International Telecommunication Union
Standardization Sector

Recommendation

ITU-T Y.4603 (03/2023)

SERIES Y: Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities

Internet of things and smart cities and communities – Services, applications, computation and data processing

---

Requirements and functional model to support data quality management in Internet of things

ITU-T Y-SERIES RECOMMENDATIONS

GLOBAL INFORMATION INFRASTRUCTURE, INTERNET PROTOCOL ASPECTS, NEXT-GENERATION NETWORKS, INTERNET OF THINGS AND SMART CITIES

GLOBAL INFORMATION INFRASTRUCTURE	
General	Y.100–Y.199
Services, applications and middleware	Y.200–Y.299
Network aspects	Y.300–Y.399
Interfaces and protocols	Y.400–Y.499
Numbering, addressing and naming	Y.500–Y.599
Operation, administration and maintenance	Y.600–Y.699
Security	Y.700–Y.799
Performances	Y.800–Y.899
INTERNET PROTOCOL ASPECTS	
General	Y.1000–Y.1099
Services and applications	Y.1100–Y.1199
Architecture, access, network capabilities and resource management	Y.1200–Y.1299
Transport	Y.1300–Y.1399
Interworking	Y.1400–Y.1499
Quality of service and network performance	Y.1500–Y.1599
Signalling	Y.1600–Y.1699
Operation, administration and maintenance	Y.1700–Y.1799
Charging	Y.1800–Y.1899
IPTV over NGN	Y.1900–Y.1999
NEXT GENERATION NETWORKS	
Frameworks and functional architecture models	Y.2000–Y.2099
Quality of Service and performance	Y.2100–Y.2199
Service aspects: Service capabilities and service architecture	Y.2200–Y.2249
Service aspects: Interoperability of services and networks in NGN	Y.2250–Y.2299
Enhancements to NGN	Y.2300–Y.2399
Network management	Y.2400–Y.2499
Computing power networks	Y.2500–Y.2599
Packet-based Networks	Y.2600–Y.2699
Security	Y.2700–Y.2799
Generalized mobility	Y.2800–Y.2899
Carrier grade open environment	Y.2900–Y.2999
FUTURE NETWORKS	Y.3000–Y.3499
CLOUD COMPUTING	Y.3500–Y.3599
BIG DATA	Y.3600–Y.3799
QUANTUM KEY DISTRIBUTION NETWORKS	Y.3800–Y.3999
INTERNET OF THINGS AND SMART CITIES AND COMMUNITIES	
General	Y.4000–Y.4049
Definitions and terminologies	Y.4050–Y.4099
Requirements and use cases	Y.4100–Y.4249
Infrastructure, connectivity and networks	Y.4250–Y.4399
Frameworks, architectures and protocols	Y.4400–Y.4549
Services, applications, computation and data processing	**Y.4550–Y.4699**
Management, control and performance	Y.4700–Y.4799
Identification and security	Y.4800–Y.4899
Evaluation and assessment	Y.4900–Y.4999

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T Y.4603

Requirements and functional model to support data quality management in Internet of things

Summary

Data quality management comprises the mature processes, tools, and in-depth understanding of data you need to make decisions or solve problems to minimize risk and impact to your organization or customers. Data quality management in Internet of things (IoT) is the practice of using that IoT data to serve your purposes with flexibility and agility for IoT applications. To do this, it is necessary to assess what data you have today and the processes and tools that use or support data against purposes and requirement of IoT applications. The requirements for data and its quality vary from IoT application to application or organization in different contexts. Data quality management practice in IoT makes data a holistic asset, by which it means that data is the input and output of every task and transaction performed according to the IoT application for a business.

Recommendation ITU-T Y.4603 specifies key requirements with respect to data quality management in IoT and important elements to fulfil these requirements. This Recommendation indicates the requirements and functional model in terms of the scopes of the following: data quality management in IoT, requirements of data quality management in IoT and a functional model to support data quality management.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T Y.4603	2023-03-28	20	11.1002/1000/15487

Keywords

Data quality, data quality classification, data quality management, IoT.

* To access the Recommendation, type the URL http://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID. For example, http://handle.itu.int/11.1002/1000/11830-en.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents/software copyrights, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the appropriate ITU-T databases available via the ITU-T website at http://www.itu.int/ITU-T/ipr/.

Table of Contents

Recommendation ITU-T Y.4603

Requirements and functional model to support data quality management in Internet of things

1 Scope

This Recommendation identifies requirements and a functional model for data quality management in Internet of things (IoT). The scope of this Recommendation covers several key requirements with respect to data quality management in IoT, many important elements to fulfil these requirements and a functional model. Specifically, it covers the following:

– Overview of data quality;

– Data quality management in IoT;

– Requirements of data quality management in IoT;

– Functional components to support data quality management.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T Y.4000] Recommendation ITU-T Y.4000/Y.2060 (2013), *Overview of the Internet of things*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following term defined elsewhere:

3.1.1 Internet of things (IoT) [ITU-T Y.4000]: A global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 critical data: The data to be required for the successful completion of a task in an application or service within a specific business context. It is characterized by defined attributes and priorities.

3.2.2 data quality: The degree to which the characteristics of data satisfy stated and implied needs when used under specified conditions.

3.2.3 non-critical data: Low business value and impact data.

