

# Data Quality Management in IoT

Subjects: [Computer Science](#), [Information Systems](#)

Contributor: LINA ZHANG

The concept of the Internet of Things (IoT) was first introduced by Ashton o describe the ability of sensors to connect to the Internet and provide new services. Ma also defined the IoT as a network that connects an ordinary physical object with an identifiable address to provide intelligent services. Nowadays, IoT is being used in more and more application areas and the importance of IoT data quality is widely recognized by practitioners and researchers. The requirements for data and its quality vary from application to application or organization in different contexts. Many methodologies and frameworks include techniques for defining, assessing, and improving data quality. However, due to the diversity of requirements, it can be a challenge to choose the appropriate technique for the IoT system.

- data quality
- data quality management
- data quality methodology
- data quality standard
- IoT data

## 1. Data Quality in IoT

It is evident from previous research <sup>[1]</sup> that there is usually an important phase in the initial stage of the DQ management technique that is used to define the data and DQ. This includes the context and type of data. In addition, analyzing the factors in the system that potentially affect DQ has an impact on DQ assessment. This chapter summarizes the classification of general data and IoT data, describes the definition of DQ, and provides a discussion of the potential factors that affect IoT DQ.

### 1.1. Definition of Types of Data

Data are “abstract representations of selected characteristics of real-world objects, events, and concepts, expressed and understood through explicitly definable conventions related to their meaning, collection, and storage” <sup>[2]</sup>. The term information is used in some studies and is interchangeable with data, without clear distinction. In <sup>[3]</sup>, the authors use data to denote structured data in databases, and other broader types of data are described by information except linked open data, and Big Data. In some studies <sup>[4][5]</sup>, the generic term information is used to indicate that the study may involve any type of data, without specifying a specific type of data. Data can be categorized into different types depending on how they are used in different areas. Researchers proposed several classifications of data, as shown in **Table 1**. In the existing studies, the most widely used data classification is that based on data structure.

**Table 1.** Classifications for data.

Ref.	Basis	Data Types	Description
<sup>[1][6][7]</sup>	Structure	Structured data	Data with formal schema definition; (e.g., relation tables)
<sup>[8][9]</sup>		Unstructured data	Generic sequence of symbols (e.g., video)
<sup>[10][11]</sup>		Semi-structured data	Data partly structured or have a descriptive without schema (e.g., XML file)
<sup>[7][12]</sup>	Change frequency	Stable data	Data impossible to change
		Long-term changing data	Data with very low frequency of change
	Product	Frequently changing data	Dramatically changing data, (e.g., real-time traffic information)
<sup>[1][6][7][13]</sup>		Raw data items	Data that have not been processed
		Information products	Results of manufacturing activities
<sup>[7][14]</sup>	Nature	Component data items	Semi-processed information
		Federated data	Data from different heterogeneous sources
		Web data	Data from the Web
		High-dimensional data	Big data
		Descriptive data	Consists of many tables with complex interrelationships.

Ref.	Basis	Data Types	Description
		Longitudinal data	Time series data
		Streaming data	Data generated sequentially at a higher rate in a single source

The International Organization for Standardization (ISO)000-2 [15] defines data as “an interpretive representation of information in a appropriate format”. ISO 8000-100 [16] lists special data including but not limited to: master data, transaction data, and measurement data. Master data are data that an organization hold to characterize entities that are both separate and essential to that organization and are cited for the purpose of executing its transactions, consist of reference data and feature data. Transaction data are data that represents business transactions. Measurement data are the data to record the measurement results.

IoT data collects smart things over networks, with special characteristics such as distribution, volume, velocity, variety, veracity, value, spatio-temporality and dynamicity. Different characteristics and different sources of IoT data may have different data management methods. Before carrying out DQ management or assessment, it is very important to determine the type of data. Fathy et al. summarized the three classifications of IoT data, and explained each category, including numerical (referring to data with numerical values) vs. symbolic (data that have string/text values), discrete (data with finite values or finite sets of values) vs. continuous (data with unlimited sets of values), and static(data that do not change) vs. streaming(data that changes over time) [17][18]. Klein et al. [19] present findings that show streaming data includes the numerical, discretized and digitized data. Cooper and James classified IoT data into Radio Frequency Identification (RFID) address/unique identifiers, descriptive data, positional and environmental data, sensor data, historical data, the physics model, and command data based on domain [20].

From the perspective of data semantics, the data in the IoT can be the underlying raw data and the high-level generalized data. Different data formats introduce the basis of data polymorphism and data heterogeneity [21]. Kim et al. [22] divide the data generated and used in the IoT into six categories, including sensor data (sensor-generated data); observed metadata (describe sensor data behavior); device metadata (describe the characteristics of the device or sensor); business data (for business purposes); external data (provide additional information for product capabilities, such as weather) and technical metadata (data standards and physical data storage structures). Perez-Castillo et al. consider the dependence on various data sources and classify data involved in the IoT into four categories [23]: sensor data, which is generated by sensors and digitized into machine-readable data (For example, the reading of temperature sensors); device data: metadata of sensor observations and IoT devices (for example, the timestamp of the observation and the device manufacturer); general data: IoT-device-generated or device-related data (for example, sensor observations stored in a database); IoT data: in an IoT system, all data other than the raw data generated by sensors are collectively referred to as IoT data, which is a collection of general data and device data. Many studies have been published on sensor DQ [24][18][25][26][27][28][29][30][31][32] and streaming data DQ management [18][19][26][33]. We summarize various classification methods for IoT data mentioned in the literature, as shown in Figure 1.

