

© ISO/IEC 202X – All rights reserved

ISO/IEC FDIS 5259-2:202X(X)

ISO/IEC JTC 1/SC 42/AWG 2

Secretariat: ANSI

Date: 2024-06-18

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

Part 2:  
Data quality measures

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

ISO/IEC FDIS 5259-2

<https://standards.itih.ai/catalog/standards/iso/a1df5595-1f27-4bcb-835f-3e7fle5354b8/iso-iec-fdis-5259-2>

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.

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.

Edited DIS - MUST BE USED FOR FINAL DRAFT

ISO #####-#:####(X)

*To help you, this guide on writing standards was produced by the ISO/TMB and is available at <https://www.iso.org/iso/how-to-write-standards.pdf>*

*A model manuscript of a draft International Standard (known as “The Rice Model”) is available at <https://www.iso.org/iso/model-document-rice-model.pdf>*

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

[ISO/IEC FDIS 5259-2](https://standards.iteh.ai/catalog/standards/iso/a1df5595-1f27-4bcb-835f-3e7f1e5354b8/iso-iec-fdis-5259-2)

<https://standards.iteh.ai/catalog/standards/iso/a1df5595-1f27-4bcb-835f-3e7f1e5354b8/iso-iec-fdis-5259-2>

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

[ISO/IEC FDIS 5259-2](https://standards.itih.ai/catalog/standards/iso/a1df5595-1f27-4bcb-835f-3e7f1e5354b8/iso-iec-fdis-5259-2)

<https://standards.itih.ai/catalog/standards/iso/a1df5595-1f27-4bcb-835f-3e7f1e5354b8/iso-iec-fdis-5259-2>

Edited DIS - MUST BE USED FOR FINAL DRAFT

ISO/IEC FDIS 5259-2:202X(X2024(en)

© 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](mailto:copyright@iso.org)  
Website: [www.iso.org](http://www.iso.org)

Published in Switzerland

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

ISO/IEC FDIS 5259-2

<https://standards.iteh.ai/catalog/standards/iso/a1df5595-1f27-4bcb-835f-3e7f1e5354b8/iso-iec-fdis-5259-2>