3.2.4 poor data: The data classified as low quality during a data quality assessment process.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

AI Artificial Intelligence

DL Deep Learning

DQM Data Quality Management

IoT Internet of Things

ML Machine Learning

PII Personally Identifiable Information

5 Conventions

In this Recommendation:

The keywords **"is required to"** indicate a requirement which must be strictly followed and from which no deviation is permitted if conformance to this document is to be claimed.

6 Overview of data quality

6.1 Concept of data quality

For data that is collected through sensing to be fit for use, it should possess certain features so that it can satisfy a set of system requirements.

There is an increasing awareness of the criticality of data to making informed decisions and of how inaccurate data can lead to disastrous consequences. The challenge lies in ensuring that enterprises collect relevant data for their businesses, manage or govern that data in a meaningful and sustainable way, ensure high quality for critical data, and analyse the high quality data to accomplish stated business objectives.

The data quality will be attributed to a quality supporting business processes using analysis techniques, and it will rise according to whether a data quality of the existing data would be worthwhile or plausible. Thus, data quality is necessarily designed, measured and assessed with this consideration in mind, and the possible risks of deploying this data in business will be overcome.

6.2 Data quality issues

Data quality is very important for product or information consumers because without high data quality, a business operator or service provider will not be able to provide the requested service capabilities. This is because the characteristics of the data are not related to satisfying the stated requirements of a service. Furthermore, data quality cannot be specified as the degree to which the characteristics of the data satisfy the stated and implied needs when used under specified conditions.

Many data quality issues have been raised, and common issues are listed in Table 6-1.

Table 6-1 – Common issues in data quality

Data quality issues	Description
Loss of event	Gaps in the data, such as the data for some time period is missing
Out of range values	The physiological sensors' values are out of range with respect to the domain. These issues could be in individual time series data and in aggregated and mean data.

Table 6-1 – Common issues in data quality

Data quality issues	Description
Value spikes	The values changed suddenly and are considered incorrect from a sensor and domain.
Wrong timestamp	Differing timestamps from sensors in the same location
Signal noise	Errors in the measurement of states
Data not updated as required	Sensors continuously send the same value over a long time. This may indicate some issues in the sensors or no change in the real-word event
Un-availability of data sources	The source of data (such as sensors) is not available for the service
Errors in correlated data	Values which are normally correlated but behave unexpectedly
Data formats	Due to different data formats such as different units, or different measurement frequencies, creates problems in data aggregation and takes more processing time
Class imbalance	Biasness in the data samples which affects the accuracy of machine learning deep learning models
Inconsistent data	The level of data errors and noise changes over time from different sensors
Duplicated data	Same data reception over a long time from one data source or multiple data sources
Different precision levels	Different precision levels for the same type of data in various business services

6.3 Data quality classification

Based on the concept of data quality, the data quality classification has been identified as shown in Figure 6-1:

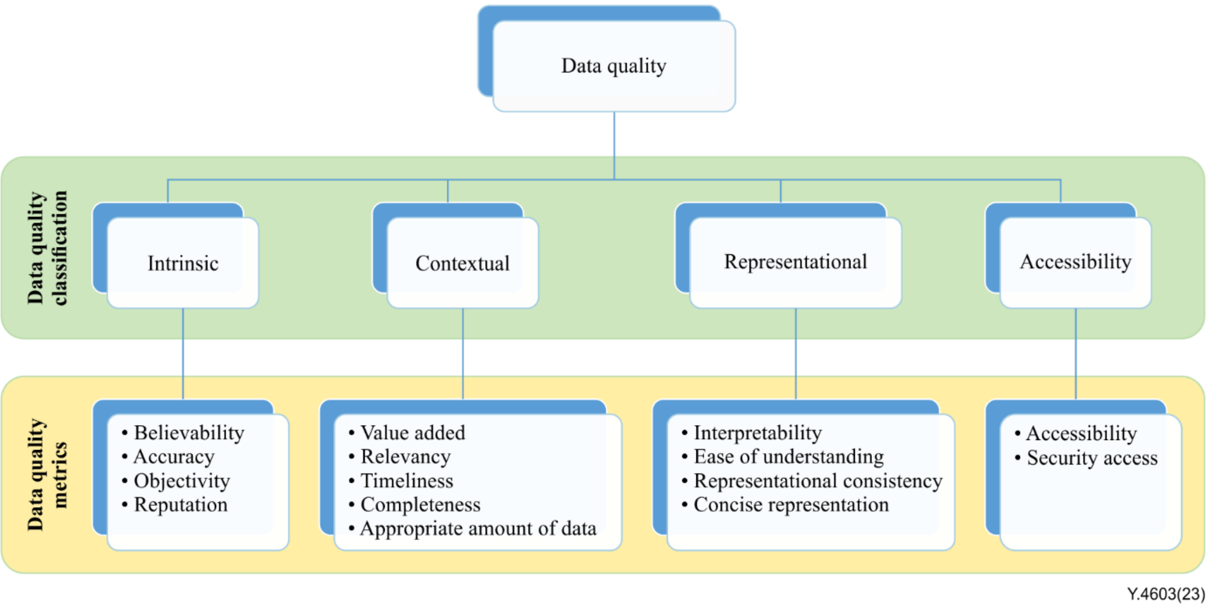

Figure 6-1 – Classification of data quality

6.3.1 Intrinsic data quality

The intrinsic data quality that is related to the actual values of data regardless of the context or data components (e.g., accuracy).

