in *Quality measures in data mining*. 2007, Springer. p. 101-126.
18. Hipp, J., U. Güntzer, and U. Grimmer. *Data Quality Mining-Making a Virute of Necessity*. in *DMKD*. 2001.
19. Merino, J., et al., *A data quality in use model for big data.* Future Generation Computer Systems, 2016. **63**: p. 123-130.
20. Banerjee, T. and A. Sheth, *Iot quality control for data and application needs.* IEEE Intelligent Systems, 2017. **32**(2): p. 68-73.
21. Sicari, S., et al., *A secure and quality-aware prototypical architecture for the Internet of Things.* Information Systems, 2016. **58**: p. 43-55.
22. Lee, Y.W., et al., *AIMQ: a methodology for information quality assessment.* Information & management, 2002. **40**(2): p. 133-146.
23. Karkouch, A., et al., *A model-driven framework for data quality management in the Internet of Things.* Journal of Ambient Intelligence and Humanized Computing, 2018. **9**(4): p. 977-998.
24. Strong, D.M., Y.W. Lee, and R.Y. Wang, *Data quality in context.* Communications of the ACM, 1997. **40**(5): p. 103-110.
25. Byabazaire, J., G. O'Hare, and D. Delaney, *Data Quality and Trust: Review of Challenges and Opportunities for Data Sharing in IoT.* Electronics, 2020. **9**(12): p. 2083.
26. Baqa, H., et al. *Quality of Information as an indicator of Trust in the Internet of Things*. in *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*. 2018. IEEE.
27. Guptill, S.C. and J.L. Morrison, *Elements of spatial data quality*. 2013: Elsevier.

28. Metzger, A., et al. *Research challenges on online service quality prediction for proactive adaptation*. in *2012 First International Workshop on European Software Services and Systems Research-Results and Challenges (S-Cube)*. 2012. IEEE.
29. Guo, B., et al., *Opportunistic IoT: Exploring the harmonious interaction between human and the internet of things.* Journal of Network and Computer Applications, 2013. **36**(6): p. 1531-1539.
30. Barnaghi, P. and A. Sheth, *On searching the Internet of Things: Requirements and challenges.* IEEE Intelligent Systems, 2016. **31**(6): p. 71-75.
31. Klein, A. and W. Lehner. *How to optimize the quality of sensor data streams*. in *2009 Fourth International Multi-Conference on Computing in the Global Information Technology*. 2009. IEEE.
32. Togneri, R., et al. *Foundations of data quality assurance for IoT-based smart applications*. in *2019 IEEE Latin-American Conference on Communications (LATINCOM)*. 2019. IEEE.
33. Farooqi, M.M., H.A. Khattak, and M. Imran. *Data quality techniques in the Internet of Things: random forest regression*. in *2018 14th International Conference on Emerging Technologies (ICET)*. 2018. IEEE.
34. Firmani, D., et al., *On the meaningfulness of "big data quality".* Data Science and Engineering, 2016. **1**(1): p. 6-20.
35. Barnaghi, P., M. Bermudez-Edo, and R. Tönjes, *Challenges for quality of data in smart cities.* Journal of Data and Information Quality (JDIQ), 2015. **6**(2-3): p. 1-4.
36. Wang, R.Y. and D.M. Strong, *Beyond accuracy: What data quality means to data consumers.* Journal of management information systems, 1996. **12**(4): p. 5-33.
37. Wolfswinkel, J.F., E. Furtmueller, and C.P. Wilderom, *Using grounded theory as a method for rigorously reviewing literature.* European journal of information systems, 2013. **22**(1): p. 45-55.
38. Ramaswamy, L., V. Lawson, and S.V. Gogineni. *Towards a quality-centric big data architecture for federated sensor services*. in *2013 ieee international congress on big data*. 2013. IEEE.
39. Wei, X., et al., *A survey on quality-assurance approximate stream processing and applications.* Future Generation Computer Systems, 2019. **101**: p. 1062-1080.
40. Anagnostopoulos, C. and K. Kolomvatsos, *A delay-resilient and quality-aware mechanism over incomplete contextual data streams.* Information Sciences, 2016. **355**: p. 90-109.
41. Setia, P., et al., *Leveraging digital technologies: How information quality leads to localized capabilities and customer service performance.* Mis Quarterly, 2013: p. 565-590.
42. Mavrogiorgou, A., et al., *Analyzing data and data sources towards a unified approach for ensuring end-to-end data and data sources quality in healthcare 4.0.* Computer methods and programs in biomedicine, 2019. **181**: p. 104967.
43. Sicari, S., et al., *Toward data governance in the internet of things*, in *New advances in the internet of things*. 2018, Springer. p. 59-74.
44. Cai, L. and Y. Zhu, *The challenges of data quality and data quality assessment in the big data era.* Data science journal, 2015. **14**.
45. Chabridon, S., et al., *A survey on addressing privacy together with quality of context for context management in the Internet of Things.* Annals of telecommunications-annales des télécommunications, 2014. **69**(1): p. 47-62.
46. Karkouch, A., et al. *A model-driven architecture-based data quality management framework for the internet of Things*. in *2016 2nd International Conference on Cloud Computing Technologies and Applications (CloudTech)*. 2016. IEEE.
47. Cappiello, C. and F.A. Schreiber. *Quality-and energy-aware data compression by aggregation in wsn data streams*. in *2009 IEEE International Conference on Pervasive Computing and Communications*. 2009. IEEE.
48. Klein, A. and W. Lehner, *Representing data quality in sensor data streaming environments.* Journal of Data and Information Quality (JDIQ), 2009. **1**(2): p. 1-28.