Contents

**Foreword**.....viii

**Introduction**.....ix

**1 Scope**.....1

**2 Normative references**.....1

**3 Terms and definitions**.....1

**4 Symbols and abbreviated terms**.....5

**5 Data quality components and data quality models for analytics and machine learning**.....6

**5.1 Data quality components in data life cycle**.....6

**5.2 Data quality model**.....7

**6 Data quality characteristics and quality measures**.....10

**6.1 General**.....10

**6.2 Inherent data quality characteristics**.....11

**6.2.1 Accuracy**.....11

**6.2.2 Completeness**.....12

**6.2.3 Consistency**.....14

**6.2.4 Credibility**.....15

**6.2.5 Currentness**.....16

**6.3 Inherent and system-dependent data quality characteristics**.....17

**6.3.1 Accessibility**.....17

**6.3.2 Compliance**.....17

**6.3.3 Efficiency**.....18

**6.3.4 Precision**.....18

**6.3.5 Traceability**.....19

**6.3.6 Understandability**.....19

**6.4 System-dependent data quality characteristics**.....20

**6.4.1 Availability**.....20

**6.4.2 Portability**.....20

**6.4.3 Recoverability**.....21

**6.5 Additional data quality characteristics**.....21

**6.5.1 Auditability**.....21

**6.5.2 Balance**.....22

**6.5.3 Diversity**.....24

**6.5.4 Effectiveness**.....25

**6.5.5 Identifiability**.....26

**6.5.6 Relevance**.....27

**6.5.7 Representativeness**.....28

**6.5.8 Similarity**.....28

**6.5.9 Timeliness**.....30

**7 Implementing a data quality model and data quality measures for an analytics or ML task**30

**8 Data quality reporting**.....31

**8.1 Data quality reporting framework**.....31

**8.2 Data quality measure information**.....31

**8.3 Guidance to organizations**.....32

**Annex A (informative) Design and document of a measurement function**.....34

**Annex B (informative) UML model of data quality measure framework**.....36

**Annex C (informative) Overview of data quality characteristics**.....38

<b>Annex D (informative) Alternative groups of data quality characteristics .....</b>	<b>40</b>
<b>Annex E (informative) Comparison between data quality characteristics of ISO/IEC 25012 and ISO/IEC 5259-2 .....</b>	<b>42</b>
<b>Bibliography .....</b>	<b>44</b>
<b>Foreword .....</b>	<b>v</b>
<b>Introduction .....</b>	<b>vi</b>
<b>1 Scope .....</b>	<b>1</b>
<b>2 Normative references .....</b>	<b>1</b>
<b>3 Terms and definitions .....</b>	<b>1</b>
<b>4 Symbols and abbreviated terms .....</b>	<b>4</b>
<b>5 Data quality components and data quality models for analytics and machine learning .....</b>	<b>5</b>
5.1 Data quality components in data life cycle .....	5
5.2 Data quality model .....	6
<b>6 Data quality characteristics and quality measures .....</b>	<b>8</b>
6.1 General .....	8
6.2 Inherent data quality characteristics .....	8
6.2.1 Accuracy .....	8
6.2.2 Completeness .....	9
6.2.3 Consistency .....	11
6.2.4 Credibility .....	12
6.2.5 Currentness .....	13
6.3 Inherent and system-dependent data quality characteristics .....	14
6.3.1 Accessibility .....	14
6.3.2 Compliance .....	15
6.3.3 Efficiency .....	15
6.3.4 Precision .....	15
6.3.5 Traceability .....	16
6.3.6 Understandability .....	16
6.4 System-dependent data quality characteristics .....	17
6.4.1 Availability .....	17
6.4.2 Portability .....	17
6.4.3 Recoverability .....	18
6.5 Additional data quality characteristics .....	18
6.5.1 Auditability .....	18
6.5.2 Balance .....	19
6.5.3 Diversity .....	21
6.5.4 Effectiveness .....	22
6.5.5 Identifiability .....	23
6.5.6 Relevance .....	24
6.5.7 Representativeness .....	25
6.5.8 Similarity .....	25
6.5.9 Timeliness .....	27
<b>7 Implementing a data quality model and data quality measures for an analytics or ML task .....</b>	<b>27</b>
<b>8 Data quality reporting .....</b>	<b>28</b>
8.1 Data quality reporting framework .....	28
8.2 Data quality measure information .....	28
8.3 Guidance to organizations .....	29

ISO/IEC FDIS 5259-2:202X(X2024)(en)

Annex A (informative) Design and document of a measurement function.....	30
Annex B (informative) UML model of data quality measure framework.....	32
Annex C (informative) Overview of data quality characteristics.....	33
Annex D (informative) Alternative groups of data quality characteristics.....	35
Annex E (informative) Comparison between data quality characteristics of ISO/IEC 25012 and ISO/IEC 5259-2.....	37
Bibliography.....	39

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

ISO/IEC FDIS 5259-2

<https://standards.itih.ai/catalog/standards/iso/a1df5595-1f27-4bcb-835f-3e7f1e5354b8/iso-iec-fdis-5259-2>

## 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](http://www.iso.org/directives) or [www.iec.ch/members\\_experts/refdocs](http://www.iec.ch/members_experts/refdocs)).

Field Code Changed

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](http://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](http://www.iso.org/iso/foreword.html). In the IEC, see [www.iec.ch/understanding-standards](http://www.iec.ch/understanding-standards).

Field Code Changed

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.

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](http://www.iso.org/members.html) and [www.iec.ch/national-committees](http://www.iec.ch/national-committees).

