International Standard

ISO/IEC 5259-3

First edition
2024-07

Artificial intelligence — Data quality for analytics and machine learning (ML) —

Part 3: Data quality management requirements and guidelines

Intelligence artificielle — Qualité des données pour les analyses de données et l'apprentissage automatique —

Partie 3: Exigences et lignes directrices pour la gestion de la qualité des données

iTeh Standards
(https://standards.iteh.ai)
Document Preview
ISO/IEC 5259-3:2024
https://standards.iteh.ai/catalog/standards/iso/cda43a2b-10d5-4b3d-b2c2-c4593045a3b4/iso-iec-5259-3-2024

Reference number
ISO/IEC 5259-3:2024(en)

© ISO/IEC 2024

iTeh Standards
(https://standards.iteh.ai)
Document Preview

ISO/IEC 5259-3:2024
https://standards.iteh.ai/catalog/standards/iso/cda43a2b-10d5-4b3d-b2c2-c4593045a3b4/iso-iec-5259-3-2024

COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2024

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

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 ..... **v**
Introduction ..... **vi**
1 Scope **1**
2 Normative references **1**
3 Terms and definitions **1**
4 Symbols and abbreviated terms **2**
5 Intended usage **2**
6 Overall data quality management **2**
6.1 Objective 2
6.2 General 2
6.3 Requirements and recommendations 3
6.3.1 General 3
6.3.2 Data quality culture 3
6.3.3 Management of data quality issues 3
6.3.4 Competence management 3
6.3.5 Resource management 4
6.3.6 Management system integration 4
6.3.7 Documentation 4
6.3.8 Data quality audit and assessment 4
6.3.9 Confirmation review and data quality measures 5
6.3.10 Project-specific data quality management 5
6.4 Work products 5
7 Life cycle-specific data quality management **6**
7.1 Objective 6
7.2 General 6
7.2.1 Data quality management life cycle 6
7.2.2 Data quality management life cycle stages 7
7.2.3 Project-independent tailoring of the data quality management life cycle 8
7.2.4 Horizontal aspects of the data quality management life cycle 8
7.3 Requirements and recommendations 9
7.3.1 Data motivation and conceptualization 9
7.3.2 Data specification 9
7.3.3 Data planning 11
7.3.4 Data acquisition 11
7.3.5 Data preprocessing 13
7.3.6 Data augmentation 13
7.3.7 Data provisioning 14
7.3.8 Data decommissioning 16
7.4 Work products 17
7.4.1 Work products of data motivation and conceptualization stage 17
7.4.2 Work products of data specification stage 17
7.4.3 Work products of data planning stage 17
7.4.4 Work products of data acquisition stage 17
7.4.5 Work products of data preprocessing stage 17
7.4.6 Work products of data augmentation stage 18
7.4.7 Work products of data provisioning stage 18
7.4.8 Work products of data decommissioning stage 18
8 Horizontal processes **18**
8.1 Objective 18
8.2 General 18
8.3 Requirements and recommendations 18
8.3.1 Verification and validation 18

iTeh Standards
(https://standards.iteh.ai)
Document Preview
ISO/IEC 5259-3:2024
https://standards.iteh.ai/catalog/standards/iso/cdc4ca2b-1d45-4b3d-b4e2-a4593045a3b4/iso-iec-5259-3-2024

© ISO/IEC 2024 – All rights reserved

8.3.2 Configuration management 19
8.3.3 Change management 19
8.3.4 Risk management 20
8.4 Work products 21
8.4.1 Work products of verification and validation 21
8.4.2 Work products of configuration management 21
8.4.3 Work products of change management 21
8.4.4 Work products for risk management 21

9 Management of data quality in supply chains 22
9.1 Objective 22
9.2 Requirements and recommendations 22
9.3 Work products 22

10 Management of data processing tools 23
10.1 Objective 23
10.2 Requirements and recommendations 23
10.3 Work products 23

11 Management of data quality dependencies 23
11.1 Objective 23
11.2 Requirements and recommendations 23
11.3 Work products 23

