Technical Specification

ISO/TS 8000-230

First edition
2026-05

Data quality —

Part 230: Sensor data — Guidelines for data cleansing

Qualité des données —

Partie 230: Données des capteurs — Lignes directrices relatives au nettoyage des données

Reference number
ISO/TS 8000-230:2026(en)

COPYRIGHT PROTECTED DOCUMENT

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

ISO draws attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO takes no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents. ISO shall not be held responsible for identifying any or all such patent rights.

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 184, *Automation systems and integration*, Subcommittee SC 4, *Industrial data*.

A list of all parts in the ISO 8000 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found atwww.iso.org/members.html.

Introduction

0.1 Foundations of the ISO 8000 series

Digital data deliver value by enhancing all aspects of organizational performance including:

— operational effectiveness and efficiency;

— safety and security;

— reputation with customers and the wider public;

— compliance with statutory regulations;

— innovation;

— consumer costs, revenues and stock prices.

In addition, many organizations are now addressing these considerations with reference to the United Nations Sustainable Development Goals[1)].

The influence on performance originates from data being the formalized representation of information. ISO 8000-2 [1] defines information as "knowledge concerning objects, such as facts, events, things, processes, or ideas, including concepts, that within a certain context has a particular meaning". This information enables organizations to make reliable decisions. This decision making can be performed by human beings directly and also by automated data processing including artificial intelligence systems.

Through widespread adoption of digital computing and associated communication technologies, organizations become dependent on digital data. This dependency amplifies the negative consequences of lack of quality in these data. These consequences are the decrease of organizational performance.

The biggest impact of digital data comes from two key factors:

— the data having a structure that reflects the nature of the subject matter;

EXAMPLE 1 A research scientist writes a report using a software application for word processing. This report includes a table that uses a clear, logical layout to show results from an experiment. These results indicate how material properties vary with temperature. The report is read by a designer, who uses the results to create a product that works in a range of different operating temperatures.

— the data being computer processable (machine readable) rather than just being for a person to read and understand.

EXAMPLE 2 A research scientist uses a database system to store the results of experiments on a material. This system controls the format of different values in the data set. The system generates an output file of digital data. This file is processed by a software application for engineering analysis. The application determines the optimum geometry when using the material to make a product.

ISO 9000 [2] explains that quality is not an abstract concept of absolute perfection. Quality is actually the conformance of characteristics to requirements. This actuality means that any item of data can be of high quality for one purpose but not for a different purpose. The quality is different because the requirements are different between the two purposes.

EXAMPLE 3 Time data are processed by calendar applications and also by control systems for propulsion units on spacecraft. These data include start times for meetings in a calendar application and activation times in a control system. These start times require less precision than the activation times.

1) https://sdgs.un.org/goals

The nature of digital data is fundamental to establishing requirements that are relevant to the specific decisions made by each organization.

EXAMPLE 4 ISO 8000-1 [3] identifies that data have syntactic (format), semantic (meaning) and pragmatic (usefulness) characteristics.

To support the delivery of high-quality data, the ISO 8000 series addresses:

— data governance, data quality management and maturity assessment;

EXAMPLE 5 ISO 8000-61 [4] specifies a process reference model for data quality management.

— creating and applying requirements for data and information;

EXAMPLE 6 ISO 8000-110 [5] specifies how to exchange characteristic data that are master data.

— monitoring and measuring information and data quality;

EXAMPLE 7 ISO 8000-8 [6] specifies approaches to measuring information and data quality.

— improving data and, consequently, information quality;

EXAMPLE 8 ISO/TS 8000-81 [7] specifies an approach to data profiling, which identifies opportunities to improve data quality.

— issues that are specific to the type of content in a data set.

EXAMPLE 9 ISO/TS 8000-311 [8] specifies how to address quality considerations for product shape data.

Data quality management covers all aspects of data processing, including creating, collecting, storing, maintaining, transferring, exploiting and presenting data to deliver information.

Effective data quality management is systemic and systematic, requiring an understanding of the root causes of data quality issues. This understanding is the basis for not just correcting existing nonconformities but for also implementing solutions that prevent future reoccurrence of those nonconformities.