Field Code Changed



## Introduction

Data-supported decision-making brings new challenges to data quality management in data analytics and artificial intelligence (AI) based on machine learning (ML). Issues in data quality, such as incomplete, false or outdated data, can adversely affect analytics and ML processes and outcomes. Data from various sources, including structured data (e.g. relational databases) and unstructured data (e.g. documents, images, audios), can be directly consumed into the data life cycle for analytics and ML model development. Data are transformed in each stage of the data life cycle of analytics and ML. A holistic standardized approach to control, produce and deliver sufficient high-quality data is necessary for data analytics and ML models to be safe, reliable and interoperable. To develop credible data quality management for analytics and ML, intrinsic data quality International Standards, including concepts and use cases, characteristics and measurements, management requirements, and process framework, can be considered.

This document is a part of the ISO/IEC 5259 series. This document builds upon the ISO 8000 series, ISO/IEC 25012 and ISO/IEC 25024. The purpose of this document is to describe a data quality model through the definition of data quality characteristics and data quality measures based on ISO/IEC 25012 and ISO/IEC 25024. -Data quality models can be extended or modified according to this document.

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

[ISO/IEC FDIS 5259-2](#)

<https://standards.iteh.ai/catalog/standards/iso/a1df5595-1f27-4bcb-835f-3e7f1e5354b8/iso-iec-fdis-5259-2>



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

## Part 2: Data quality measures

### 1 Scope

This document specifies a data quality model, data quality measures and guidance on reporting data quality in the context of analytics and machine learning (ML).

This document is applicable to all types of organizations who want to achieve their data quality objectives.

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

ISO/IEC 25024, *Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Measurement of data quality*

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 5259-1, ISO/IEC 22989 and the following apply.

ISO and IEC maintain ~~terminological~~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 <http://www.electropedia.org/><https://www.electropedia.org/>

#### 3.1 data

re-interpretable representation of information in a formalized manner suitable for communication, interpretation, or processing

Note\_1\_to\_entry: Data can be processed by humans or by automatic means.

[SOURCE: ISO/IEC 2382:2015, 2121272]

#### 3.2 data frame

set of data records represented by a specific domain or purpose, with a shared structure of data items

## ISO/IEC FDIS 5259-2:2024(en)

Note\_1\_to\_entry: A data frame is two-dimensional, like a table with rows and columns. The term is specifically used in analytics and ML, e.g. in the R language, while other languages use “data set” to mean the same thing. In this document, “dataset” has a more generic meaning.

### 3.3 data type

categorization of an abstract set of possible values, characteristics, and set of operations for an attribute

Note\_1\_to\_entry: Examples of data types are character strings, texts, dates, numbers, images and sounds.

[SOURCE: ISO/IEC 25024:2015, 4.16]

### 3.4 data value

content of data item

Note\_1\_to\_entry: In ISO/IEC 25012:2008, 5.1.1, it is specified that from the “~~Inherent~~inherent” point of view, data quality refers to data itself such as data domain values and possible restrictions.

Note\_2\_to\_entry: Number or category assigned to an attribute of a target entity by making a measurement.

[SOURCE: ISO/IEC 25024:2015, 4.17]

### 3.5 empty data item

data item whose *data value* ~~(3.4)~~(3.4) has no value, i.e. Null or None

Note\_1\_to\_entry: This definition in general signifies non-existence of a data value (i.e. Null or None). A data item with string data type can be an empty data item by using either the empty string or Null. However, there is an exception for some application a string can be empty (e.g. “”) but not Null and hence not imply an empty data item.

### 3.6 entity

concrete or abstract thing in the domain under consideration [ISO/IEC FDIS 5259-2](#)

[SOURCE: ISO 8000-2:2022, 3.3.3]

### 3.7 raw data

data in its originally acquired, direct form from its source before subsequent processing

[SOURCE: ISO 5127:2017, 3.1.10.04]

### 3.8 target data

~~data (3.1)~~data (3.1) used in an analytics or ML task whose quality is measured

### 3.9 target population

population of interest in the analytics or ML project to which inferences are to be made

### 3.10 data quality subject

entity ~~(3.6)~~(3.6) affected by data quality

**3.11**  
**quality measure element**

measure defined in terms of a property and the measurement method for quantifying it, including optionally the transformation by a mathematical function

[SOURCE: ISO/IEC 25024:2015, 4.32]

**3.12**  
**quantity**

property of a phenomenon, body, or substance, where the property has a magnitude that can be expressed as a number and a reference

[SOURCE: ISO/IEC Guide 99:2007, 1.1, modified — Notes to entry deleted.]