12 Project-specific data quality management 24
12.1 Objective 24
12.2 Requirements and recommendations 24
12.2.1 Context and intended use 24
12.2.2 Objective 24
12.2.3 Requirements and recommendations 24
12.3 Specification and management of data quality requirements 24
12.3.1 Objective 24
12.3.2 Requirements and recommendations 25
12.4 Roles and responsibilities in data quality management 25
12.4.1 Objective 25
12.4.2 Requirements and recommendations 25
12.4.3 Work products 25
12.5 Tailoring of the data quality activities 25
12.6 Planning and coordination of the data quality activities 26
12.6.1 General 26
12.6.2 Data quality plan 26
12.6.3 Planning of processes 26
12.7 Progression of the data quality life cycle 26
12.8 Data quality justification 26
12.9 Decommissioning 27
12.10 Work products 27

Bibliography 28

iTeh Standards
(https://standards.iteh.ai)
Document Preview
ISO/IEC 5259-3:2024
https://standards.iteh.ai/catalog/standards/sist/...

© ISO/IEC 2024 – All rights reserved

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

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 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 or www.iec.ch/members_experts/refdocs).

ISO and IEC draw attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO and IEC take 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 and IEC 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 and https://patents.iec.ch. ISO and IEC 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. In the IEC, see www.iec.ch/understanding-standards.

iTeh Standards
(https://standards.iteh.ai)
Document Preview

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 42, *Artificial intelligence*.

A list of all parts in the ISO/IEC 5259 series can be found on the ISO and IEC websites.

ISO/IEC 5259-3:2024

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 at www.iso.org/members.html and www.iec.ch/national-committees.

© ISO/IEC 2024 – All rights reserved

Introduction

The quality of analytics and machine learning (ML) based products and services depends on the quality of data used to train ML models. Hence, data quality management is essential as it often helps to ensure the success of analytics and ML technology.

The adoption of a data quality management system facilitates managing the quality of products and services that employ analytics and ML technologies. This document defines vocabulary, requirements and guidelines for communication, alignment and agreement for managing data quality. The data quality management system provides transparency and auditability, either through self-assessment or third party assessment. It facilitates achieving relevant stakeholder satisfaction and managing quality, performance and self-declaration requirements. Specifically, this document defines requirements for a data quality management system with references to data quality measures that are relevant for the most commonly used analytics and ML technologies.

As data quality requirements vary with context and application domain, this document provides a generic set of requirements and recommendations relating to common data life cycle stages. A data life cycle is typically tightly integrated with the accompanying AI system life cycle and therefore has several dependencies. This document does not prescribe what AI system life cycle to use. Instead, it provides generic interfaces that allow users of this document the flexibility to interface with several life cycle models as long as the life cycle processes can be mapped.

ISO/IEC 5259-1 describes the data quality terminology and concepts used in this document.

ISO/IEC 5259-2[1)] describes the data quality model and data quality measures used in this document.

ISO/IEC 5259-4 describes the data quality process framework used in this document.

ISO/IEC 5259-5[2)] provides a data quality governance framework as guidance for governing bodies.

ISO/IEC TR 5259-6[3)] describes a visualization framework for data quality in analytics and ML.

iTeh Standards
(https://standards.iteh.ai)
Document Preview

ISO/IEC 5259-3:2024
https://standards.iteh.ai/catalog/standards/iso/cda43a2b-10d5-4b3d-b2c2-c4593045a3b4/iso-iec-5259-3-2024

1) Under preparation. Stage at the time of publication: ISO/IEC FDIS 5259-2:2024.

2) Under preparation. Stage at the time of publication: ISO/IEC DIS 5259-5:2023.

3) Under preparation. Stage at the time of publication: ISO/IEC CD TR 5259-6:2023.

© ISO/IEC 2024 – All rights reserved

Artificial intelligence — Data quality for analytics and machine learning (ML) —

Part 3:
Data quality management requirements and guidelines

1 Scope

This document specifies requirements and provides guidance for establishing, implementing, maintaining and continually improving the quality of data used in the areas of analytics and machine learning.