EXAMPLE 10 If a data set includes dates in multiple formats including “yyyy-mm-dd”, “mm-dd-yy” and “dd-mm-yy”, then data cleansing can correct the consistency of the values. Such cleansing requires additional information, however, to resolve ambiguous entries (such as “04-05-20”). The cleansing also cannot address any process issues and people issues, including training, that have caused the inconsistency.

0.2 Understanding more about the ISO 8000 series

ISO 8000-1 [3] provides a detailed explanation of the structure and scope of the ISO 8000 series.

ISO 8000-2 [1] specifies the single, common vocabulary for the ISO 8000 series. This vocabulary is ideal reading material by which to understand the overall subject matter of data quality. ISO 8000-2 [1] presents the vocabulary structured by a series of topic areas (e.g. terms relating to quality and terms relating to data and information).

ISO has identified ISO 8000-1 [3], ISO 8000-2 [1] and ISO 8000-8 [6] as horizontal deliverables, i.e. deliverable dealing with a subject relevant to a number of committees or sectors or of crucial importance to ensure coherence across standardization deliverables.

0.3 Role of this document

As a contribution to the overall capability of the ISO 8000 series, this document addresses guidelines to improve the quality of sensor data by cleansing data anomalies that affect low quality characteristics. The guidelines include principles, the process and implementation requirements for sensor data cleansing. The process performs sensor data cleansing using data quality characteristics and anomalies defined in ISO 8000-210 [9] and data quality measures defined in ISO 8000-220 [10]. To help users understand, they also present methods and examples of cleansing data anomalies. Through this document, users will learn

procedures and methods for improving the quality of sensor data collected from IoT or sensor network environments prior to data analysis and exploitation.

This document supports activities that affect:

— one or more information systems;

— data flows within the organization and with external organizations;

— any phase of the data life cycle.

Organizations can use this document on its own or in conjunction with other parts in the ISO 8000 series. Annex A contains an identifier that conforms to ISO/IEC 8824-1 [11]. The identifier unambiguously identifies this document in an open information system.

0.4 Benefits of the ISO 8000 series

By implementing parts of the ISO 8000 series to improve organizational performance, an organization achieves the following benefits:

— objective validation of the foundations for digital transformation of the organization;

— a sustainable basis for data in digital form becoming a fundamental asset class the organization relies on to deliver value;

— securing evidence-based trust from other parties (including supply chain partners and regulators) about the repeatability and reliability of data and information processing in the organization;

— portability of data with resulting protection against loss of intellectual property and re-usability across the organization and applications;

— effective and efficient interoperability between all parties in a supply chain to achieve traceability of data back to original sources;

— readiness to acquire or supply services where the other party expects to work with common understanding of explicit data requirements.

Data quality —

Part 230:
Sensor data — Guidelines for data cleansing

1 Scope

This document specifies guidelines to improve data quality by cleansing sensor data anomalies that affect low inherent quality characteristics.

The following are within the scope of this document:

— principles for sensor data cleansing;

— the process for sensor data cleansing;

— implementation requirements for sensor data cleansing;

— list of data anomaly detection and repair methods (see Annex B);

— examples of sensor data cleansing (see Annex C).

The following are outside the scope of this document:

— algorithms or detailed methods to detect and repair data anomalies;

— the process of sensor data cleansing for real time processing.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 8000-2, *Data quality — Part 2: Vocabulary*

ISO 8000-210, *Data quality – Part 210: Sensor data: Data quality characteristics*

ISO 8000-220, *Data quality – Part 220: Sensor data: Quality measurement*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 8000-2 and the following apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1 Terms relating to sensor data

3.1.1
sensor
device that observes and measures a property of a natural phenomenon, system or human-made process and converts that measurement into a signal

Note 1 to entry: A sensor can exist not only in a single physical form but also in a sensor-based variant such as a virtual sensor.

[SOURCE: ISO/IEC 29182-2:2013 [12], 2.1.5, modified — "system" has been added to the definition, "man" changed to "human", and "physical" deleted from the definition. Note 1 to entry has been changed.]