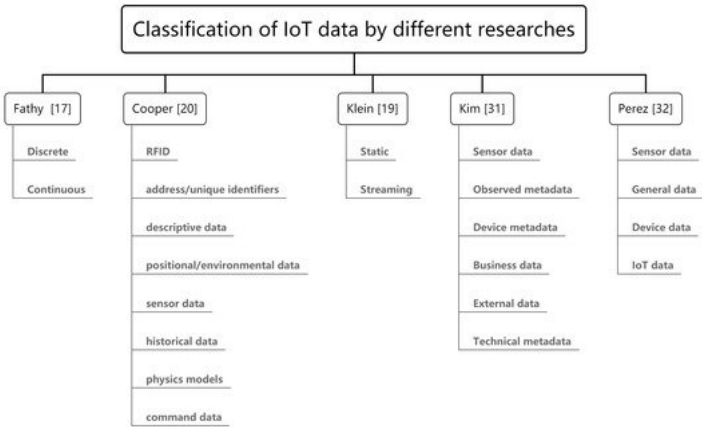


Figure 1. Classification of IoT data by different research.

1.2. Definition of Data Quality

DQ has been defined differently in various fields and time periods. The understanding of the concept of DQ mainly includes the following two perspectives: first, it focuses on measuring DQ from a practical perspective, i.e., it is judged from a user perspective, emphasizing user satisfaction, and also from data producers and managers; second, it focuses on evaluation from a system-oriented perspective, considering DQ as a comprehensive concept and a multidimensional concept. It is

necessary to measure its basic quality elements from multiple perspectives, such as accuracy, timeliness, completeness, and consistency.

One of the first people to define a basis DQ was Wang et al. [34], who wrote: “data that are fit for use by data consumers”. Juran et al. [35] provided a new definition of DQ: “data are of high quality if they are fit for their intended uses in operations, decision making, and planning”. The ISO 9000 standard [36] defines quality as “the degree to which a set of inherent characteristics fulfills a need or expectation that is stated, generally implied, or obligatory”.

The ISO 8000-2 [15] present DQ as the degree to which the inherent characteristics of data meet the demands. The ISO 8000-8 classifies information and DQ into three categories: syntactic quality, semantic quality and pragmatic quality. Syntactic quality refers to the extent to which the data conform to the specified syntax, such as consistency with metadata. Semantic quality describes how well the data correspond to the content it stands for. Pragmatic quality refers to the extent to which the data are appropriate and valuable for a specific objective [37]. As defined by the ISO 8000, DQ includes the following principles:

- The data are fit for purpose;
- Being in the right place at the right time, with the right data;
- Meet the data requirements agreed by the customer;
- Preventing duplication and eliminating waste through enhancement phases and preventing the recurrence of data defects.

The definition of IoT DQ is basically aligned with the definition of DQ mentioned above. A further definition is given by Karkouch [38], who describes whether the data collected (from IoT devices) is appropriate for IoT users to provide ubiquitous services. IoT devices typically monitor a variable of interest in the physical world, such as temperature, sleep habits, and so on.

DQ research and practices can be categorized into top-down and bottom-up approaches [39]. The top-down approach usually proposes a DQ framework with DQ dimensions, and then by integrating with specific requirements in the application, more detailed DQ dimensions are constructed, while the bottom-up approach starts by refining a series of DQ dimensions from specific requirements, and through the demonstration of practical applications, the DQ framework is finally generalized.

### 1.3. Issues of Data Quality

Data suffering from quality issues are not representative of the true situation and may negatively impact the decision making and operational levels of any business or organization. The challenges facing the IoT data directly inherit and even amplify the characteristics of the Internet because of large-scale deployments of IoT devices, information flows, and indirect user involvement [40].

Lee et al. concluded ten root causes of DQ problems: multiple data sources, subjective judgments during data generation, insufficient computational resources, the balance of security and accessibility, cross-disciplinary encoding of data, complex data representation, data volume, input rules that are overly restrictive or ignored, evolving data demands, and distributed heterogeneous systems, respectively [41]. These ten root causes are equally applicable in IoT systems. Jeffery et al. [25] summarized two types of DQ problems “missed readings” and “unreliable readings” generated by the IoT devices. For example, the sensor average delivery is only 42% in an IoT experiment, which will lead to dropped data. IoT data may come from multiple different objects and have different formats, which will lead to inconsistencies in multi-source data [42]. Additionally, problems such as data duplication [43], data leakage, and time calibration of multiple data sources were reported in the studies.

To better examine and appreciate the DQ problems and challenges in the IoT, we describe the features and problems of the IoT via a three-layer structure [44]. A typical three-layer IoT system consists of the perception layer, the network layer, and the application layer. In the perception layer, which is also known as the device layer [45], the physical objects and IoT devices, such as DHT11, which includes temperature and humidity sensors, measure and collect the observed temperature and humidity results. Next, the network layer is used to send the observation results via wireless technologies, such as LoRa [46] and Bluetooth. Then, the application layer receives observation results from the previous layer, where data processing, analysis, and storage are all carried out, and provide users with ubiquitous services. Perez-Castill et al. propose a three-layer conceptual framework for IoT DQ, as shown in **Figure 2**, with each layer focusing on both device DQ and general DQ [47][48].

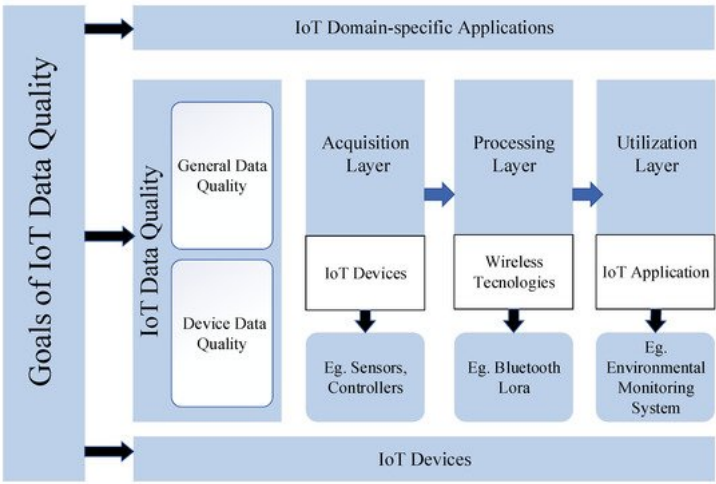


Figure 2. IoT DQ conceptual framework [48].

