



FINAL DRAFT International Standard

ISO/IEC FDIS 5259-5

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

Part 5: Data quality governance framework

ISO/IEC JTC 1/SC 42

Secretariat: **ANSI**

Voting begins on:
2024-10-29

Voting terminates on:
2024-12-24

iTeh Standards
(<https://standards.itih.ai>)
Document Preview

[ISO/IEC FDIS 5259-5](https://standards.itih.ai/catalog/standards/iso/745e0014-e40f-4160-b41a-4e2712730582/iso-iec-fdis-5259-5)

<https://standards.itih.ai/catalog/standards/iso/745e0014-e40f-4160-b41a-4e2712730582/iso-iec-fdis-5259-5>

RECIPIENTS OF THIS DRAFT ARE INVITED TO SUBMIT, WITH THEIR COMMENTS, NOTIFICATION OF ANY RELEVANT PATENT RIGHTS OF WHICH THEY ARE AWARE AND TO PROVIDE SUPPORTING DOCUMENTATION.

IN ADDITION TO THEIR EVALUATION AS BEING ACCEPTABLE FOR INDUSTRIAL, TECHNOLOGICAL, COMMERCIAL AND USER PURPOSES, DRAFT INTERNATIONAL STANDARDS MAY ON OCCASION HAVE TO BE CONSIDERED IN THE LIGHT OF THEIR POTENTIAL TO BECOME STANDARDS TO WHICH REFERENCE MAY BE MADE IN NATIONAL REGULATIONS.

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

[ISO/IEC FDIS 5259-5](https://standards.iteh.ai/catalog/standards/iso/745e0014-e40f-4160-b41a-4e2712730582/iso-iec-fdis-5259-5)

<https://standards.iteh.ai/catalog/standards/iso/745e0014-e40f-4160-b41a-4e2712730582/iso-iec-fdis-5259-5>



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	iv
Introduction	v
1 Scope	1
2 Normative references	1
3 Terms and definitions	1
4 Abbreviated terms	3
5 Data quality governance in the context of analytics and ML	4
5.1 Foundation.....	4
5.2 Ambiguous responsibilities for data.....	4
5.3 Purpose and justification.....	4
6 Data quality governance framework	5
6.1 General.....	5
6.2 DQ guiding principles.....	6
6.3 Strategies and policies for DQ.....	6
6.4 Business planning for DQ.....	6
6.5 DQ accountabilities.....	7
6.6 DQ risk management.....	7
6.7 Management processes for DQ.....	7
7 Responsibilities of governing body	8
7.1 Understand the strategic importance of data quality.....	8
7.2 Establish enabling environment for data quality governance.....	8
7.3 Formulate data quality strategies.....	9
7.4 Business planning for data quality.....	10
7.5 Data quality risk management capability.....	10
7.6 Set policies to ensure data quality.....	10
7.7 Establish oversight mechanisms.....	12
8 Responsibilities of management	12
8.1 Implement data quality strategies.....	12
8.2 Establish and enforce comprehensive data quality policies.....	12
8.3 Implement data quality management processes.....	12
8.4 Establishing internal risk control as part of management process.....	14
Bibliography	15

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/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.

This document was prepared by 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.

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.

Introduction

To address data quality properly without wasting critical resources, the organization's governing body can set the strategic direction for the use of analytics and machine learning (ML) and can oversee the quality of the needed data.

The data quality governance framework for analytics and ML assists the governing body in establishing a data quality governance within its organization with adequate controls across different layers of the organization throughout the data life cycle (DLC).

The framework can be used by both the governing body and management to interact and ensure the establishment of an effective data quality governance for analytics and ML at all levels in the organization.