3.1.2
sensor network
system of spatially distributed *sensor* (3.1.1) nodes interacting with each other and, depending on applications, possibly with other infrastructure in order to acquire, process, transfer, and provide information extracted from its environment with a primary function of information gathering and possible control capability

Note 1 to entry: Distinguishing features of a sensor network can include wide area coverage, use of radio networks, flexibility of purpose, self-organization, openness, and providing data for multiple applications.

[SOURCE: ISO/IEC 29182-2:2013 [12], 2.1.6]

3.1.3
sensor node
sensor network (3.1.2) element that includes at least one *sensor* (3.1.1) and, optionally actuators with communication capabilities and data processing capabilities

Note 1 to entry: It can include additional application capabilities.

Note 2 to entry: A hybrid sensor composed of multiple sensors is considered a sensor node that includes multiple sensors.

[SOURCE: ISO/IEC 29182-2:2013 [12], 2.1.8, modified — Note 2 to entry has been added to the definition.]

3.1.4
sensor data
data produced by a *sensor node* (3.1.3)

Note 1 to entry: Sensor data consist of a stream of digital values converted from *sensor* (3.1.1) signals, and information such as the identification of each sensor and timestamps of data acquired by the sensor node.

3.1.5
internet of things
IoT
infrastructure of interconnected entities, people, systems and information resources together with services which processes and reacts to information from the physical world and virtual world

[SOURCE: ISO/IEC 20924:2024 [13], 3.2.4]

3.2 Terms relating to data quality

3.2.1
data anomaly
item of data in a data set, where the item deviates from the expected pattern for items in the data set

3.2.2
quality characteristic
inherent characteristic of an object related to a requirement

Note 1 to entry: ISO 8000-8 [6] uses the term quality dimension as a synonym for quality characteristics that determine the pragmatic quality of data.

[SOURCE: ISO 9000:2015 [14], 3.10.2, modified — Note 1 to entry has been added.]

3.2.3
data cleansing
process used to improve data quality by detecting and repairing defects and errors in data

Note 1 to entry: In ISO 8000-2 [1], data error is defined as non-fulfilment of a data requirement and also noted as synonymous with data nonconformity.

Note 2 to entry: In ISO 9000 [2], defect is defined as non-fulfilment of a requirement related to an intended or specified use.

Note 3 to entry: In ISO 8000-61 [4], data cleansing is specified as a sub-process of data quality improvement.

[SOURCE: ISO 13008:2022 [15], 3.4, modified — “correcting (or removing)" is changed to “repairing” and Notes 1, 2 and 3 to entry are added.]

3.2.4
data profiling
activities that are performed to understand the data structures and system rules that affect the extraction of audit data

[SOURCE: ISO 21378:2019 [16], 3.6]

3.3 Terms relating to measurement

3.3.1
data quality measure
quality measure
variable to which a value is assigned as the result of measuring a data *quality characteristic* (3.2.2)

Note 1 to entry: Adapted from ISO/IEC 25012:2008 [17], 4.5.

4 Principles for sensor data cleansing

— When a data anomaly occurs due to sensor or system errors, the quality of the data shall be improved by deleting or modifying the anomalous data.

— When a data anomaly reflects actual phenomena in the field, whether to maintain, delete, or modify the anomalous data shall be decided according to the stated purpose of an intended or specified use.

— When the cause of data anomaly is not clearly identified, data deletion or modification shall be minimized to avoid changing the original correct data.

— When a data anomaly cannot be deleted or modified for any reason, a flag or mark shall be placed on the data so that the person in charge of the data can recognize it and take appropriate actions.

— Data cleansing shall be carried out with the consent of stakeholders.

5 Process for sensor data cleansing

5.1 General

The sensor data cleansing process is designed with the following considerations in mind:

— The plan-do-check-act concept used to define the data quality management process in ISO 8000-61 [4] is applied to the data cleansing process. In other words, the process is designed with the following activities: provide a quality measurement plan (plan), measure data quality (do and check) and improve data quality (act). In addition, once the plan is provided, activities of measurement (do and check) and improvement (act) are repeatedly performed to determine whether the sensor data satisfy quality requirements.

— This process is designed for post processing (or offline mode), not for real time processing (or online mode).

NOTE 1 As sensor data are collected in the form of streams in real time and the amount is very large, it takes time to cleanse them. Therefore, real-time data cleansing is not realistic in the environment for rapid decision-making. Real-time data cleansing can only be performed in special environments where data anomalies are already known and do not need checked or verified.