49. Van der Togt, R., P.J. Bakker, and M.W. Jaspers, *A framework for performance and data quality assessment of Radio Frequency IDentification (RFID) systems in health care settings.* Journal of Biomedical Informatics, 2011. **44**(2): p. 372-383.
50. Sicari, S., et al. *A NFP model for internet of things applications*. in *2014 IEEE 10th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*. 2014. IEEE.
51. Fan, W., *Data quality: From theory to practice.* Acm Sigmod Record, 2015. **44**(3): p. 7-18.
52. Lawson, V. and L. Ramaswamy. *Data quality and energy management tradeoffs in sensor service clouds*. in *2015 IEEE International Congress on Big Data*. 2015. IEEE.
53. Karkouch, A., et al. *Data quality enhancement in Internet of Things environment*. in *2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)*. 2015. IEEE.
54. Liono, J., et al., *QDaS: Quality driven data summarisation for effective storage management in Internet of Things.* Journal of Parallel and Distributed Computing, 2019. **127**: p. 196-208.
55. Perez-Castillo, R., et al. *Data quality best practices in IoT environments*. in *2018 11th International Conference on the Quality of Information and Communications Technology (QUATIC)*. 2018. IEEE.
56. Byabazaire, J., G. O'Hare, and D. Delaney. *Data Quality and Trust: A Perception from Shared Data in IoT*. in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*. 2020. IEEE.
57. Sidi, F., et al. *Data quality: A survey of data quality dimensions*. in *2012 International Conference on Information Retrieval & Knowledge Management*. 2012. IEEE.
58. Azimi, S. and C. Pahl. *Continuous Data Quality Management for Machine Learning based Data-as-a-Service Architectures*. in *CLOSER*. 2021.
59. Aggarwal, C.C., N. Ashish, and A. Sheth, *The internet of things: A survey from the data-centric perspective*, in *Managing and mining sensor data*. 2013, Springer. p. 383-428.
60. Chen, K.-C. and S.-Y. Lien, *Machine-to-machine communications: Technologies and challenges.* Ad Hoc Networks, 2014. **18**: p. 3-23.
61. Pipino, L.L., Y.W. Lee, and R.Y. Wang, *Data quality assessment.* Communications of the ACM, 2002. **45**(4): p. 211-218.
62. Jara, A.J., L. Ladid, and A.F. Gómez-Skarmeta, *The Internet of Everything through IPv6: An Analysis of Challenges, Solutions and Opportunities.* J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl., 2013. **4**(3): p. 97-118.
63. Sezer, O.B., E. Dogdu, and A.M. Ozbayoglu, *Context-aware computing, learning, and big data in internet of things: a survey.* IEEE Internet of Things Journal, 2017. **5**(1): p. 1-27.
64. Bao, F., R. Chen, and J. Guo. *Scalable, adaptive and survivable trust management for community of interest based internet of things systems*. in *2013 IEEE eleventh international symposium on autonomous decentralized systems (ISADS)*. 2013. IEEE.
65. Yan, Z., P. Zhang, and A.V. Vasilakos, *A survey on trust management for Internet of Things.* Journal of network and computer applications, 2014. **42**: p. 120-134.
66. Kang, K., et al., *An interactive trust model for application market of the internet of things.* IEEE Transactions on Industrial Informatics, 2014. **10**(2): p. 1516-1526.
67. Al-Fuqaha, A., et al., *Toward better horizontal integration among IoT services.* IEEE Communications Magazine, 2015. **53**(9): p. 72-79.
68. Čolaković, A. and T. Čaršimamović, *The corresponding options of TCP variants for fairness problem in ad hoc networks.* International Journal of Soft Computing and Engineering, 2015. **5**(2): p. 1-8.

