SLOVENSKI STANDARD
kSIST-TP FprCEN/CLC/TR 18115:2024

01-september-2024

Upravljanje in kakovost podatkov za UI v evropskem okviru

Data governance and quality for AI within the European context

Datenmanagement und -qualität für KI im europäischen Kontext

Gouvernance et qualité des données pour l'IA dans le contexte européen

Ta slovenski standard je istoveten z: FprCEN/CLC/TR 18115

ICS:

35.240.01	Uporabniške rešitve informacijske tehnike in tehnologije na splošno	Application of information technology in general

kSIST-TP FprCEN/CLC/TR 18115:2024   **en,fr,de**

TECHNICAL REPORT

RAPPORT TECHNIQUE

TECHNISCHER REPORT

FINAL DRAFT
FprCEN/CLC/TR 18115

June 2024

ICS 35.240.01

English version

Data governance and quality for AI within the European context

This draft Technical Report is submitted to CEN members for Vote. It has been drawn up by the Technical Committee CEN/CLC/JTC 21.

CEN and CENELEC members are the national standards bodies and national electrotechnical committees of Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Republic of North Macedonia, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Türkiye and United Kingdom.

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.

Warning : This document is not a Technical Report. It is distributed for review and comments. It is subject to change without notice and shall not be referred to as a Technical Report.

CEN-CENELEC Management Centre:
Rue de la Science 23, B-1040 Brussels

Ref. No. FprCEN/CLC/TR 18115:2024 E

FprCEN/CLC/TR 18115:2024 (E)

Contents

Page

Figures

Tables

FprCEN/CLC/TR 18115:2024 (E)

European foreword

This document (FprCEN/CLC/TR 18115:2024) has been prepared by Technical Committee CEN/JTC 21 "Artificial Intelligence", the secretariat of which is held by DS.

This document is currently submitted to the vote on TR.

Introduction

This document aims to provide an overview of the relevant regulations in the European context and connected international standards, paying particular attention to data governance and data quality topics. Relevant regulations considered are:

— "Council of Europe" Ad hoc Committee on AI (CAI) that produced "Recommendation CM/Rec (2020) of the Committee of Ministers to member States on the human rights impact of algorithmic systems" and the deliverable "possible elements of a legal framework on Artificial Intelligence, based on the Council of Europe's standards on human rights, democracy and the rules of law" (2021) [1].

— "European strategy for data" (2020), which is essential to govern new technologies and create business opportunities.

— "Artificial Intelligence Act" (2024 upcoming final version), which aims to ensure that AI systems placed on the market and used in the EU are safe and respect fundamental rights. Attention is given specifically to:

  — Article 10 "Data and data governance" describing the quality criteria specifying aspects of training, validation and testing of data sets.

  — Article 15 "Accuracy, robustness, and cybersecurity" describing essential quality characteristics that can be extended to a general data quality model; consistency between terms and definitions is a common goal of this document, as well as of future TS and EN standards.

  — Articles where standard quality characteristics are mentioned (see Figure 5).

— "Data Governance Act" (2022) providing a framework aiming:

  — to increase trust in data sharing across areas;

  — to develop common European data spaces in strategic domains (e.g. health, environment, energy, agriculture, mobility, finance, manufacturing, public administration;

  — to strengthen mechanism to increase data availability and overcome technical obstacles to the reuse of data.

— "Data Act" (2023): key elements include the reinforced data portability and data sharing, rules governing the processing data shared, model contracts, access and use data held by private companies, data and cloud interoperability, databases containing data from IoT, restriction on data sharing.

— "Open data Directive" (EU 2019/1024): provides common rules for a European market for government-held data, including the re-use of public sector information.