Many researchers have found that DQ problems may occur in different layers of the IoT structure and affect the DQ of IoT platform, which are: the deployment scale, resource constraints, network, sensors, environment, vandalism, fail-dirty, privacy preservation processing, security vulnerability, and data stream processing [38][49][18][31][50][51][52][53][54][55][56].

Teh et al. identified eight types of sensor data errors: anomalies, missing values, deviations, drift, noise, constant value, uncertainty, and stuck-at-zero [57]. The most common error is outliers, values that lie above the thresholds or significantly deviate from the normal behavior provided by the model. The second most common error in sensor data is missing data, also known as incomplete data in relational databases. Li and Parker [58] believed that missing data are due to a variety of factors, such as an unstable wireless connection caused by network congestion, the power failure of sensor devices caused by limited battery life, and environmental interference such as artificial blockage, walls, weather conditions and vandalism.

In each layer of the data transmitting process, there may be different DQ issues due to various impacts. As shown in Table 2, we summarized causes and the error types that may result from each layer [38][59], which should be detected and corrected to improve IoT DQ. While some issues affect only one layer, many cross multiple layers.

Table 2. Layered distribution of factors threatening IoT DQ.

Layer	Affecting Factors	Examples	Error Types
Perception layer	<ul style="list-style-type: none"><li>• Sensors</li><li>• Environment</li><li>• Security</li><li>• Privacy</li><li>• Network</li></ul>	<ul style="list-style-type: none"><li>• Battery problems</li></ul>	<ul style="list-style-type: none"><li>• Missing value [59]</li><li>• Incorrect value</li></ul>
		<ul style="list-style-type: none"><li>• Precision limitation</li></ul>	
		<ul style="list-style-type: none"><li>• Mechanical failures</li></ul>	
		<ul style="list-style-type: none"><li>• Bad weather</li></ul>	
		<ul style="list-style-type: none"><li>• Device upgrades</li></ul>	
		<ul style="list-style-type: none"><li>• Unstable network</li></ul>	
Network layer	<ul style="list-style-type: none"><li>• Network</li><li>• Environment</li><li>• Security</li><li>• Privacy</li></ul>	<ul style="list-style-type: none"><li>• Non-encrypted</li></ul>	<ul style="list-style-type: none"><li>• Missing value</li><li>• Incorrect value</li></ul>
		<ul style="list-style-type: none"><li>• Unstable network</li></ul>	
		<ul style="list-style-type: none"><li>• Bad weather</li></ul>	
		<ul style="list-style-type: none"><li>• Security attacks</li></ul>	
Application layer	<ul style="list-style-type: none"><li>• Streaming processing</li></ul>	<ul style="list-style-type: none"><li>• Manually errors</li></ul>	<ul style="list-style-type: none"><li>• Wrong schema definition</li></ul>

Layer	Affecting Factors	Examples	Error Types
	<ul style="list-style-type: none"><li>• Security</li></ul>	<ul style="list-style-type: none"><li>• Obsolete schema definition</li></ul>	<ul style="list-style-type: none"><li>• Misplaced value</li></ul>
	<ul style="list-style-type: none"><li>• Privacy</li></ul>	<ul style="list-style-type: none"><li>• Streaming operators</li></ul>	<ul style="list-style-type: none"><li>• Broken join relationship</li><li>• Misplaced column values</li><li>• Missing record</li></ul>

**2. Data Quality Management Techniques**

**2.1. Methodology and Frameworks Appropriate for IoT Data**

Batini et al. defined a DQ methodology [1] as “a set of guidelines and techniques that define a rational process for assessing and improving DQ, starting from describing the input information for a given application context, defines a rational process to assess and improve the quality of data”. A framework is considered as a theory-building and practice-oriented tool [60], providing a structure for using QA theory and methods [61][62]. The terms DQ methodology and DQ framework are often used interchangeably in related research. In this chapter, we review the general DQ management methodologies and frameworks, comparing them in terms of research objectives, management phases, applicable data types, the number of DQ dimensions, and whether they can be extended, respectively. Most of the research in DQ methodology has focused on structured and semi-structured data, while only a few of them also involve semi-structured data. Methodology and framework, in many studies, refer to the same thing.

Early on, Wang [63] proposed a general methodology “Total Data Quality Management (TDQM)”, which is one of the most famous complete and general methodologies. The TDQM treats data as information products and presents a comprehensive set of associated dimensions and enhancements, which can be applied to different contexts. However, the structure of the processable data is not specified. The goal of TDQM is to continuously enhance the quality of information products through a cycle of defining, measuring, analyzing and enhancing data and the process of managing them, without appropriate steps specified in the assessment process.

English [4] described a methodology of “Total Information Quality Management (TIQM)” applied to data warehouse projects. Later, due to its detailed design and universality, it became a generic information quality management methodology that can be customized for many backgrounds and different data types, including structured data, unstructured data, and semi-structured data, the latter two of which are not mentioned in the study but can be inferred. The TIQM cycle includes evaluation, improvement, and improvement management and monitoring. Compared with other methodology, TIQM is original and more comprehensive in terms of cost–benefit analysis and the management perspective [3]. However, during the evaluation phase, TIQM manages a fixed set of DQ dimensions, with a number of DQ dimensions of 13 and their solution strictly follows these dimensions. TIQM is one of the few methodologies that considers the cost dimension and provides detailed classifications for costs.