— The process is represented by the IDEF0 (integration definition for function modelling) functional model defined by ISO/IEC/IEEE 31320-1 [18]. This model breaks down a process into hierarchical activities to show what activities are performed and how. It helps analyse and design processes by clearly showing the inputs, outputs, controls, and mechanisms of each activity.

NOTE 2 A functional model is identified by a model name, an IDEF0 box is identified by a box name, and an IDEF0 arrow segment is identified by an arrow label. An identifier is written in title case, i.e. the first letter of each word is capitalized. See ISO/IEC/IEEE 31320-1 [18] for details on the notation in the functional model.

5.2 Functional model of sensor data cleansing

5.2.1 Perform sensor data cleansing (A0)

The functional model of the sensor data cleansing process is represented by the A-0 context diagram for perform sensor data cleansing (see Figure 1).

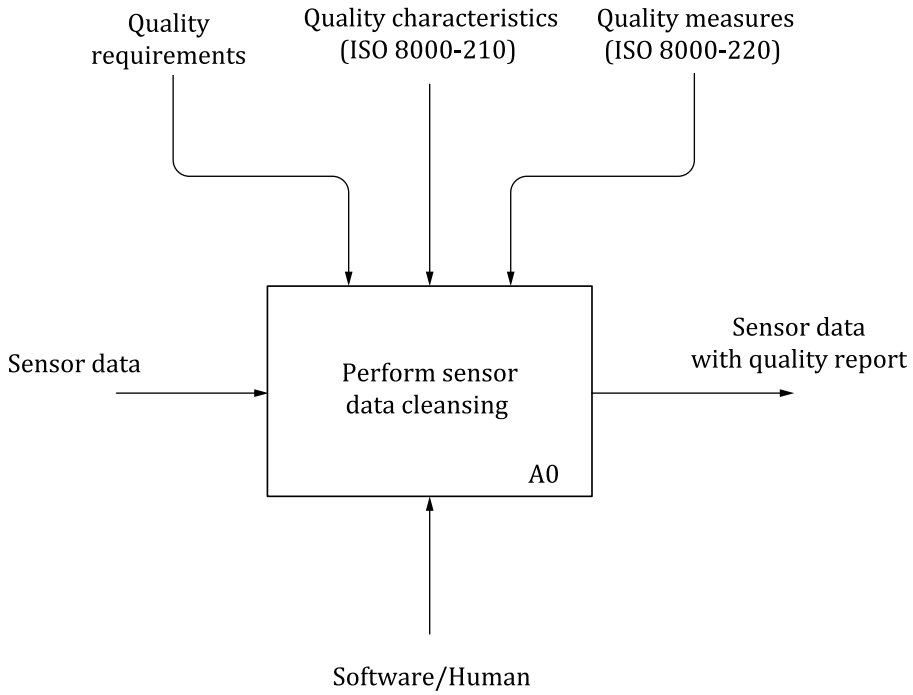

Figure 1 — A-0 context diagram for perform sensor data cleansing (model diagram A0)

This process is to perform data cleansing to improve the quality of sensor data prior to data analysis or exploitation. By accepting sensor data and considering quality requirements, quality characteristics defined in ISO 8000-210 [19], and quality measures defined in ISO 8000-220 [20], the process provides sensor data with a quality report as an output.