Intrinsic data quality metrics:

– Believability: The believability is the extent to which data are accepted or regarded as true, real, and credible;
– Accuracy: The degree to which data correctly describes the "real world" object or event being described;
– Objectivity: The extent to which data is unbiased, unprejudiced, and impartial;
– Reputation: The extent to which data is highly regarded in terms of its source or content.

6.3.2 Contextual data quality

The contextual data quality is related to other data components within a certain context (e.g., completeness, timeliness, and consistency).

Contextual data quality metrics:

– Value added: The extent to which data is beneficial and provides advantages from its use;
– Relevancy: The extent to which data is applicable and helpful for the task at hand;
– Timeliness: The degree to which data represents reality from the required point in time;
– Completeness: The proportion of stored data against the potential of "100% complete";
– Appropriate amount of data: This means that a sufficient amount of data is available for use to compute a result of data items.

6.3.3 Representational data quality

The representational data quality captures aspects related to the design of the data.

Representational data quality metrics:

– Interpretability: The extent to which data is in appropriate languages, symbols, units, and the definitions are clear;
– Ease of understanding: The ease of understanding means the extent to which data is easily comprehended;
– Representational consistency: The extent to which a whole dataset is presented in the data structure and format;
– Concise representation: The extent to which data is compactly represented.

6.3.4 Accessibility data quality

This category deals with the data quality aspects related to data infrastructure such as accessibility, security access, data retrieval.

Representational data quality metrics:

– Accessibility: The extent to which data is available, or easily and quickly retrievable;
– Security access: The security access metrics measures that the data is safeguarded from unauthorized access and preventing data loss.

7 Data quality management in IoT

7.1 Overview of data quality management

Due to advanced information and communication technologies, a very large amount of data is generated in the operation, production, and management of IoT business. However, the quality and reliability of that data is questionable. This newly collected data from such sources and already available data in the organization are pre-processed in order to check the data completeness, accuracy, and consistency for improved and high quality decision making services. Data quality management is an administration type that incorporates the role establishment, role deployment, policies, responsibilities and processes with regard to the acquisition, maintenance, disposition and distribution of data. Data quality management initiative promotes a strong partnership between technology groups and business. Building and controlling the entire data environment, that is, architecture, systems, technical establishments, and databases will be promoted through the overall environments to acquire, maintain, disseminate and dispose of an organization's electronic data assets. Key motivations for data quality management in IoT:

– Data quality management is the set of processes and activities of combing all of the entities in an organization including people, technology and culture to promote the common goal of data quality provision to IoT data consumers;
– Data quality management is an open system where data quality professionals need to interact with data consumers and other data stakeholders freely in order to identify the information needs and data quality requirements;
– Data quality management is also crucial when it comes to choosing one dataset for business IoT applications over other available datasets;
– Data quality will be attributed to a quality supporting business processes and analysis techniques, and it will be raised according to whether a data quality of the existing data is worthwhile or plausible;
– Information in data applications can be derived from the processing of data to give meaning and sense to it with multiple types of metadata.

To achieve data quality for IoT applications, it is necessary to collect and pre-process the data, perform a quality assessment of existing data, evaluate the results and improve it, and if necessary, perform a ranking and monitoring of data for sustainable data quality management:

– Data acquisition: The data acquisition supports internal and external data with consistency, storage efficiency, retrieval efficiency and security efficiency.
– Data quality assessment: To assess the existing data quality, common data quality metrics are applied;
– Data quality evaluation: The data quality evaluation focuses on the economic advantages or technical feasibility to choose data quality improvement solutions;
– Data quality improvement: To improve the data quality a designed data quality plan is applied for the data quality improvement;
– Data quality ranking: Rank the data as per achieved data quality level, the ranking of data for further use in the business services.

In order to support ongoing data quality monitoring and management, an additional process to monitor data quality will be prepared in the data quality management function.

An overview of various data quality management process is shown in Figure 7-1.

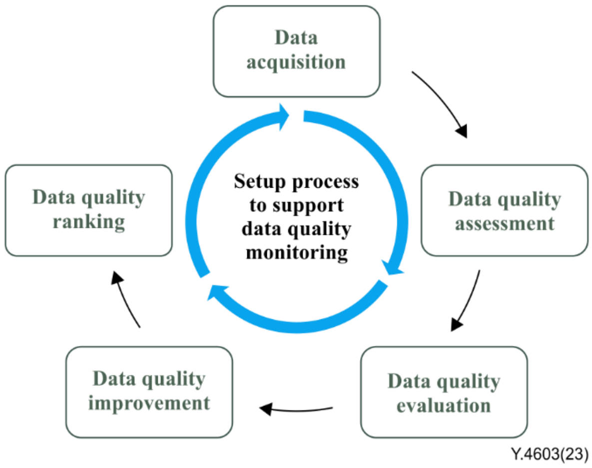

Figure 7-1 – Overview of data quality management

7.2 Data quality management characteristics

Based on the data quality concept and data quality management overview, a data quality management framework is necessary to comprise the following characteristics:

– The quality of data should be managed in a way that data should be accurate, available, consistent, confident, integrated, relevant, reusable, current, and complete;
– The quality of data is measured with relevant data quality metrics;
– The data quality management framework contains data provenance information;
– The management system contains methodologies to maintain the metadata;
– Data quality management has the capability to manage reference data;
– The system supports the process to improve data quality if the received data has poor quality;
– A central data modelling system is provided in the data quality management framework;
– The ranking of the data with respect to its data quality indicators is supported.