Lee et al. [5] presented “A Methodology for Information Quality Assessment (AIMQ)”, which is the first quality management method that focuses on benchmarking and will provide objective and domain-independent generic quality assessment techniques. The methodology designs a PSP/IQ model that provides a standard list of quality dimensions and attributes that can be used to categorize quality dimensions according to importance from a user and an administrator perspective. The AIMQ cycle includes the measurement, analysis, and interpretation of an assessment, and lacks guidance on activities to improve DQ. AIMQ uses questionnaires applicable to structured data for qualitative assessments but can be applied to other data types, including unstructured data and semi-structured data. Similar to TIQM, during the measurement phase, AIMQ manages a fixed group of DQ dimensions (metrics), with a number of dimensions of 15, and their solution strictly follows these dimensions.

Monica et al. [64] present a cooperative framework “DaQuinCIS” for DQ by applying TDQM, which is one of the rare methodologies that focuses on semi-structured data. This approach proposes a model, called data and data quality (D2Q). The model associates DQ values with XML documents, and can be used to verify the accuracy, currency, completeness, and consistency of the data. Another contribution of DaQuinCIS is the degree of flexibility that each organization has to export the quality of its data because of the semi-structured model.

Batini et al. [10] proposed a “Comprehensive Data Quality methodology (CDQ)” that extends the steps and techniques originally developed for all types of organizational data. CDQ integrates the phases, techniques and tools from other methodologies and overcomes some of the limitations in those methodologies. The CDQ cycle includes state reconstruction, assessment, and improvement. All data types, both structured and semi-structured, should be investigated in the state reconstruction step. CDQ manages four DQ dimensions and considers the cost of alternative improvement activities to compare and evaluate the minimum-cost improvement processes.

Cappiello [66] described a “Hybrid Information Quality Management (HIQM) methodology”, which supported error detection and correction management at runtime and improved the traditional DQ management cycle by adding the user perspective. For example, HIQM defines DQ by considering the needs of not only companies and suppliers, but also user end consumers to determine DQ requirements. The HIQM cycle includes definition, quality measurement, analysis and monitoring, and improvement. However, in the measurement stage, only the need for measurement algorithms for each DQ dimension is expressed, without defining specific metrics. In particular, TIQM designed a warning interface that represents an efficient way to analyze and manage problems and warnings that appear, and considers whether to recommend a recovery operation by analyzing the details of the warning message.

Caballero [66] proposed “A Methodology Based on ISO/IEC 15939 to Draw up Data Quality Measurement Process (MMPRO)”, which is based on the ISO/IEC 15939 standard [67] for software quality assessment and can also be used for DQ assessment. The MMPRO cycle includes the DQ Measurement Commitment, Plan the DQ Measurement Process, Perform the DQ Measurement Process and Evaluate the DQ Measurement Process. Although the approach does not categorize DQ measures or provide a set of behaviors for improving data quality, its structure helps to incorporate DQ issues into the software.

Maria et al. [68] described “A Data Quality Practical Approach (DQPA)”, which described a DQ framework in a heterogeneous multi-database environment and applied it with a use case. The DQPA cycle consists of seven phases, the including identification of DQ issues, identification of relevant data that has a direct impact on the business, evaluation, the determination of the business impact through DQ comparison, cleansing of data, monitoring the DQ, and carrying out the assessment stage regularly. In DQPA, the authors propose the Measurement Model based on [69][70][71], which extends the DQ assessment metrics into metrics for evaluating primary data sources and metrics for evaluating derived data. The model can be used at different levels of granularity for databases, relationships, tuples, and attributes.

Batini et al. [44] presented a “Heterogenous Data Quality Methodology (HDQM)”, which can be used to evaluate and improve the DQ, and has been verified by using cases. The HDQM cycle includes state reconstruction, assessment, and improvement. The HDQM recommends considering all types of data in the state reconstruction phase by using a model that describes the information according to the level of abstraction. In the assessment phase, HDQM defines a method that can be easily generalized to any dimension. Furthermore, the DQ dimensions of the HDQM measurement and improvement phase can be applied to different data types. A major contribution of HDQM is based on the techniques of the cost–benefit analysis in TIQM, COLDQ and CDQ, presenting a more qualitative approach to guide the selection of appropriate improvement techniques.

Laura et al. [2] described a “Data Quality Measurement Framework (DQAF)”, which provides a comprehensive set of objective DQ metrics for DQ assessment organizations to choose from, comprising 48 universal measurement types based on completeness, timeliness, validity, consistency, and integrity. In DQAF, the authors introduce a concept of “measurement type” that is a generic form suitable for a particular metric, and develop some strategies to describe six aspects of each measure type, including definition, business concerns, measurement methodology, programming, support processes and a skills and measurement logical model. The DQAF cycle includes define, measure, analyze, improve, and control. Specifically, the authors focus on comparing the results of the DQ assessment with assumptions or expectations, and continuously monitoring the data to ensure that it continues to meet the requirements.

Carretero et al. [72] developed an “Alarcos Data Improvement Model (MADM Framework)” that can be applied in many fields, which can provide a Process Reference Model and evaluation and improvement methods. Finally, it was verified with an applied hospital case. The MADM Framework cycle consists of a two-stage Process Reference Model based on the ISO 8000-61 standard and an Assessment and Improvement Model based on ISO/IEC 33000. The MADM Process Reference Model consists of 21 processes that can be used in the areas of data management, data governance, and DQ management quality. The assessment model is a methodology that consists of five steps and a maturity model.