69. Rizzardi, A., et al., *AUPS: An open source AUthenticated Publish/Subscribe system for the Internet of Things.* Information Systems, 2016. **62**: p. 29-41.
70. Čolaković, A. and M. Hadžialić, *Internet of Things (IoT): A review of enabling technologies, challenges, and open research issues.* Computer Networks, 2018. **144**: p. 17-39.
71. Luo, T., et al., *Improving IoT data quality in mobile crowd sensing: A cross validation approach.* IEEE Internet of Things Journal, 2019. **6**(3): p. 5651-5664.
72. Davenport, T. and T. Redman, *Build data quality into the internet of things.* Wall Str. J, 2015.
73. Gil, D., et al., *Internet of things: A review of surveys based on context aware intelligent services.* Sensors, 2016. **16**(7): p. 1069.
74. Monga, M. and S. Sicari. *Assessing data quality by a cross-layer approach*. in *2009 International Conference on Ultra Modern Telecommunications & Workshops*. 2009. IEEE.
75. Li, F., S. Nastic, and S. Dustdar. *Data quality observation in pervasive environments*. in *2012 IEEE 15th International Conference on Computational Science and Engineering*. 2012. IEEE.
76. Jing, Q., et al., *Security of the Internet of Things: perspectives and challenges.* Wireless Networks, 2014. **20**(8): p. 2481-2501.
77. Chandola, V., A. Banerjee, and V. Kumar, *Anomaly Detection A Survey. ACM Computing Surveys.* CSUR), 41 (3), 2009. **15**.
78. Pinto-Valverde, J.M., et al., *HDQM2: healthcare data quality maturity model.* 2013.
79. Janssen, M., H. van der Voort, and A. Wahyudi, *Factors influencing big data decision-making quality.* Journal of Business Research, 2017. **70**: p. 338-345.
80. Simpson, M., et al. *A Platform for the Analysis of Qualitative and Quantitative Data about the Built Environment and its Users*. in *2017 IEEE 13th International Conference on e-Science (e-Science)*. 2017. IEEE.
81. Kamilaris, A., A. Kartakoullis, and F.X. Prenafeta-Boldú, *A review on the practice of big data analysis in agriculture.* Computers and Electronics in Agriculture, 2017. **143**: p. 23-37.
82. Zhang, Y., N. Meratnia, and P. Havinga, *Outlier detection techniques for wireless sensor networks: A survey.* IEEE communications surveys & tutorials, 2010. **12**(2): p. 159-170.
83. Yang, S., *IoT stream processing and analytics in the fog.* IEEE Communications Magazine, 2017. **55**(8): p. 21-27.
84. Vilenski, E., P. Bak, and J.D. Rosenblatt, *Multivariate anomaly detection for ensuring data quality of dendrometer sensor networks.* Computers and Electronics in Agriculture, 2019. **162**: p. 412-421.
85. Fan, W., et al. *Detecting inconsistencies in distributed data*. in *2010 IEEE 26th International Conference on Data Engineering (ICDE 2010)*. 2010. IEEE.
86. Asbury, J.-E., *Overview of focus group research.* Qualitative health research, 1995. **5**(4): p. 414-420.
87. Mohanta, B.K., et al., *Survey on IoT security: challenges and solution using machine learning, artificial intelligence and blockchain technology.* Internet of Things, 2020: p. 100227.
88. Sengupta, J., S. Ruj, and S.D. Bit, *A comprehensive survey on attacks, security issues and blockchain solutions for IoT and IIoT.* Journal of Network and Computer Applications, 2020. **149**: p. 102481.
89. Hossain, E., et al., *Application of big data and machine learning in smart grid, and associated security concerns: A review.* IEEE Access, 2019. **7**: p. 13960-13988.
90. Ghosh, A., D. Chakraborty, and A. Law, *Artificial intelligence in Internet of things.* CAAI Transactions on Intelligence Technology, 2018. **3**(4): p. 208-218.
91. Bamakan, S.M.H., N. Faregh, and A. ZareRavasan, *Di-ANFIS: an integrated blockchain–IoT–big data-enabled framework for evaluating service supply chain performance.* Journal of Computational Design and Engineering, 2021. **8**(2): p. 676-690.