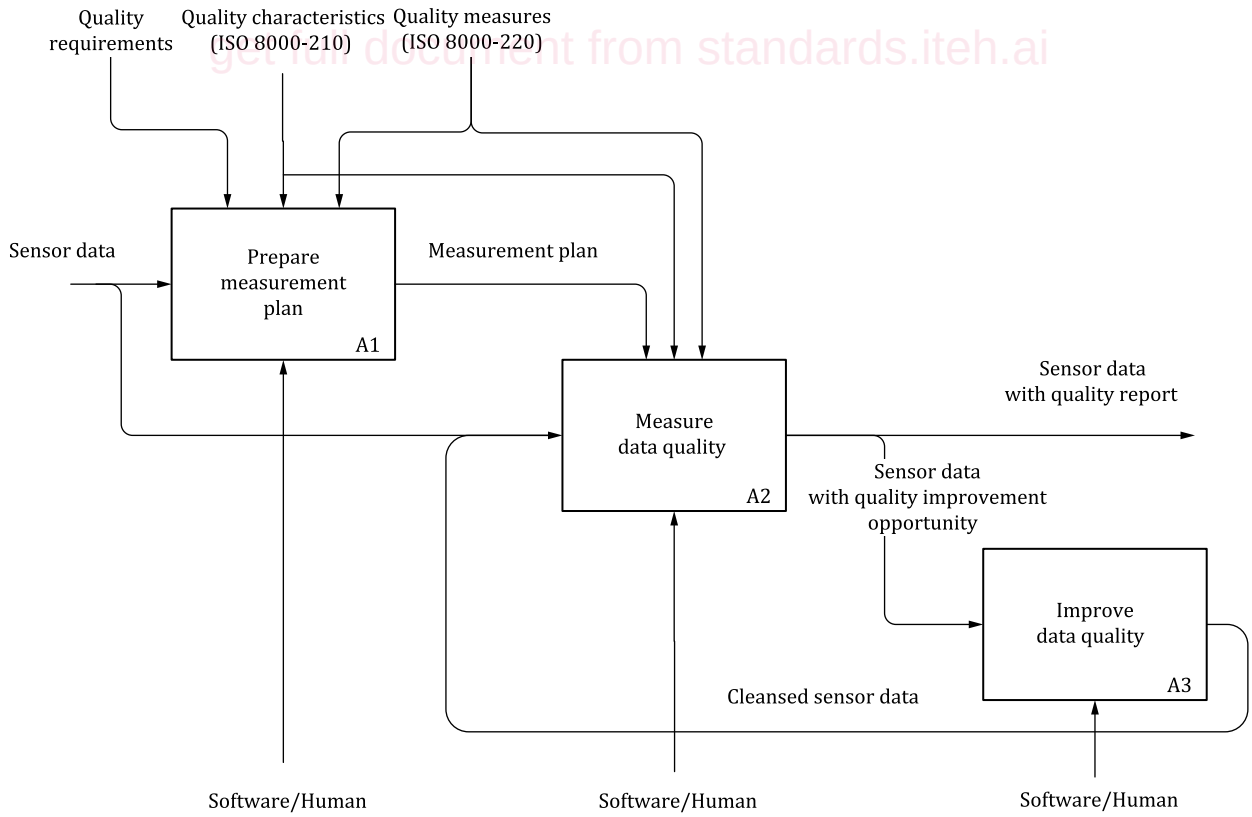

Figure 2 — Perform sensor data cleansing (model diagram A0)

As in Figure 2, this process consists of three activities, prepare measurement plans (A1), measure data quality (A2) and improve data quality (A3).

NOTE 1 Figure 2 is a child diagram of Figure 1.

Each activity at the lowest level of the process is described by the following elements:

— a title, which is a descriptive heading for an activity (modified from ISO/IEC TR 24774:2010 [21]);

— a purpose, which describes the goal of performing an activity (modified from ISO/IEC TR 24774:2010 [21]);

— tasks, which are required, recommended, or permissible actions, intended to contribute to the achievement of the goal of an activity (modified from ISO/IEC/IEEE 24774:2021 [22]);

— inputs, which are items transformed into output by an activity (modified from ISO/IEC/IEEE 31320-1 [18]);

— outputs, which are product, result or service produced by an activity (modified from ISO/IEC/IEEE 24774:2021 [22]);

— controls, which are conditions or constraints required for an activity to produce correct output (modified from ISO/IEC/IEEE 31320-1 [18]);

— mechanisms, which are the means used by an activity to transform input into output (modified from ISO/IEC/IEEE 31320-1 [18]).

NOTE 2 These elements are adapted from those of process description in ISO/IEC TR 24774:2010 [21], ISO/IEC/IEEE 24774:2021 [22] and those of functional model in ISO/IEC/IEEE 31320-1 [18] to fit the activity definition.

5.2.2 Prepare measurement plan (A1)

5.2.2.1 General

This activity is intended to prepare a plan for measuring sensor data quality based on quality requirements, quality characteristics, quality measures and sensor data.