Reza et al. [73] introduced an “observe–orient–decide–act (OODA)” framework to identify and improve DQ through the cyclic application of the OODA method, which is adaptive and can be used across industries, organizational types, and organizational scales. The OODA framework cycle includes observe, orient, decide and act. Only the need for a metric algorithm for each DQ dimension is indicated, and the OODA DQ approach refers to the use of existing DQ metrics and tools for metrics. Although the OODA DQ methodology does not involve any formal process for analysis and improvement processes, DQ issues are identified through tools such as routine reports and dashboards during the observe phase. In addition, notices alerting for possible DQ issues and feedback from external agencies are also recommended [7].

There are many more comparative perspectives on these 12 general DQ management methodologies/frameworks, such as flexibility in the choice of dimensions [2], the use of subjective or objective measurements in the assessment phase, specific steps in the assessment/improvement phase, cost considerations, data-driven or process-driven, etc. There is not much

research on IoT DQ assessment yet, and a beginner may have some difficulties on aspects such as how to make decisions, so start with the question, what are the general requirements of data users? If the user needs to manage IoT data in a holistic way that supports the definition, assessment and improvement process without resorting to some tool or software, the generic DQ management methodology/framework mentioned in this section can be chosen.

2.2. ISO Standards Related to Data Quality

Another important area of DQ in industry and academia is the research and standardization of DQ standards. By developing uniform DQ standards, DQ can be managed more efficiently across countries, organizations, and departments, thereby facilitating data storage, delivery, and sharing, and reducing errors in judgment and decision making due to data incompatibility, data redundancy, and data deficiencies. Since IoT systems are distributed in nature, the use of international standards can have a positive effect on improving the performance of business processes by aligning various organizations with the same foundation, addressing interoperability issues, and finally working in a seamless manner.

The ISOas made a great deal of effort in this regard and has developed several standards to regulate international data quality. The ISO 8000 DQ standard has been developed [74] to address the increasingly important issue of DQ and data management. ISO 8000 covers the quality characteristics of data throughout the product life cycle, from conceptual design to disposal. ISO 8000 describes a framework for improving the DQ of a particular data, which can be used independently or in cooperation with a quality management system.

The ISO 8000-6x family of standards provides a value-driven approach to DQ management. Several of the IoT data assessment frameworks reviewed in the next section are based on this standard. This series of standards provides a set of guidelines for the overall management of DQ that can be customized for different domains. It describes a DQ management structure derived from ISO 9001's Plan-Do-Check-Act (PDCA), a life cycle that includes DQ planning, DQ control, DQ assurance, and DQ improvement. However, it is not primarily intended as a methodology for DQ management, but merely to serve as a process reference model. Figure 3 depicts the components of the ISO 8000 DQ standard.



Figure 3. Components of the ISO 8000.

Before ISO 8000 DQ standards were published, a more mature management system of product quality standards existed—ISO 9000 [75]. Initially published by the ISO in 1987 and refined several times, the ISO 9000 family of standards was designed to help organizations ensure that they meet the needs of their customers and other stakeholders while meeting the legal and regulatory requirements associated with their products. It is a general requirement and guide for quality management that helps organizations to effectively implement and operate a quality management system. While ISO 9000 is concerned with product quality, ISO 8000 is focused on DQ. ISO 8000 is designed to improve data-based quality management systems, a standard that addresses the gap between ISO 9000 standards and data products [76].

In addition, international standards related to DQ include ISO/IEC 25012 Software Product Quality Requirements and Assessment Data Quality Model [77], ISO/IEC 25024 Quality Requirements and Evaluation of Systems and Software [78]—Measurement of Data Quality, etc. ISO/IEC 25012 standard proposes a DQ model called Software Product Quality Requirements and Evaluation (SQuaRE) that can be used to manage any type of data. It emphasizes the view of DQ as a part of the information system and defines quality features for the subject data. In the following, we compare the following 5 ISO standards that are often used in DQ management studies, as shown in Table 3.

Table 3. ISO standards related to data quality.

Standards	Components	Scope of Application
ISO/IEC 33000	Terminology related to process assessment; a framework for process quality assessment.	Information Technology Domain Systems

Standards	Components	Scope of Application
ISO/IEC 25000	A general DQ model; 15 data quality characteristics.	Structured data
ISO/IEC 15939	Activities of the measurement process; a suitable set of measures.	System and software engineering
ISO 9000	A quality management system; 7 quality management principles.	Quality management system
ISO 8000	Characteristics related to information and DQ; a framework for enhancing the quality of specific types of data methods for managing, measuring and refining information and DQ.	Partially available for all types of data, partially available for specified data types

The benefits of customizing and using international standards in the IoT context are: (1) the number of issues and system failures in the IoT environment will be reduced and all stakeholders will be aligned. (2) It is easier to apply DQ solutions on a global scale due to reduced heterogeneity. (3) DQ research in the IoT can be aligned with international standards to provide standardized solutions. (4) It enables better communication between partners.