In addition, Regulation (EU) 2016/679 of the European Parliament and the Council on the protection of natural persons about the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC – GDPR, it is also considered in this document. The General Data Protection Regulation – GDPR, entered into force on May 2016, creates a harmonized set of rules applicable to processing of all European personal data. The objective of GDPR is to ensure that personal data enjoys a high standard of protection everywhere in the EU, increasing legal certainty for both individuals and organizations proceeding data, and offering a higher degree of protection for individuals and their fundamental rights. According to ISO/IEC 22989 types of organizations are e.g. commercial

FprCEN/CLC/TR 18115:2024 (E)

enterprises, governments agencies, not-for-profit organizations. The objective of GDPR is to provide a consistent and high level of protection of natural persons regarding the processing of personal data and the free movement of such data and to remove the obstacles to the flow of personal data within the Union. In addition, GDPR ensures a common level of protection of the rights and freedoms of natural persons concerning the processing of such data all over the Member States, increasing legal certainty for both individuals and organizations proceeding with data and offering a higher degree of protection for individuals and their fundamental rights.

GDPR takes into consideration also the activity of processing personal data by Artificial Intelligence systems (see processing reported in 3.2.10), as we will see explaining characteristics of data quality containing specific requirements on this topic strongly related to some principles of GDPR and as can also be seen in some documents of the Council of Europe COE [1].

Another important aspect of quality underlined in this document it is related to accessibility for disabled users. In this case also we will describe the concepts explaining characteristics of data quality the value of accessibility, and understandability of data. The accessibility quality characteristic related to a European legislative regulation is a good example of governance of data that are obtained with a global vision by monitoring the activities in progress in each Country. A similar approach of governance, global and local, can be extended in the future to the large applications of AI, developing specific EN Standards or Technical specification.

Finally, some considerations on ethics are reported to reinforce some aspects related to data use.

The European Commission and the Member States put forward a 'Coordinated Plan on Artificial Intelligence' - COM (2018) 795 - with the stated goal of maximizing AI investments impact both at European and national levels and strengthening synergies and cooperation among Member States. To this end, Member States were strongly encouraged to develop their own national AI strategies (e.g. with Guidelines and monitoring specifications) to achieve these aims, in conformance with laws.
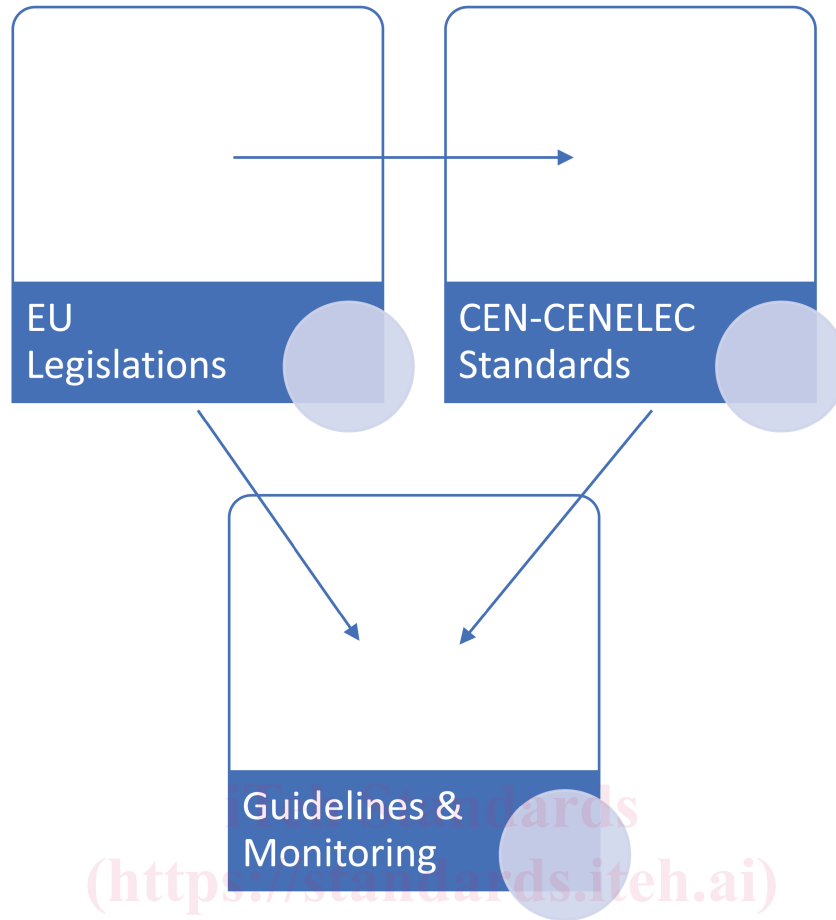
Figure 1 — Connections of legislations, standards, guidelines, monitoring specifications