The framework can be applicable regardless of an organization's size and type; and used in conjunction with other parts of the ISO/IEC 5259 series.

iTeh Standards (<https://standards.itih.ai>) Document Preview

[ISO/IEC FDIS 5259-5](https://standards.itih.ai/catalog/standards/iso/745e0014-e40f-4160-b41a-4e2712730582/iso-iec-fdis-5259-5)

<https://standards.itih.ai/catalog/standards/iso/745e0014-e40f-4160-b41a-4e2712730582/iso-iec-fdis-5259-5>

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

Part 5: Data quality governance framework

1 Scope

This document provides a data quality governance framework for analytics and machine learning (ML) to enable governing bodies of organizations to direct and oversee the implementation and operation of data quality measures, management, and related processes with adequate controls throughout the data life cycle (DLC) model according to ISO/IEC 5259-1.

This document can be applied to any analytics and ML. This document does not define specific management requirements or process requirements according to ISO/IEC 5259-3 and ISO/IEC 5259-4 respectively.

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

ISO/IEC 22989:2022, *Information technology — Artificial intelligence — Artificial intelligence concepts and terminology*

ISO/IEC 38505-1, *Information technology — Governance of IT — Governance of data — Part 1: Application of ISO/IEC 38500 to the governance of data*

ISO/IEC 38507:2022, *Information technology — Governance of IT — Governance implications of the use of artificial intelligence by organizations*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 5259-1, ISO/IEC 22989, ISO/IEC 38505-1 and ISO/IEC 38507, 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 creator

role within an organization responsible for generating, collecting and curating data from data sources

3.2

data owner

organization that is in the position to obtain, create, and have significant control over the content, access and distribution of data

Note 1 to entry: A data owner does not necessarily have a legal status with respect to data.

[SOURCE: ISO/TR 14872:2019, 3.4 — modified, Note 1 to entry replaced]

3.3

data steward

role within an organization responsible for ensuring that data-related work is performed according to policies and practices as established through data governance

[SOURCE: ISO/IEC TS 38505-3:2021, 3.9]

3.4

direct

communicate desired purposes and outcomes

Note 1 to entry: Within the context of governance of IT, directing involves setting objectives, strategies, and policies to be adopted by the members of the organization, to ensure that the use of IT meets organization's business objectives.

Note 2 to entry: Objectives, strategies, and policies can be set by management if they have the relevant authority delegated to them by the governing body.

[SOURCE: ISO/IEC 38500:2024, 3.1]

3.5

executive manager

person who has authority delegated from the governing body for implementation of strategies and policies to fulfil the purpose of the organization

Note 1 to entry: Executive managers can include roles which report to the governing body or the head of the organization or have overall accountability for major reporting function, for example Chief Executive Officers (CEOs), Heads of Government Organizations, Chief Financial Officers (CFOs), Chief Operating Officers (COOs), Chief Information Officers (CIOs), and similar roles.

Note 2 to entry: In management standards, executive managers can be referred to as top management.

3.6

governance

human-based system comprising directing, overseeing and accountability

[SOURCE: ISO/IEC 38500:2024, 3.3]

3.7

governing body

person or group of people who have ultimate accountability for the whole organization

Note 1 to entry: Every organizational entity has one governing body, whether or not it is explicitly established. When the organization is not an organizational entity, the term governing group is applicable where "governing body" is used throughout this document.

Note 2 to entry: A governing body can be explicitly established in a number of formats including, but not limited to, a board of directors, supervisory board, sole director, joint and several directors, or trustees.

Note 3 to entry: ISO management system standards make reference to the term "top management" to describe a role that, depending on the standard and organizational context, reports to, and is held accountable by, the governing body.

[SOURCE: ISO/IEC 37000:2021, 3.3.4]

3.8
monitor

review as a basis for appropriate decisions and adjustments

Note 1 to entry: Monitor involves routinely obtaining information about progress against plans as well as the periodic examination of overall achievements against agreed strategies and outcomes to provide a basis for decision making and adjustments to plans.

Note 2 to entry: Monitor includes reviewing compliance with relevant legislation, regulations, and organizational policies.