This document does not define a detailed process, methods or metrics. Rather it defines the requirements and guidance for a quality management process along with a reference process and methods that can be tailored to meet the requirements in this document.

The requirements and recommendations set out in this document are generic and are intended to be applicable to all organizations, regardless of type, size or nature.

2 Normative references

iTeh Standards
(https://standards.iteh.ai)
Document Preview

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/IEC 5259-1:2024, *Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 1: Overview, terminology, and examples*

ISO/IEC 5259-3:2024

ISO/IEC 5259-2[4)], *Artificial Intelligence — Data quality for analytics and machine learning (ML) — Part 2: Data quality measures*

ISO/IEC 22989, *Information technology — Artificial intelligence — Artificial intelligence concepts and terminology*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 22989, ISO/IEC 5259-1 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
data quality claim
statement to what degree data satisfy a data quality requirement

4) Under preparation. Stage at the time of publication: ISO/IEC FDIS 5259-2:2024.

© ISO/IEC 2024 – All rights reserved

3.2
data quality plan
specification of practices, processes and allocation of resources to achieve data quality objectives as the outcome of data quality planning

3.3
data quality planning
part of data quality management focused on setting data quality objectives and specifying necessary operational processes and related resources to achieve the quality objectives

[SOURCE: ISO 8000-2:2022, modified — example removed]

3.4
development interface agreement
DIA
agreement between customer and supplier in which the responsibilities for activities to be performed, evidence to be reviewed, or work products to be exchanged by each party related to the development of items or elements are specified

Note 1 to entry: While DIA applies to the development phase, supply agreement applies to production.

[SOURCE: ISO 26262-1:2018]

4 Symbols and abbreviated terms

DQMLC data quality management life cycle

iTeh Standards
(https://standards.iteh.ai)
Document Preview

5 Intended usage

This document may be used in one or more of the following modes:

— by an organization to establish and tailor a data quality management process for the use of data in analytics and ML, and continually improve processes;

— by an ML project to define, trace and evaluate data quality requirements;

ISO/IEC 5259-3:2024
https://standards.iteh.ai/catalog/standards/sist/[illegible]3045a3b4/iso-iec-5259-3-2024

— by a data user and data holder to establish a common understanding of data quality characteristics, and to ensure that agreed requirements have been met, facilitating an agreement for transacting data.

NOTE The organization can request assurances of confidentiality and proper use for supporting evidence.

6 Overall data quality management

6.1 Objective

The objective of a data quality management process is to establish appropriate (i.e. repeatable and auditable) processes to manage the quality of data and reliably meet a given set of requirements set by the organization.

6.2 General

Data quality impacts outcomes of analytics and ML algorithms. Data quality has an inherent constituent and a system-dependent constituent. Data can be suitable for one application but not suitable for another. This document helps to establish and maintain data quality for each analytics and ML application.

© ISO/IEC 2024 – All rights reserved

6.3 Requirements and recommendations

6.3.1 General

The following requirements and recommendations apply to the whole organization.

6.3.2 Data quality culture

The organization should sustain a data quality culture.

The organization shall:

a) have rules and processes to achieve quality (according to this document) taking into account the data quality model as applied to the applicable products and services;

b) define and implement data quality management processes, and perform related data quality activities;

c) integrate the data quality management processes and activities, to the extent appropriate, into other management processes and activities, such as general quality management and risk management;

d) document the performed activities;

e) provide resources sufficient to perform data quality management;

f) monitor, and to the extent necessary, review and improve the data quality management processes;

g) provide the required authority to involved personnel;

h) communicate data quality policies within the organization.

iTeh Standards
(https://standards.iteh.ai)
Document Preview

6.3.3 Management of data quality issues

The organization shall meet data quality requirements by:

a) having processes for communicating, analysing, evaluating, resolving and closing data quality issues;

b) documenting closed issues;

ISO/IEC 5259-3:2024
https://standards.iteh.ai/catalog/standards/iso/cda43a2b-10d5-4b3d-b2c2-c4593045a3b4/iso-iec-5259-3-2024

c) escalating or delegating issues that cannot be closed.