8 Requirements to support data quality management in IoT

To support the business goal of data quality management in the IoT, many aspects are required to be considered. Based on clause 7.1, several types of requirements are considered to support data quality management:

– General requirements of data quality management;
– Requirements in the data acquisition phase;
– Requirements in the data quality assessment phase;
– Requirements in the data quality evaluation phase;
– Requirements in the data quality improvement phase;
– Requirements in the data quality ranking phase;
– Requirements in the data quality monitoring phase.

8.1 General requirements of data quality management

To receive the appropriate level of data quality the following general aspects are required:

– The system is required to be scalable in order to handle the growing large volume of IoT data;
– The data quality management is required to be available during the processing of data;

– It is required to define the set of validations that need to be developed to measure the quality of data;
– The data quality management process is required to be interoperable in order to handle data received from heterogeneous sources;
– For ongoing improvement in the data quality, the system is required to be able to monitor data quality over some length of time;
– Privacy of users is required to be considered during the processing of data in order to estimate and improve the data quality;
– Privacy protection is required to be performed during the data collection process;
– In order to increase the data quality completeness aspects, handling of missing data with sophisticated techniques is required to be provided;
– To increase the accuracy and confidence of the data, the data quality management function is required to detect and correct false and suspect data;
– It is required to support the ability to produce ad hoc reports indicating gaps in data for any attributes selected;
– It is required to support the ability to identify the number of data points that are not populated for each required attribute field;
– It is required that the data collection scheduler balances the data collection interval in order to increase efficiency (for example: the battery life of IoT sensing objects);
– The data quality management framework is required to support time alignment in the collection of data in the case of a multi-sources data collection request for the IoT service provision;
– To estimate and improve the quality of data efficiently, the data collection methodologies in the data quality framework are required to support the collection of metadata along with the data;
– The data quality management framework is required to support a coordinated workflow mechanism in order to handle the collection of duplicated data.

8.2 Requirements in data acquisition phase

To acquire the IoT data for business needs, the data quality management framework is required to support the following functionalities:

– It is required to ensure that no extra data is collected with respect to the required policy. The collection of extra data reduces customer's privacy and introduces the data leakage issues of data quality;
– Acquiring and validation of data from external sources;
– Acquiring and validation of data from internal sources;
– Data masking during the data acquiring phase in order to mitigate personally identifiable information (PII) aspects.

8.3 Requirements in data quality assessment phase

To measure the quality of data with optimal precision, data quality management framework is required to support the following functionalities:

– Due to the large volume of IoT data, it is difficult to estimate the quality of data within a reasonable amount of time, therefore the data quality management framework is required to be efficient and scalable to support efficient data quality assessment;

– As there is no unified standard to estimate the quality of data for various IoT applications, the framework is therefore required to support measurement of the quality of data with numerous data quality metrics;
– The framework is required to support the estimation of data quality with common metrics initially;
– The data quality management framework is required to support a basic set of data quality measurement methodologies;
– It is required to support the addition of a new data quality indicator and measurement methodology in the framework;
– The measurement of data quality is required to be considered according to various interest groups;
– It is required to provide an update in the measurement methodology of existing data quality indicators;
– Data quality assessment function is required to support individual quality measurement for each data source;
– Data quality assessment function is required to support aggregated measurement for data received from multiple sources;
– Periodic data quality assessment capability is required to be supported for non-critical data;
– Continuous data quality assessment function is required to be supported for critical data.

8.4 Requirements in data quality evaluation phase

The evaluation of data quality management procedures is mandatory because it also focuses on the economic advantages or technical feasibility to choose data quality improvement solutions. To ensure a suitable level of data quality evaluation, the following functionalities are required to be provided by the data quality management framework:

– The system is required to locate critical areas of data which affect the quality of data;
– The system is required to optimize the time used in the data quality assessment;
– The system is also required to check the data reputation after its usage in the IoT applications so that other services can get the benefit;
– The system is required to evaluate the direct and indirect costs of the data quality process;
– The system is required to check that the results of the data quality assessment are up to standard for the IoT applications and services;
– The system is required to recommend improving the grey areas of data which weaken the level of data quality;
– The system is required to support evaluation of the business rules from time to time defined against the various data sources.

8.5 Requirements in data quality improvement phase

To ensure the choice of suitable methodologies to improve data quality, the following functionalities are required to be provided by the data quality management framework:

– The system is required to be able to perform data various types of data interpolation in order to handle missing data;
– The system is required to support detection and correction of data outliers in the streaming data;
– The system is required to be able to rectify and improve the data affected by malware attacks;

– It is required that the framework should support data deduplication;

– Data is required to be transformed in encrypted format in order to increase security and confidence of data;

– The system is required to support the common data representation formats in order to increase the data interpretability;

– The data quality management model is required to have high power infrastructure in order to support data availability in peak hours and emergency situations;

– The system is required to have a predefined data threshold in order to validate data accuracy;

– The data quality management framework is required to support a proactive approach to improve the data quality if possible.