92. Cha, S.-C., et al., *A blockchain connected gateway for BLE-based devices in the internet of things.* IEEE Access, 2018. **6**: p. 24639-24649.
93. Casado-Vara, R., et al. *Blockchain framework for IoT data quality via edge computing*. in *Proceedings of the 1st Workshop on Blockchain-enabled Networked Sensor Systems*. 2018.
94. Azaria, A., et al. *Medrec: Using blockchain for medical data access and permission management*. in *2016 2nd International Conference on Open and Big Data (OBD)*. 2016. IEEE.
95. Kar, S., et al., *Risk analysis of blockchain application for aerospace records management*. 2019, SAE Technical Paper.

Appendix I. The list of sources used in the review

Liu, C., et al., *Data quality and the Internet of Things.* Computing, 2020. **102**(2): p. 573-599.

Côrte-Real, N., P. Ruivo, and T. Oliveira, *Leveraging internet of things and big data analytics initiatives in European and American firms: Is data quality a way to extract business value?* Information & Management, 2020. **57**(1): p. 103141.

Karkouch, A., et al., *Data quality in internet of things: A state-of-the-art survey.* Journal of Network and Computer Applications, 2016. **73**: p. 57-81.

Sicari, S., et al., *A secure and quality-aware prototypical architecture for the Internet of Things.* Information Systems, 2016. **58**: p. 43-55.

Lee, Y.W., et al., *AIMQ: a methodology for information quality assessment.* Information & management, 2002. **40**(2): p. 133-146.

Byabazaire, J., G. O'Hare, and D. Delaney, *Data Quality and Trust: Review of Challenges and Opportunities for Data Sharing in IoT.* Electronics, 2020. **9**(12): p. 2083.

Metzger, A., et al. *Research challenges on online service quality prediction for proactive adaptation*. in *2012 First International Workshop on European Software Services and Systems Research-Results and Challenges (S-Cube)*. 2012. IEEE.

Guo, B., et al., *Opportunistic IoT: Exploring the harmonious interaction between human and the internet of things.* Journal of Network and Computer Applications, 2013. **36**(6): p. 1531-1539.

Farooqi, M.M., H.A. Khattak, and M. Imran. *Data quality techniques in the Internet of Things: random forest regression*. in *2018 14th International Conference on Emerging Technologies (ICET)*. 2018. IEEE.

Barnaghi, P., M. Bermudez-Edo, and R. Tönjes, *Challenges for quality of data in smart cities.* Journal of Data and Information Quality (JDIQ), 2015. **6**(2-3): p. 1-4.

Wang, R.Y. and D.M. Strong, *Beyond accuracy: What data quality means to data consumers.* Journal of management information systems, 1996. **12**(4): p. 5-33.