## References

1. Batini, C.; Cappiello, C.; Francalanci, C.; Maurino, A. Methodologies for data quality assessment and improvement. *ACM Comput. Surv. (CSUR)* 2009, 41, 1–52.
2. Sebastian-Coleman, L. *Measuring Data Quality for Ongoing Improvement: A Data Quality Assessment Framework*; Morgan Kaufmann Publishers Inc.: Francisco, CA, USA, 2012.
3. Zhu, H.; Madnick, S.; Lee, Y.; Wang, R. *Data and Information Quality Research*; Springer: Cham, Switzerland, 2014; pp. 16–1–16–20.
4. English, L.P. *Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1999.
5. Lee, Y.W.; Strong, D.M.; Kahn, B.K.; Wang, R.Y. AIMQ: A methodology for information quality assessment. *Inf. Manag.* 2002, 40, 133–146.
6. Batini, C.; Scannapieca, M. *Data Quality: Concepts, Methodologies and Techniques*; Springer: Berlin/Heidelberg, Germany, 2006.
7. Cichy, C.; Rass, S. An overview of data quality frameworks. *IEEE Access* 2019, 7, 24634–24648.
8. Abiteboul, S. Querying semi-structured data. In *International Conference on Database Theory*; Springer: Berlin/Heidelberg, Germany, 1997; pp. 1–18.
9. Abiteboul, S.; Buneman, P.; Suciu, D. *Data on the Web: From Relations to Semistructured Data and XML*; Morgan Kaufmann: San Francisco, CA, USA, 2000.
10. Batini, C.; Cabitza, F.; Cappiello, C.; Francalanci, C. A comprehensive data quality methodology for web and structured data. *Int. J. Innov. Comput. Appl.* 2008, 1, 205–218.
11. Carlo, B.; Daniele, B.; Federico, C.; Simone, G. A data quality methodology for heterogeneous data. *J. Database Manag. Syst.* 2011, 3, 60–79.
12. Bouzeghoub, M. A framework for analysis of data freshness. In *Proceedings of the 2004 International Workshop on Information Quality in Information Systems*, Paris, France, 18 June 2004; pp. 59–67.
13. Shankaranarayanan, G.; Wang, R.Y.; Ziad, M. IP-MAP: Representing the Manufacture of an Information Product. *IQ* 2000, 2000, 1–16.
14. Dasu, T.; Johnson, T. *Exploratory Data Mining and Data Cleaning*; John Wiley & Sons: Hoboken, NJ, USA, 2003; Volume 479.
15. ISO. ISO 8000-2:2017 Data Quality—Part 2: Vocabulary; Standard, International Organization for Standardization/TC 184/SC 4 Industrial Data (2017); ISO: Geneva, Switzerland, 2017.
16. ISO. ISO 8000-100:2016 Data Quality—Part 100: Master Data: Exchange of Characteristic Data: Overview; Standard, International Organization for Standardization/TC 184/SC 4 Industrial Data (2016); ISO: Geneva, Switzerland, 2016.
17. Fathy, Y.; Barnaghi, P.; Tafazolli, R. Large-scale indexing, discovery, and ranking for the Internet of Things (IoT). *ACM Comput. Surv. (CSUR)* 2018, 51, 1–53.



18. Klein, A.; Do, H.H.; Hackenbroich, G.; Karnstedt, M.; Lehner, W. Representing data quality for streaming and static data. In Proceedings of the 2007 IEEE 23rd International Conference on Data Engineering Workshop, Istanbul, Turkey, 17–20 April 2007; pp. 3–10.
19. Klein, A.; Lehner, W. Representing data quality in sensor data streaming environments. *J. Data Inf. Qual. (JDIQ)* 2009, 1, 1–28.
20. Cooper, J.; James, A. Challenges for database management in the internet of things. *IETE Tech. Rev.* 2009, 26, 320–329.
21. Vongsingthong, S.; Smachet, S. A review of data management in internet of things. *Asia Pac. J. Sci. Technol.* 2015, 20, 215–240.
22. Kim, S.; Del Castillo, R.P.; Caballero, I.; Lee, J.; Lee, C.; Lee, D.; Lee, S.; Mate, A. Extending data quality management for smart connected product operations. *IEEE Access* 2019, 7, 144663–144678.
23. Perez-Castillo, R.; Carretero, A.G.; Caballero, I.; Rodriguez, M.; Piattini, M.; Mate, A.; Kim, S.; Lee, D. DAQUA-MASS: An ISO 8000-61 based data quality management methodology for sensor data. *Sensors* 2018, 18, 3105.
24. De Aquino, G.R.; De Farias, C.M.; Pirmez, L. Data quality assessment and enhancement on social and sensor data. *CEUR Workshop Proc.* 2018, 2247, 1–7.
25. Jeffery, S.R.; Alonso, G.; Franklin, M.J.; Hong, W.; Widom, J. Declarative support for sensor data cleaning. In *International Conference on Pervasive Computing*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 83–100.
26. Klein, A.; Lehner, W. How to optimize the quality of sensor data streams. In Proceedings of the 2009 Fourth International Multi-Conference on Computing in the Global Information Technology, Cannes/La Bocca, France, 23–29 August 2009; pp. 13–19.
27. Kuemper, D.; Iggena, T.; Toenjes, R.; Pulvermueller, E. Valid. IoT: A framework for sensor data quality analysis and interpolation. In Proceedings of the 9th ACM Multimedia Systems Conference, Amsterdam, The Netherlands, 12–15 June 2018; pp. 294–303.
28. Okafor, N.U.; Alghorani, Y.; Delaney, D.T. Improving Data Quality of Low-cost IoT Sensors in Environmental Monitoring Networks Using Data Fusion and Machine Learning Approach. *ICT Express* 2020, 6, 220–228.
29. Aggarwal, C.C. *Managing and Mining Sensor Data*; Springer Science & Business Media: New York, NY, USA, 2013.
30. Qin, Z.; Han, Q.; Mehrotra, S.; Venkatasubramanian, N. Quality-aware sensor data management. In *The Art of Wireless Sensor Networks*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 429–464.
31. Branch, J.W.; Giannella, C.; Szymanski, B.; Wolff, R.; Kargupta, H. In-network outlier detection in wireless sensor networks. *Knowl. Inf. Syst.* 2013, 34, 23–54.
32. Sanyal, S.; Zhang, P. Improving Quality of Data: IoT Data Aggregation Using Device to Device Communications. *IEEE Access* 2018, 6, 67830–67840.
33. Geisler, S.; Quix, C.; Weber, S.; Jarke, M. Ontology-based data quality management for data streams. *J. Data Inf. Qual. (JDIQ)* 2016, 7, 1–34.
34. Wang, R.Y.; Strong, D.M. Beyond accuracy: What data quality means to data consumers. *J. Manag. Inf. Syst.* 1996, 12, 5–33.
35. Juran, J.M.; Godfrey, A.B. *Juran's Quality Handbook*, 5th ed.; McGraw-Hill Companies: New York, NY, USA, 1999.
36. Hoyle, D. *ISO 9000 Quality Systems Handbook-Updated for the ISO 9001:2008 Standard*; Putting ISO 9000 in Context; ISO: Geneva, Switzerland, 2009; pp. 3–21.
37. ISO. *ISO 8000-8:2015 Data Quality—Part 8: Information and Data Quality: Concepts and Measuring*; Standard, International Organization for Standardization/TC 184/SC 4 Industrial Data (2015); ISO: Geneva, Switzerland, 2015.
38. Karkouch, A.; Mousannif, H.; Al Moatassime, H.; Noel, T. Data quality in internet of things: A state-of-the-art survey. *J. Netw. Comput. Appl.* 2016, 73, 57–81.
39. Wang, Z.; Yang, Q. Research on Scientific Data Quality and Its Standardization. *Stand. Sci.* 2019, 3, 25–30.

