ISO/IEC FDIS 5259-3:20##(X)

ISO/IEC JTC 1/SC 42/WG 2

Secretariat: ANSI

Date: 2024-03-15

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

Part 3:
Data quality management requirements and guidelines

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

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

FDIS stage

Warning for WDs and CDs

This document is not an ISO International Standard. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an International Standard.

© ISO/IEC 202X – All rights reserved

Edited DIS - MUST BE USED FOR FINAL DRAFT

~~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.~~

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

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

 © ISO #### – All rights reserved

© ISO/IEC – All rights reserved

~~© ISO 202X~~

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

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

© 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

~~Fax: +41 22 749 09 47~~

~~Email~~E-mail: copyright@iso.org
Website: ~~www.iso.org~~www.iso.org

Published in Switzerland

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

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

Contents

Foreword viii
Introduction ix
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 3
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 6
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 8
7.2.3 Project-independent tailoring of the data quality management life cycle 9
7.2.4 Horizontal aspects of the data quality management life cycle 9
7.3 Requirements and recommendations 10
7.3.1 Data motivation and conceptualization 10
7.3.2 Data specification 11
7.3.3 Data planning 12
7.3.4 Data acquisition 13
7.3.5 Data preprocessing 15
7.3.6 Data augmentation 15
7.3.7 Data provisioning 16

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

© ISO/IEC 202X 2024 – All rights reserved

Edited DIS - MUST BE USED FOR FINAL DRAFT

7.3.8 Data decommissioning ... 18
7.4 Work products ... 19
7.4.1 Work products of data motivation and conceptualization stage ... 19
7.4.2 Work products of data specification stage ... 19
7.4.3 Work products of data planning stage ... 19
7.4.4 Work products of data acquisition stage ... 20
7.4.5 Work products of data preprocessing stage ... 20
7.4.6 Work products of data augmentation stage ... 20
7.4.7 Work products of data provisioning stage ... 20
7.4.8 Work products of data decommissioning stage ... 20
8 Horizontal processes ... 21
8.1 Objective ... 21
8.2 General ... 21
8.3 Requirements and recommendations ... 21
8.3.1 Verification and validation ... 21
8.3.2 Configuration management ... 21
8.3.3 Change management ... 22
8.3.4 Risk management ... 23
8.4 Work products ... 23
8.4.1 Work products of verification and validation ... 23
8.4.2 Work products of configuration management ... 24
8.4.3 Work products of change management ... 24
8.4.4 Work products for risk management ... 24
9 Management of data quality in supply chains ... 24
9.1 Objective ... 24
9.2 Requirements and recommendations ... 24
9.3 Work products ... 25
10 Management of data processing tools ... 25
10.1 Objective ... 25
10.2 Requirements and recommendations ... 25
10.3 Work products ... 26
11 Management of data quality dependencies ... 26
11.1 Objective ... 26
11.2 Requirements and recommendations ... 26
11.3 Work products ... 26
12 Project-specific data quality management ... 26
12.1 Objective ... 26
12.2 Requirements and recommendations ... 26
12.2.1 Context and intended use ... 26

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

 © ISO/IEC ~~202X~~ 2024 – All rights reserved

12.2.2 Objective 27
12.2.3 Requirements and recommendations 27
12.3 Specification and management of data quality requirements 27
12.3.1 Objective 27
12.3.2 Requirements and recommendations 27
12.4 Roles and responsibilities in data quality management 28
12.4.1 Objective 28
12.4.2 Requirements and recommendations 28
12.4.3 Work products 28
12.5 Tailoring of the data quality activities 28
12.6 Planning and coordination of the data quality activities 29
12.6.1 General 29
12.6.2 Data quality plan 29
12.6.3 Planning of processes 29
12.7 Progression of the data quality life cycle 29
12.8 Data quality justification 29
12.9 Decommissioning 30
12.10 Work products 30
Bibliography 31

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

© ISO/IEC 202X 2024 – All rights reserved

Edited DIS - MUST BE USED FOR FINAL DRAFT

Foreword

ISO (the International Organization for Standardization) ~~is a~~ and IEC (the International Electrotechnical Commission) form the specialized system for worldwide ~~federation of national standards~~standardization. National bodies ~~(~~that are members of ISO ~~member bodies). The work of preparing~~or IEC participate in the development of 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~~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. ~~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~~www.iso.org/directives or www.iec.ch/members_experts/refdocs).

ISO ~~draws~~and IEC draw attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO ~~takes~~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. ISO~~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.

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

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~~) see www.iso.org/iso/foreword.html. In the IEC, see www.iec.ch/understanding-standards.

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

© ISO/IEC ~~202X~~ 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.

This document is part of the ISO/IEC 5259 series. Other parts of the ISO/IEC 5259 series include:

— ISO/IEC 5259-1[1], *Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 1: Overview, terminology, and examples*;
— ISO/IEC 5259-2[2], *Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 2: Data quality measures*;
— ISO/IEC 5259-4[3], *Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 4: Data quality process framework*;
— ISO/IEC 5259-5[4], *Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 5: Data quality governance framework*;
— ISO/IEC 5259-6[5], *Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 6: Visualization framework for data quality.*

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

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

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

[1] Under preparation. Stage at the time of publication: ISO/IEC FDIS 5259-1:2024.
[2] Under preparation. Stage at the time of publication: ISO/IEC DIS 5259-2:2023.
[3] Under preparation. Stage at the time of publication: ISO/IEC FDIS 5259-4:2024.
[4] Under preparation. Stage at the time of publication: ISO/IEC DIS 5259-5:2023.
[5] Under preparation. Stage at the time of publication: ISO/IEC WD TR 5259-6:2023.

© ISO/IEC 202X 2024 – All rights reserved

Edited DIS - MUST BE USED FOR FINAL DRAFT

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

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

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

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

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

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

© 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~~

DQMLC data quality management life cycle

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

5 Intended usage

ISO/IEC FDIS 5259-3
https://standards.iteh.ai/catalog/standards/sist/[illegible]-b2c2-c4593045a3b4/iso-iec-fdis-5259-3

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

— by an organization to establish and tailor a data quality management ~~system~~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;

— 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 ~~system~~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.