8.6 Requirements in data quality ranking phase

To ensure a suitable level of data quality ranking, the following functionalities are required to be provided by the data quality management framework:

– In consideration of fitness for use for the task at hand, the system is required to assign weights to the preferred data quality metrics for measuring the ranking of data;

– In the IoT services sometimes the data of an individual source is used and other times data from multiple sources are used collectively. Therefore, the data quality management framework is required to have the provision of individual aggregative data quality ranking;

– For ongoing improvement and provision of services, the quality of data is required to be ranked in different time slots;

– The data quality management framework is required to support estimated or predicted data quality ranking if the actual data quality ranking is under processing, in order to support prediction of IoT service behaviour modelling;

– It is required to provide the reporting of the quality of data in terms of its business usages and technical rankings;

– For the sustainable data quality management, it is required to support monitoring and improvement of the data at the source level.

8.7 Requirements in data quality monitoring phase

To support ongoing data quality monitoring, the following functionalities are required to be provided by the data quality management framework:

– It is required to support periodic monitoring that provides feedback on the data quality management process and enable dynamic tuning;

– The system is required to generate alerts when the level of data quality decreases below a certain threshold;

– The data quality management framework is required to have the capability of periodical reporting of quality;

– For the sustainable data quality management, the system is required to support monitoring of the quality of data at the source level;

– The data quality management framework is required to support multiple schedulers to monitor ongoing data quality;

– The framework is required to support appropriate visualization data quality monitoring results;

– The framework is required to support tracking levels of data quality over time for monitoring of the ongoing process and improvement.

9 Functional component to support data quality management

The functional components of data quality management provide detailed mechanisms to support the data quality management of IoT. Data quality management is processed in accordance with the functions in the data quality management phases and data quality classifications as indicated Figure 9-1. Figure 9-1 indicates the functional components to support data quality management in accordance with data quality classification and business objectives presented in Figure 7-1.

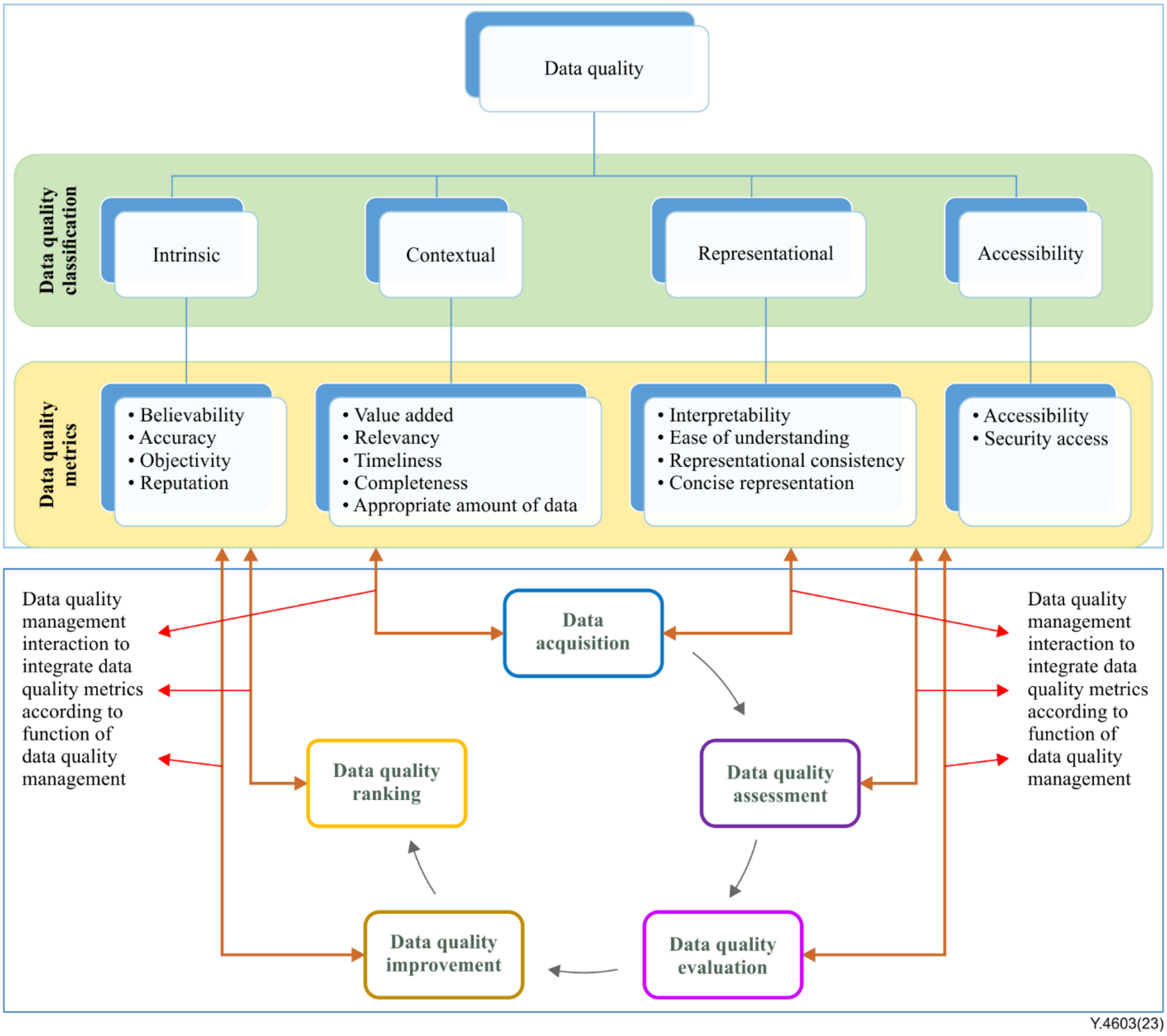

Figure 9-1 – Functional components to support data quality management in accordance with data quality classification and business objectives

9.1 Data acquisition function

The data acquisition function supports bringing of data that has been created by a source outside the organization, into the organization, for business use. The data acquisition function provides internal and external data with consistency, storage efficiency, retrieval efficiency and security efficiency. Figure 9-2 shows the four sub-functions to perform data acquisition capability with respect to data quality. In the defining of business goals in terms of data of IoT applications, it is necessary to clarify a correlation between the business processes and data quality management.