40. Chen, Q.; Britto, R.; Erill, I.; Jeffery, C.J.; Liberzon, A.; Magrane, M.; Onami, J.I.; Robinson-Rechavi, M.; Sponarova, J.; Zobel, J.; et al. Quality matters: Biocuration experts on the impact of duplication and other data quality issues in biological databases. *Genom. Proteom. Bioinform.* 2020, 18, 91.
41. Lee, Y.W.; Pipino, L.L. *Journey to Data Quality*; MIT Press: Cambridge, UK, 2006.
42. Mishra, N.; Lin, C.C.; Chang, H.T. A cognitive oriented framework for IoT big-data management prospective. In *Proceedings of the 2014 IEEE International Conference on Communication Problem-solving*, Beijing, China, 5–7 December 2014; pp. 124–127.
43. Amadeo, M.; Campolo, C.; Molinaro, A. Multi-source data retrieval in IoT via named data networking. In *Proceedings of the 1st ACM Conference on Information-Centric Networking*, Paris, France, 24–26 September 2014; pp. 67–76.
44. Yan, Z.; Zhang, P.; Vasilakos, A.V. A survey on trust management for Internet of Things. *J. Netw. Comput. Appl.* 2014, 42, 120–134.
45. Khan, R.; Khan, S.U.; Zaheer, R.; Khan, S. Future Internet: The Internet of Things Architecture, Possible Applications and Key Challenges. In *Proceedings of the International Conference on Frontiers of Information Technology*, Islamabad, Pakistan, 17–19 December 2012.
46. Bor, M.; Vidler, J.; Roedig, U. *LoRa for the Internet of Things*; Junction Publishing: Graz, Austria, 2016.
47. Alrae, R.; Nasir, Q.; Abu Talib, M. Developing House of Information Quality framework for IoT systems. *Int. J. Syst. Assur. Eng. Manag.* 2020, 11, 1294–1313.
48. Perez-Castillo, R.; Carretero, A.G.; Rodriguez, M.; Caballero, I.; Piattini, M.; Mate, A.; Kim, S.; Lee, D. Data quality best practices in IoT environments. In *Proceedings of the 2018 11th International Conference on the Quality of Information and Communications Technology (QUATIC)*, Coimbra, Portugal, 4–7 September 2018; pp. 272–275.
49. Sathe, S.; Papaioannou, T.G.; Jeung, H.; Aberer, K. A survey of model-based sensor data acquisition and management. In *Managing and Mining Sensor Data*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 9–50.
50. Erguler, I. A potential weakness in RFID-based Internet-of-things systems. *Pervasive Mob. Comput.* 2015, 20, 115–126.
51. Jeffery, S.R.; Garofalakis, M.; Franklin, M.J. Adaptive Cleaning for RFID Data Streams. 2006, Volume 6, pp. 163–174. Available online: <https://www2.eecs.berkeley.edu/Pubs/TechRpts/2006/EECS-2006-29.pdf> (accessed on 1 June 2021).
52. Said, O.; Masud, M. Towards internet of things: Survey and future vision. *Int. J. Comput. Netw.* 2013, 5, 1–17.
53. Ukil, A.; Sen, J.; Koilakonda, S. Embedded security for Internet of Things. In *Proceedings of the 2011 2nd National Conference on Emerging Trends and Applications in Computer Science*, Shillong, India, 4–5 March 2011; pp. 1–6.
54. Zeng, D.; Guo, S.; Cheng, Z. The web of things: A survey. *JCM* 2011, 6, 424–438.
55. Benabbas, A.; Nicklas, D. Quality-Aware Sensor Data Stream Management in a Living Lab Environment. In *Proceedings of the 2019 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2019*, Kyoto, Japan, 11–15 March 2019; pp. 445–446.
56. Wang, Z.; Talburt, J.R.; Wu, N.; Dagtas, S.; Zozus, M.N. A Rule-Based Data Quality Assessment System for Electronic Health Record Data. *Appl. Clin. Inform.* 2020, 11, 622–634.
57. Teh, H.Y.; Kempa-Liehr, A.W.; Kevin, I.; Wang, K. Sensor data quality: A systematic review. *J. Big Data* 2020, 7, 11.
58. Li, Y.; Parker, L.E. Nearest neighbor imputation using spatial-temporal correlations in wireless sensor networks. *Inf. Fusion* 2014, 15, 64–79.
59. Song, S.; Zhang, A. IoT Data Quality. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, Galway, Ireland, 19–23 October 2020; pp. 3517–3518.
60. Eppler, M.J.; Wittig, D. Conceptualizing Information Quality: A Review of Information Quality Frameworks from the Last Ten Years. *IQ* 2000, 20, 83–96.