Wei, X., et al., *A survey on quality-assurance approximate stream processing and applications.* Future Generation Computer Systems, 2019. **101**: p. 1062-1080.

Anagnostopoulos, C. and K. Kolomvatsos, *A delay-resilient and quality-aware mechanism over incomplete contextual data streams.* Information Sciences, 2016. **355**: p. 90-109.

Sicari, S., et al., *Toward data governance in the internet of things*, in *New advances in the internet of things*. 2018, Springer. p. 59-74.

Cai, L. and Y. Zhu, *The challenges of data quality and data quality assessment in the big data era.* Data science journal, 2015. **14**.

Chabridon, S., et al., *A survey on addressing privacy together with quality of context for context management in the Internet of Things.* Annals of telecommunications-annales des télécommunications, 2014. **69**(1): p. 47-62.

Karkouch, A., et al. *A model-driven architecture-based data quality management framework for the internet of Things*. in *2016 2nd International Conference on Cloud Computing Technologies and Applications (CloudTech)*. 2016. IEEE.

Klein, A. and W. Lehner, *Representing data quality in sensor data streaming environments.* Journal of Data and Information Quality (JDIQ), 2009. **1**(2): p. 1-28.
Van der Togt, R., P.J. Bakker, and M.W. Jaspers, *A framework for performance and data quality assessment of Radio Frequency IDentification (RFID) systems in health care settings.* Journal of Biomedical Informatics, 2011. **44**(2): p. 372-383.
Sicari, S., et al. *A NFP model for internet of things applications*. in *2014 IEEE 10th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*. 2014. IEEE.
Fan, W., *Data quality: From theory to practice.* Acm Sigmod Record, 2015. **44**(3): p. 7-18.
Lawson, V. and L. Ramaswamy. *Data quality and energy management tradeoffs in sensor service clouds*. in *2015 IEEE International Congress on Big Data*. 2015. IEEE.
Karkouch, A., et al. *Data quality enhancement in Internet of Things environment*. in *2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)*. 2015. IEEE.
Liono, J., et al., *QDaS: Quality driven data summarisation for effective storage management in Internet of Things.* Journal of Parallel and Distributed Computing, 2019. **127**: p. 196-208.
Perez-Castillo, R., et al. *Data quality best practices in IoT environments*. in *2018 11th International Conference on the Quality of Information and Communications Technology (QUATIC)*. 2018. IEEE.
Byabazaire, J., G. O'Hare, and D. Delaney. *Data Quality and Trust: A Perception from Shared Data in IoT*. in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*. 2020. IEEE.
Sidi, F., et al. *Data quality: A survey of data quality dimensions*. in *2012 International Conference on Information Retrieval & Knowledge Management*. 2012. IEEE.
Azimi, S. and C. Pahl. *Continuous Data Quality Management for Machine Learning based Data-as-a-Service Architectures*. in *CLOSER*. 2021.
Qin, Y., et al., *When things matter: A survey on data-centric internet of things.* Journal of Network and Computer Applications, 2016. **64**: p. 137-153.
Aggarwal, C.C., N. Ashish, and A. Sheth, *The internet of things: A survey from the data-centric perspective*, in *Managing and mining sensor data*. 2013, Springer. p. 383-428.
Chen, K.-C. and S.-Y. Lien, *Machine-to-machine communications: Technologies and challenges.* Ad Hoc Networks, 2014. **18**: p. 3-23.
Pipino, L.L., Y.W. Lee, and R.Y. Wang, *Data quality assessment.* Communications of the ACM, 2002. **45**(4): p. 211-218.
Jara, A.J., L. Ladid, and A.F. Gómez-Skarmeta, *The Internet of Everything through IPv6: An Analysis of Challenges, Solutions and Opportunities.* J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl., 2013. **4**(3): p. 97-118.
Sezer, O.B., E. Dogdu, and A.M. Ozbayoglu, *Context-aware computing, learning, and big data in internet of things: a survey.* IEEE Internet of Things Journal, 2017. **5**(1): p. 1-27.
Bao, F., R. Chen, and J. Guo. *Scalable, adaptive and survivable trust management for community of interest based internet of things systems*. in *2013 IEEE eleventh international symposium on autonomous decentralized systems (ISADS)*. 2013. IEEE.
Yan, Z., P. Zhang, and A.V. Vasilakos, *A survey on trust management for Internet of Things.* Journal of network and computer applications, 2014. **42**: p. 120-134.
Kang, K., et al., *An interactive trust model for application market of the internet of things.* IEEE Transactions on Industrial Informatics, 2014. **10**(2): p. 1516-1526.
Al-Fuqaha, A., et al., *Toward better horizontal integration among IoT services.* IEEE Communications Magazine, 2015. **53**(9): p. 72-79.
Čolaković, A. and T. Čaršimamović, *The corresponding options of TCP variants for fairness problem in ad hoc networks.* International Journal of Soft Computing and Engineering, 2015. **5**(2): p. 1-8.
Rizzardi, A., et al., *AUPS: An open source AUthenticated Publish/Subscribe system for the Internet of Things.* Information Systems, 2016. **62**: p. 29-41.