[SOURCE: ISO/IEC 38500:2024, 3.8]

3.9
principle

fundamental truth, proposition or assumption that serves as foundations for a set of beliefs or behaviours or for a chain of reasoning

[SOURCE: ISO/IEC 37000:2021, 3.2.1]

3.10
strategy

organization's overall plan of development, describing the effective use of resources in support of the organization in its future activities

Note 1 to entry: involves setting objectives and proposing initiatives for action

[SOURCE: ISO/IEC/IEEE 24765:2017, 3.4001]

3.11
management layer

organizational layer where exercise of control and supervision are performed within the constraints of governance

3.12
operational layer

organizational layer where daily routine operational tasks are performed

<https://standards.iteh.ai/catalog/standards/iso/745e0014-e40f-4160-b41a-4e2712730582/iso-iec-fdis-5259-5>

4 Abbreviated terms

CDO	chief data officer
CEO	chief executive officer
DLC	data life cycle
DQ	data quality
DQMLC	data quality management life cycle
DQP	data quality processes
DT	digital transformation
IT	information technology
ML	machine learning
PII	personally identifiable information
SW	software

HW hardware

5 Data quality governance in the context of analytics and ML

5.1 Foundation

Data for analytics and ML has its own set of unique characteristics compared to the traditional data generally used in business settings. The quality of training data plays a key role in the decision-making process using an ML model without human intervention. It is very important to produce quality outputs.

In order to safeguard data quality for analytics and ML effectively, a governing body should have adequate visibility into how data quality can impact analytics and ML.

Data quality can impact the results of analytics and ML if the input data have a problem with data quality characteristics such as accessibility, auditability, identifiability, portability, understandability, currentness, effectiveness and efficiency; and dataset quality characteristics such as accuracy, balance, consistency, scalability, diversity, effectiveness, generalizability, precision, relevance, representativeness, similarity and timeliness. A more detailed list of data and dataset quality characteristics and their definitions are described in ISO/IEC 5259-2:—. A governing body should understand these data quality characteristics and reflect them in its governance arrangements when consuming and producing data throughout the DLC model according to ISO/IEC 5259-1:2024.

5.2 Ambiguous responsibilities for data

The data used by analytics and ML can be provided by a great number of third parties and the functionality of the analytics and ML model can be primarily dependent on the data used. In this case, determining the relevant party becomes a critical issue if an ML model produces an inaccurate or incorrect inference or prediction due to an anomaly in data collected from multiple sensors or systems. In this case, an organization can have a system in place that can determine the cause (data or otherwise) if an ML model produces undesirable outcomes.

In addition, data for analytics and ML can be collected in various ways, including Internet search and posts on social media services. Despite the greater convenience from the diverse usage of such data, it is important to exercise caution due to serious problems that these kinds of data can present, such as privacy violations.

Data quality governance for analytics and ML entails a greater complexity in comparison to data quality issues that involve only a single data source. The governing body should understand that there should be clear roles and responsibilities established on how data for analytics and ML should be handled and processed throughout the entire DLC within and across organizational boundaries.

5.3 Purpose and justification

A holistic data quality governance framework for analytics and ML is needed for an organization to have adequate controls throughout the DLC model according to ISO/IEC 5259-1:2024. The data quality governance framework enables the governing body to direct and oversee the implementation and operation of data quality measures according to ISO/IEC 5259-2:—, data quality management requirements and guidelines according to ISO/IEC 5259-3:— and the data quality process framework for various types of ML according to ISO/IEC 5259-4:— throughout the DLC. The goal is to enhance trust in data for analytics and ML applications and services by mitigating data quality-related risks, making informed decisions, empowering effective and efficient operations across the organization.

The approach for an organization to enhance trust in data for analytics and ML should be to establish a robust and cross-cutting data quality governance framework across different levels of the organization with clear roles and responsibilities on how data should be handled and processed (see [Figure 1](#)). This document describes a data quality governance framework with which an organization develops its own data quality governance. The framework is applicable regardless of an organization's size and type. Individual organization's actual governance arrangement can differ according to their organizational structure, maturity and other relevant factors. Both the governing body and management can use the framework to