61. Micic, N.; Neagu, D.; Campean, F.; Zadeh, E.H. Towards a data quality framework for heterogeneous data. In Proceedings of the 2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Exeter, UK, 21–23 June 2017; pp. 155–162.
62. Khokhlov, I.; Reznik, L.; Chuprov, S. Framework for integral data quality and security evaluation in smartphones. *IEEE Syst. J.* 2020.
63. Wang, R.Y. A product perspective on total data quality management. *Commun. ACM* 1998, 41, 58–65.
64. Scannapieco, M.; Virgillito, A.; Marchetti, C.; Mecella, M.; Baldoni, R. The DaQuinCIS architecture: A platform for exchanging and improving data quality in cooperative information systems. *Inf. Syst.* 2004, 29, 551–582.
65. Cappiello, C.; Ficiaro, P.; Pernici, B. HIQM: A methodology for information quality monitoring, measurement, and improvement. In *International Conference on Conceptual Modeling*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 339–351.
66. Caballero, I.; Verbo, E.; Calero, C.; Piattini, M. MMPRO: A Methodology Based on ISO/IEC 15939 to Draw Up Data Quality Measurement Processes. 2008, pp. 326–340. Available online: [https://d1wqtxts1xzle7.cloudfront.net/66879043/MMPRO\\_A\\_Methodology\\_Based\\_on\\_ISOIEC\\_159320210504-20048-24vo05-with-cover-page-v2.pdf?Expires=1630310271&Signature=e7pa3a0Xk2RSp3J27hc84urGqh7Hc1iUxHJR-W-Ur4A5mMgJeLugkAuqaFLeFmRExAA6a-kEw-jyKfz6Li-lfJt-cjiEUEQNcQ9nlueRpFwGel-X8uyyZc7mgTuM4ysE0gTDPAO68IHxprmsaYXUANKFoJ1ydKD7tgXm42SUK9vljydkst4MsN6UklzGvSIFPair-Id=APKAJLOHF5GGSLRBV4ZA](https://d1wqtxts1xzle7.cloudfront.net/66879043/MMPRO_A_Methodology_Based_on_ISOIEC_159320210504-20048-24vo05-with-cover-page-v2.pdf?Expires=1630310271&Signature=e7pa3a0Xk2RSp3J27hc84urGqh7Hc1iUxHJR-W-Ur4A5mMgJeLugkAuqaFLeFmRExAA6a-kEw-jyKfz6Li-lfJt-cjiEUEQNcQ9nlueRpFwGel-X8uyyZc7mgTuM4ysE0gTDPAO68IHxprmsaYXUANKFoJ1ydKD7tgXm42SUK9vljydkst4MsN6UklzGvSIFPair-Id=APKAJLOHF5GGSLRBV4ZA) (accessed on 1 June 2021).
67. ISO. ISO/IEC/IEEE 15939:2017 Systems and Software Engineering—Measurement Process; Standard, ISO/IEC JTC 1/SC 7 Software and Systems Engineering; ISO: Geneva, Switzerland, 2017.
68. Angeles, M.D.P.; García-Ugalde, F. A Data Quality Practical Approach. *Int. J. Adv. Softw.* 2009, 2, 259–274.
69. Pipino, L.L.; Lee, Y.W.; Wang, R.Y. Data quality assessment. *Commun. ACM* 2002, 45, 211–218.
70. Tayi, G.K.; Ballou, D.P. Examining data quality. *Commun. ACM* 1998, 41, 54–57.
71. Motro, A.; Rakov, I. Estimating the quality of databases. In *International Conference on Flexible Query Answering Systems*; Springer: Berlin/Heidelberg, Germany, 1998; pp. 298–307.
72. Carretero, A.G.; Freitas, A.; Cruz-Correia, R.; Piattini, M. A Case Study on Assessing the Organizational Maturity of Data Management, Data Quality Management and Data Governance by Means of MAMD. 2016, pp. 75–84. Available online: [https://alarcos.esi.uclm.es/iciq2016/documents/camera\\_ready/9-mamd-iciq2016.pdf](https://alarcos.esi.uclm.es/iciq2016/documents/camera_ready/9-mamd-iciq2016.pdf) (accessed on 1 June 2021).
73. Sundararaman, A.; Venkatesan, S.K. Data quality improvement through OODA methodology. In Proceedings of the 22nd MIT International Conference on Information Quality, ICIQ, Rock, AR, USA, 6–7 October 2017; pp. 1–14.
74. ISO. ISO/TS 8000-60:2017 Data Quality—Part 60: Data Quality Management: OVERVIEW; Standard, International Organization for Standardization/TC 184/SC 4 Industrial Data (2017); ISO: Geneva, Switzerland, 2017.
75. ISO. ISO/TS 9000:Quality Management Systems; Standard, ISO/IEC JTC 1/SC 7 Software and Systems Engineering; ISO: Geneva, Switzerland, 2017.
76. Tan, Z.; Wei, H.; Yong, S. ISO 8000 (big) data quality standard and application. *Big Data Res.* 2017, 3, 2017001.
77. ISO. ISO/IEC 25012:2008 SOFTWARE Engineering—Software Product Quality Requirements and Evaluation (SQuaRE)—Data Quality Model; Standard, International Organization for Standardization/ISO/IEC JTC 1/SC 7 Software and Systems Engineering (2007); ISO: Geneva, Switzerland, 2008.
78. ISO. ISO/IEC 25024:2015 Systems and Software Engineering—Systems and Software Quality Requirements and Evaluation (SQuaRE)—Measurement of Data Quality; Standard, International Organization for Standardization/ISO/IEC JTC 1/SC 7 Software and Systems Engineering (2015); ISO: Geneva, Switzerland, 2015.

Retrieved from <https://encyclopedia.pub/entry/history/show/34975>