Čolaković, A. and M. Hadžialić, *Internet of Things (IoT): A review of enabling technologies, challenges, and open research issues.* Computer Networks, 2018. **144**: p. 17-39.
Setia, P., et al., *Leveraging digital technologies: How information quality leads to localized capabilities and customer service performance.* Mis Quarterly, 2013: p. 565-590.
Ramaswamy, L., V. Lawson, and S.V. Gogineni. *Towards a quality-centric big data architecture for federated sensor services*. in *2013 ieee international congress on big data*. 2013. IEEE.
Luo, T., et al., *Improving IoT data quality in mobile crowd sensing: A cross validation approach.* IEEE Internet of Things Journal, 2019. **6**(3): p. 5651-5664.
Davenport, T. and T. Redman, *Build data quality into the internet of things.* Wall Str. J, 2015.
Togneri, R., et al. *Foundations of data quality assurance for IoT-based smart applications*. in *2019 IEEE Latin-American Conference on Communications (LATINCOM)*. 2019. IEEE.
Gil, D., et al., *Internet of things: A review of surveys based on context aware intelligent services.* Sensors, 2016. **16**(7): p. 1069.
Monga, M. and S. Sicari. *Assessing data quality by a cross-layer approach*. in *2009 International Conference on Ultra Modern Telecommunications & Workshops*. 2009. IEEE.
Li, F., S. Nastic, and S. Dustdar. *Data quality observation in pervasive environments*. in *2012 IEEE 15th International Conference on Computational Science and Engineering*. 2012. IEEE.
Jing, Q., et al., *Security of the Internet of Things: perspectives and challenges.* Wireless Networks, 2014. **20**(8): p. 2481-2501.
Chandola, V., A. Banerjee, and V. Kumar, *Anomaly Detection A Survey. ACM Computing Surveys.* CSUR), 41 (3), 2009. **15**.
Pinto-Valverde, J.M., et al., *HDQM2: healthcare data quality maturity model.* 2013.
Ghallab, H., H. Fahmy, and M. Nasr, *Detection outliers on internet of things using big data technology.* Egyptian Informatics Journal, 2020. **21**(3): p. 131-138.
Simpson, M., et al. *A Platform for the Analysis of Qualitative and Quantitative Data about the Built Environment and its Users*. in *2017 IEEE 13th International Conference on e-Science (e-Science)*. 2017. IEEE.
Janssen, M., H. van der Voort, and A. Wahyudi, *Factors influencing big data decision-making quality*. Journal of Business Research, 2017. **70**: p. 338-345.
Yang, S., *IoT stream processing and analytics in the fog.* IEEE Communications Magazine, 2017. **55**(8): p. 21-27.
Vilenski, E., P. Bak, and J.D. Rosenblatt, *Multivariate anomaly detection for ensuring data quality of dendrometer sensor networks.* Computers and Electronics in Agriculture, 2019. **162**: p. 412-421.
Fan, W., et al. *Detecting inconsistencies in distributed data*. in *2010 IEEE 26th International Conference on Data Engineering (ICDE 2010)*. 2010. IEEE.