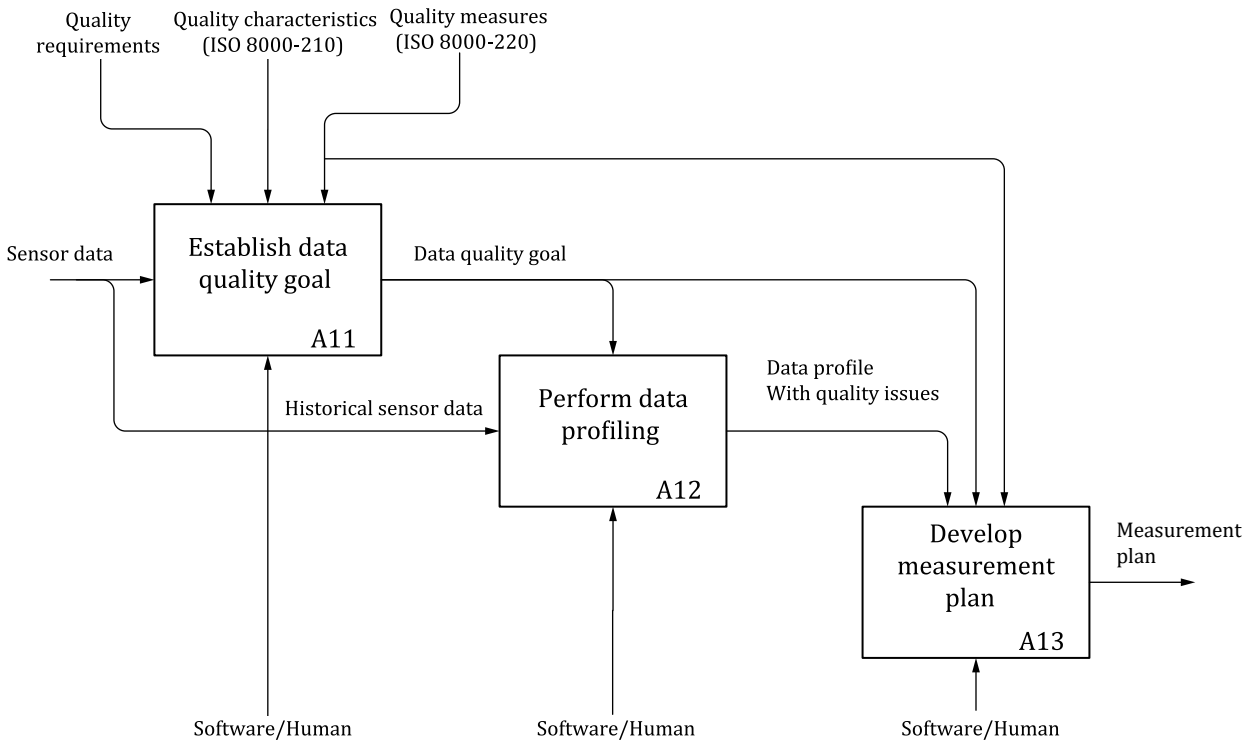

Figure 3 — Prepare measurement plan (model diagram A1)

As in Figure 3, this activity consists of three sub-activities, establish data quality goal (A11), perform data profiling (A12) and develop measurement plan (A13).

5.2.2.2 Establish data quality goal (A11)

Purpose: Establish data quality goal is to determine the data quality-related goals that reflect quality requirements of sensor data.

Task:

— gather data quality requirements from stakeholders;

— determine the goal to achieve based on data quality requirements.

Input: Sensor data collected from sensor nodes.

Output: Data quality goal represented by data quality requirements such as quality measure levels of quality characteristics in interest.

Control: Quality requirements, quality characteristics and corresponding data anomalies defined in ISO 8000-210, and quality measures defined in ISO 8000-220.

Mechanism: Software/Human

5.2.2.3 Perform data profiling (A12)

Purpose: Perform data profiling to acquire historical sensor data and perform their data profiling. Through this activity, the profile and data quality issues of sensor data are extracted from a cluster of historical occurrences of the relevant sensor data.

Task:

— collect historical sensor data;