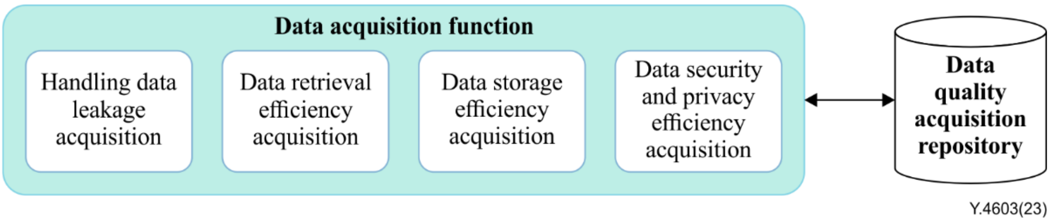

Figure 9-2 – Functional components of data acquisition function

– Handling data leakage acquisition: Data leakage issues occur when the IoT service or application acquires more data than the necessary requirements. This sub-function handles these issues when data acquisition has been performed in the data quality management platform;

– Data retrieval efficiency acquisition: This sub-function provides the capability to acquire data from the repository within an appropriate time. The data availability and timeliness aspects of data could be improved through the sub-function of data retrieval efficiency acquisition;

– Data storage efficiency acquisition: This sub-function supports the mechanism to store the acquired data within an appropriate time and consistent format;

– Data security and privacy efficiency acquisition: This sub-function provides the capability to acquire a data security and privacy level.

9.2 Data quality assessment function

The data quality assessment function offers methodologies to estimate the data quality of received and existing IoT data. The quality of data is measured with various data quality metrics. Figure 9-3 shows the four sub-functions of the data quality assessment function. According to the complexity of this function, machine learning or deep learning mechanisms may be applied to perform data quality assessment as presented in Appendix I.

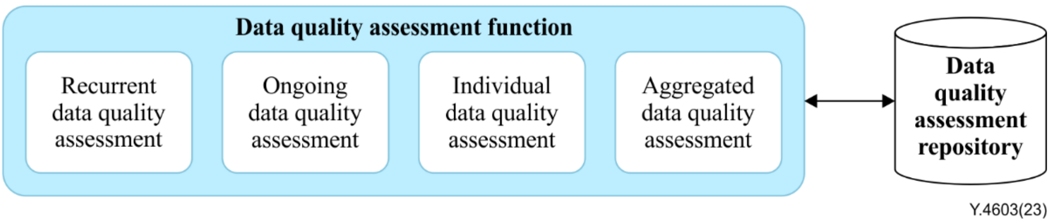

Figure 9-3 – Functional components of data quality assessment function

– Recurrent data quality assessment: The recurrent data quality assessment function supports the measurement of IoT data quality after each specific interval. This sub-function checks the quality of data with respect to various aspects as defined in the function template;

– Ongoing data quality assessment: This sub-function checks the quality of data received from IoT objects periodically. The main capability of this ongoing data quality assessment sub-function is to ensure the validation of the quality of data received from critical IoT data sources. This function also supports priority assessment among the critical data categories;

– Individual data quality assessment: This function provides the mechanism to check and measure the quality of data of each source separately. Another capability of this function is that the individual data quality assessment function can be applied as a standalone or it can be used collectively with recurrent and ongoing data quality assessment functions;

– Aggregated data quality assessment: In order to use data from various IoT data sources in a single service, the IoT service requires the aggregated data quality level for all of the data used in the service. The aggregated data quality checking function provides this capability.

In aggregated data quality checking, good or poor quality affects the overall checking results.

9.3 Data quality evaluation function

The data quality evaluation function provides capabilities in accordance with data quality classification and characteristics to support data quality management. Figure 9-4 shows the three sub-functions of the data quality evaluation function. According to the complexity of this function, machine learning or deep learning mechanisms may be applied to perform data quality evaluation as presented in Appendix I.

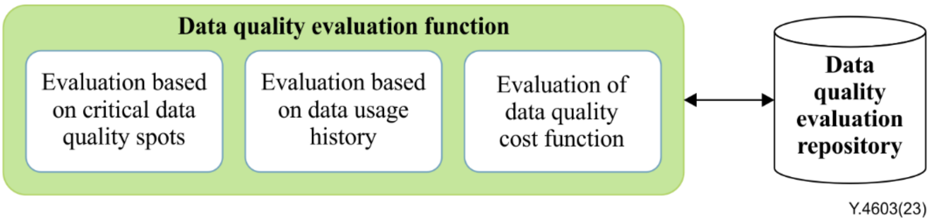

Figure 9-4 – Functional components of Data Quality Evaluation Function

– Evaluation based on critical data quality spots: This sub-function evaluates areas which hamper the overall quality of a dataset and analyses the root causes in order to avoid issues in future data;

– Evaluation based on data usage history: This sub-function enables the capability of data quality management framework to evaluate the quality of data based on its usage in the IoT service. The feedback from the users could be the input to make the evaluation of the quality of the data;

– Evaluation of data quality cost function: This sub-function performs the evaluation of the data quality process. The main focus of this sub-function is to measure the people and computational time, and system memory by considering data quality aspects, because the data quality assessment process requires high computation power and large memory due to the large volume of IoT data.