**3.13**  
**quantity value**

number and reference together expressing magnitude of a *quantity* ~~(3.12)~~(3.12)

[SOURCE: ISO/IEC Guide 99:2007, 1.9, modified — Examples deleted.]

**3.14**  
**measurement function**

algorithm or calculation performed to combine one or more *quality measure elements* ~~(3.11)~~(3.11)

[SOURCE: ISO/IEC 25021:2012, 4.7, modified — Definition revised.]

**3.15**  
**measurement result**  
**result of measurement**

set of *quantity values* ~~(3.13)~~(3.13) being attributed to a measurand together with any other available relevant information

[SOURCE: ISO/IEC Guide 99:2007, 2.9, modified — Notes to entry deleted.]

**3.16**  
**measure**

<noun> variable to which a value is assigned as the result of measurement

Note\_1\_to\_entry:- The plural form “measures” is used to refer collectively to base measures, derived measures and indicators.

[SOURCE: ISO/IEC/~~IEEE~~ 15939:2017, 3.15]

**3.17**  
**measure**

<verb> make a measurement

[SOURCE: ISO/IEC 25000:2014, 4.19]

**3.18**  
**bounding box**

rectangular region enclosing annotated object

Note\_1\_to\_entry:- The major and minor axes of the rectangle are parallel to the edges of the images. For rotated boxes, the polygon annotation is to be used.

## ISO/IEC FDIS 5259-2:2024(en)

[SOURCE: ISO/IEC 30137-4:2021, 3.3]

### **3.19 cluster**

automatically induced category of elements that are part of the dataset and that share common attributes

Note\_1\_to\_entry: Clusters do not necessarily have a name.

[SOURCE: ISO/IEC 23053:2022, 3.3.2]

### **3.20 clustering algorithm**

algorithm which groups *clusters* (3.19)(3.19) from input data

Note\_1\_to\_entry: Examples of clustering algorithms include centroid-based clustering, density-based clustering, distribution-based clustering, hierarchical clustering and graph-based clustering.

### **3.21 overfitting**

<machine learning> creating a model which fits the training data too precisely and fails to generalize on new data

Note\_1\_to\_entry: Overfitting can occur because the trained model has learned from non-essential features in the training data (i.e. features that do not generalize to useful outputs), excessive noise in the training data (e.g. excessive number of outliers), a significant mismatch between training data and production data distributions or because the model is too complex for the training data.

Note\_2\_to\_entry: Overfitting can be identified when there is a significant difference between errors measured on training data and on separate test and validation data. The performance of overfitted models is especially impacted when there is a significant mismatch between training data and production data.

[SOURCE: ISO/IEC 23053:2022, 3.1.4]

### **3.22 fidelity**

degree to which a model or simulation reproduces the state and behaviour of a real-world object or the perception of a real-world object, feature, condition, or chosen standard in a measurable or perceivable manner

[SOURCE: ISO 16781:2021, 3.1.4]

### **3.23 maintainability**

ability of a functional unit, under given conditions of use, to be retained in, or restored to, a state in which it can perform a required function when maintenance is performed under given conditions and using stated procedures and resources

Note\_1\_to\_entry: The term used in IEV 191-02-07 is "maintainability performance" and the definition is the same.

Note\_2\_to\_entry: maintainability: term and definition standardized by ISO/IEC [ISO/IEC 2382-14:1997].

Note\_3\_to\_entry: 14.01.06 (2382)

[SOURCE: ISO/IEC 2382:2015, 2123027]