NOTE 1 Resolving and closing issues of data quality can include limiting or adjusting the scope of the ML project.

NOTE 2 A data quality issue can be closed by implementing a resolution or determining a resolution based on defined acceptance criteria.

6.3.4 Competence management

The organization shall manage competence by:

a) documenting required skills and tools to process the data;

b) ensuring that involved personnel have sufficient skills to perform their activities and duties;

c) maintaining records of persons and their proficiency on the required skills and tools;

d) keeping appropriate records of training and experience that substantiate claims of appropriate skills.

The organization can use external sources of competencies.

© ISO/IEC 2024 – All rights reserved

6.3.5 Resource management

The organization shall provide the resources required for data quality management, including:

a) software applications, training and support necessary to perform data quality management;

b) IT infrastructure or services necessary to perform data quality management (e.g. compute, storage, networking);

c) personnel with the skills required to perform data quality management.

6.3.6 Management system integration

The organization should integrate its data quality management activities into its existing management system, including its management systems for product or service quality, and for the development and use of AI systems. Implications from dual roles of stakeholders should be managed by the quality management system, including mitigation of any conflicts of interest.

NOTE 1 Stakeholder management can consider the potential of multiple roles for an individual. A user of analytics and ML based products or services can also be an owner or contributor of data.

NOTE 2 Organizations can use ISO/IEC 42001 to define a management system for the development or use of AI systems.

NOTE 3 Organizations can use ISO 9001 or other sector-specific quality management systems to define their quality management system.

6.3.7 Documentation

iTeh Standards
(https://standards.iteh.ai)
Document Preview

Documentation shall be intelligible to relevant stakeholders of the project in accordance with their role. Resources in a language that is not understood by a relevant stakeholder should be accompanied by a summary in a language that the stakeholder can understand.

The documentation shall be accessible to relevant stakeholders as appropriate and authorized. Access overhead should be minimized.

ISO/IEC 5259-3:2024
https://standards.iteh.ai/catalog/standards/iso/cda43e...-...-...-

The data quality assessment plan shall be included in the specification stage. The data quality assessment shall be performed before data provisioning (see Figure 1, Stage 7: Data provisioning) or at an appropriate interval when using continuous learning or when using streaming data.

The data quality assessment may be performed on a subset of the data when it can be demonstrated that the quality of the subset is representative of the quality of the complete dataset.

6.3.9 Confirmation review and data quality measures

Data quality shall be confirmed by appropriate data quality measures in accordance with ISO/IEC 5259-2. A data quality review shall at least cover:

a) confirmation reviews of key work products. Every confirmation review:

   1) shall be finalized before data provisioning;

   2) should be based on whether the objectives of this document are achieved;

b) quality audits of the implemented processes;

c) quality assessment of the data.

All work products shall undergo confirmation reviews.

The personnel performing these reviews shall have access to the involved personnel, relevant information and required resources.

NOTE Confirmation reviews of key work products can be delegated, but the responsibility stays with the designated person.

iTeh Standards
(https://standards.iteh.ai)
Document Preview
ISO/IEC 5259-3:2024

6.3.10 Project-specific data quality management

The organization shall manage project-specific data by:

a) establishing a suitable project-specific data quality management process that meets all requirements of the specific ML project;

b) maintaining a list of relevant data quality claims. Where applicable, quantitative and qualitative benchmarks for data quality measures shall be documented;

c) adopting appropriate processes to identify and manage all data quality measures relevant for the project.

The project-specific data quality management process shall fulfil the requirements of Clause 12.

6.4 Work products

Work products of the data quality management process shall include:

a) organization-specific rules and processes for data quality (e.g. according to ISO/IEC 5259-4);

b) evidence of competence management;

c) evidence of a data quality management system;

d) identification of the used data quality measures;

e) documentation of applicable data quality measure benchmarks;

f) identified quality anomaly reports.

© ISO/IEC 2024 – All rights reserved