9.4 Data quality improvement function

The data quality improvement function provides the enhancement of the data spots whose data quality has been estimated and evaluated as poor. Figure 9-5 shows the five sub-functions of data quality improvement function.

– Data quality constraints validation: The data quality constraints validation function enables to detect missing interrelations in the IoT data. For the identification of missing interrelations in the data, this function uses the reference of dependency and integrity constraints. Due to the functional capability of this function, the data quality aspects towards data consistency and data representation could be improved in the data quality management framework;

– Data outliers: This outliers function includes identifying those values from a dataset or a single data source which are not coming within a certain range. As the data outliers is a major data quality issue in the IoT environment, by analysing and fixing data outliers, the quality of data in terms of data accuracy, and data consistence could be enhanced;

– Data interpolation: The interpolation function is an estimation of a value within two known values in a sequence of values. The data interpolation functions enable the various mechanism to improve the quality of data in terms of data completeness and accuracy. This function applies various methodologies to estimate the missing data in the IoT data streams;

– Data deduplication: The deduplication sub-function performs the process of data cleaning. The data deduplication sub-function provides the mechanism to reduce a large volume of duplicate data by detecting the same copy or instance of data for similar real-world events. The data pointers are supported by this function which refers to the unique copy of data. This sub-function enhances the data availability aspects of data quality management and reduces the size of data storage and data management efforts;

– Data representation: The data representation sub-function enables improvement of the quality of data in terms of data interoperability and provides the mechanism to translate and transform the data in the common representation and storage format. The advantages of this sub-function are the improvement of data consistency, data availability, and data accuracy.

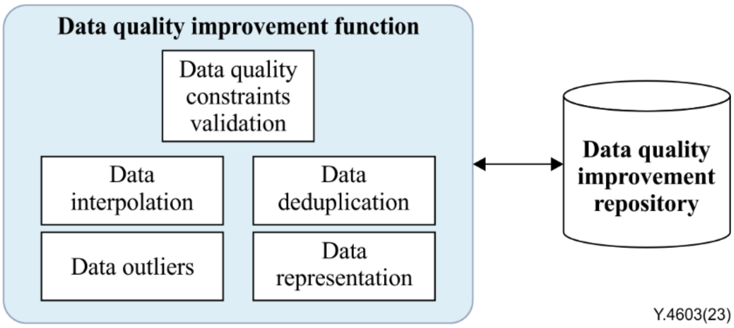

Figure 9-5 – Functional components of data quality improvement function

9.5 Data quality ranking capability

To ensure the suitable level of data quality ranking, three sub-functions of data quality ranking function perform the enhancement of the business value of the data in the data applications. Figure 9-6 shows the three sub-functions of the data quality ranking function.

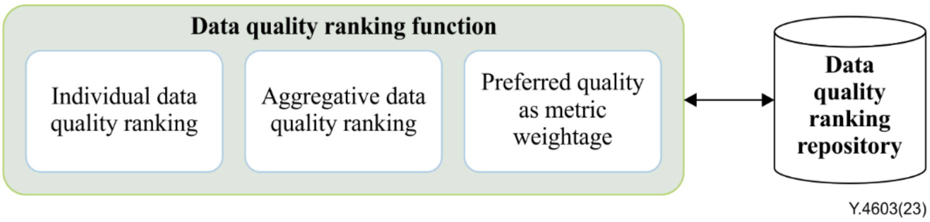

Figure 9-6 – Functional components of data quality ranking function

– Individual data quality ranking: This sub-function enables ranking of the data separately as per the data quality checking results with respect to each data source and data quality aspects such as data accuracy, data completeness, data consistency, and data availability. This sub-function provides the level of ranking to individual data applications of IoT;

– Aggregative data quality ranking: This sub-function supports the ranking of data quality collectively of all the data received from multiple sources. Further, this sub-function enables business-oriented data quality ranking such as location, time, and also data quality ranking;

– Preferred quality as metric weightage: To create and update application specific data quality ranking templates, this sub-function enables assigning of data quality weightage to preferred data quality metrics, so that the data could be categorized and used in the specific context.

9.6 Ongoing data quality monitoring function

To support the ongoing data quality monitoring for sustainable data quality, the data quality monitoring function provides four sub-functions to monitor data quality at each stage in the data quality management. Figure 9-7 shows the four functional components (sub-functions) of the data quality monitoring function for ongoing data quality monitoring.

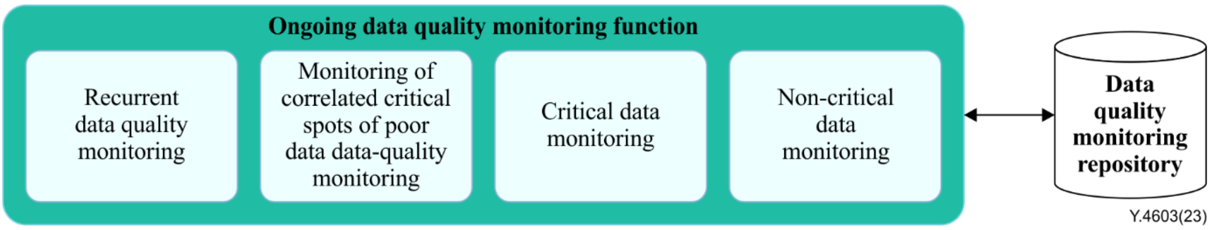

Figure 9-7 – Functional components of the ongoing data quality monitoring function

– Recurrent data quality monitoring: The recurrent data quality monitoring function ensures the monitoring of various aspects of data quality periodically. This sub-function supports various components which create many schedules for the checking of the entire process of data quality management systems. The choice of data attributes to monitor, and how to monitor them, is one of the key decisions of the design phase of the data quality management;

– Monitoring of correlated critical spots of poor data: The monitoring correlated critical spots of poor data enables to monitor the quality of data and related data whose quality estimated as weak in the data quality evaluation and ranking function. This sub-function generates reports which show the trends of correlated critical data quality improvement or deterioration from time to time;

– Critical data monitoring: This sub-function of critical data monitoring supports an important issue in the data quality management. The critical data are more important than the rest of the data in the data quality management framework such as patient surveillance data, etc. The monitoring mechanism in this sub-function focuses on critical data defined as per IoT applications;

– Non-critical data monitoring: Non-critical data also has business value and impact but not to the same extent as critical data. Similarly, as in the monitoring of critical quality data, the mechanism in this sub-function focuses on non-critical data monitoring. The difference between critical and non-critical data could be defined in the data quality monitoring templates stored in the data quality monitoring repository.

Appendix I

Intelligent data quality management using machine learning and deep learning

(This appendix does not form an integral part of this Recommendation.)

I.1 Machine learning based data quality management

The identification and correction of measurement errors often involves labour intensive case-by-case evaluations by statisticians. Machine learning will increase the efficiency and effectiveness of these evaluations. It proceeds in two steps: in the first step, a supervised learning algorithm exploits data on decisions to flag data points as erroneous to approximate the results of the human decision making process. In the second step, the algorithm applies the first-step knowledge to predict the probability of measurement errors for newly reported data points.

Further improving and maintaining high data quality is a central goal of official statistics. In the field of data quality management (DQM), the collection of data on human decisions in the DQM process creates an opportunity to increase the efficiency and effectiveness of DQM with machine learning (ML). It is necessary to predict measurement errors on the basis of data on human decisions to flag data points as erroneous. These predicted probabilities of measurement errors facilitate the work of statisticians and form the basis for a ML based approach to automate their checks.

The main focus of an application of ML to DQM is in the following two applications:

– prediction of measurement errors;

– helping to overcome data gaps.

To support DQM in both applications, the ML algorithms predict if a human decision maker would flag data points. In the application to data gaps, the algorithms predict missing values.

The ML yields accurate out-of-sample predictions and increases the efficiency of DQM. The potential of ML for official statistics is not limited to the prediction of measurement errors. Another important problem that ML can help to overcome is missing data. Out-of-sample predictions of missing values with ML algorithms can help to close data gaps in a wide range of datasets.

The use of ML based matching algorithms enables the service platform to ingest data for standardization at scale. Typical use cases include matching specific records or data sets to a common standard and transforming data to this standard, allowing for the creation of relationships and accurate links between base data and derived data. This workflow is particularly important in ML and fraud detection scenarios, where a high volume of customer's due diligence and transaction data from disparate systems necessitates extensive standardization to set flags and generate meaningful derived data.

Standardization simplifies deduplication issues and accuracy-related data quality problems. The flexibility of ML provides that it can be applied across the entire data set in a cost-effective way, reducing the overhead of moving to a new standard.

I.2 Deep learning based DQM

Deep learning (DL) in artificial intelligence (AI) might help us discover where the master data is kept. DL might be able to "spot" where the most frequently referenced data reside.

In fact, there are two other tasks in DL applied to DQM that are very different and where we do not need, and cannot use, DL.

The first task is the identification of the master data and the second concerns the enforcement of the policies that sustain it. The former steps should take no more than an hour with the right business people in the room; simply ask the business users such things as:

– What is the most important data (at a conceptual level) that is needed to make business process A work as planned?

– How much of this data is also needed to make business process B work as planned?

– How much less data can you use to make business process C work as planned?

The second task is at the other extreme; the enforcement of policy. It is the work of policy enforcement that sustains the level of data quality and the effectiveness of the workflows executed to meet that data quality and business process.

Bibliography

[b-ITU-T Y.3600] Recommendation ITU-T Y.3600 (2015), *Big data – Cloud computing based requirements and capabilities*.

[b-ITU-T Technical Paper] Technical Paper (2016), *Analysis of Digital Data Technologies Toward Future Data Eco-Society*.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	Tariff and accounting principles and international telecommunication/ICT economic and policy issues
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling, and associated measurements and tests
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	**Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities**
Series Z	Languages and general software aspects for telecommunication systems

Published in Switzerland
Geneva, 2023