Following these principles, the goal of this document is promoting a complement to the overview of common terminology and language on Artificial Intelligence to facilitate innovation, communications, coordination, planning and agreements between European countries, national visions, enterprises, projects and products realization oriented to quality. For innovation management the approach taken in the ISO 56000 family can be considered. For social motivation and responsibility, ISO 26000 can contribute to sustain the inclusiveness and ethics principles.

FprCEN/CLC/TR 18115:2024 (E)

1 Scope

This document provides an overview on AI-related standards, with a focus on data and data life cycles, to organizations, agencies, enterprises, developers, universities, researchers, focus groups, users, and other stakeholders that are experiencing this era of digital transformation.

It describes links among the many international standards and regulations published or under development, with the aim of promoting a common language, a greater culture of quality, giving an information framework.

It addresses the following areas:

— data governance;

— data quality;

— elements for data, data sets properties to provide unbiased evaluation and information for testing.

2 Normative references

There are no normative references in this document.

NOTE   For the application of this document: users and stakeholders can apply the standards listed depending on their context of use and in compliance with the laws.

3 Terms and definitions

For the purposes of this document, the following terms and definitions 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/

Note 1 to entry: Terms and definitions have been divided into General, Data Governance and Data quality.

3.1 General

3.1.1
Artificial Intelligence
AI
research and development of mechanism and applications of AI systems

Note 1 to entry: Research and development can take place across any number of fields such as computer science, data science, humanities, mathematics, and natural sciences

[SOURCE: ISO/IEC 22989:2022[1]]

---

[1] ISO/IEC 22989:2022/AMD1 is under development.

3.1.2
AI system
engineered system that generates outputs such as content, forecasts, recommendations, or decisions for a given set of human-defined objectives

Note 1 to entry: The engineered system can use various techniques and approaches related to artificial intelligence to develop a model to represent data. Knowledge, processes, etc. which can be used to conduct tasks.

Note 2 to entry: AI systems are designed to operate with varying levels of automation.

[SOURCE: ISO/IEC 22989:2022[2]]

3.1.3
element
smaller part of and architecture

EXAMPLES   records, fields, format, metadata, images, etc.

[SOURCE: ISO/IEC 25024:2015; in ISO/IEC 25024:2015, 4.19, the term is used with reference to the architecture of data and to computer program domain such as data model or data dictionary.]

3.1.4
framework
reusable design (models or code) that can be refined (specialized) and extended to provide some portion of the overall functionality of many applications

[SOURCE: IEEE 1320.2-1998 (R2004)]

3.1.7
life cycle
evolution of a system, product, service, project or other human-made entity, from conception through retirement

[SOURCE: ISO/IEC 22989:2022; ISO/IEC/IEEE 15288:2023]

3.1.8
measure
variable to which a value is assigned as the result of measurement

Note 1 to entry: the term measure is used to refer collectively to base measures, derived measures and indicators.

[SOURCE: ISO/IEC 25024:2015, 4.26, ISO/IEC 25010:2011, 4.4.5, ISO/IEC/IEEE 15939:2017]

3.1.9
measurement
set of operations having the object of determining a value of a measure

[SOURCE: ISO/IEC 25024:2015; ISO 3951-5:2006]

---

[2] ISO/IEC 22989:2022/AMD1 is under development.

FprCEN/CLC/TR 18115:2024 (E)

3.1.10
metric
defined measurement method and measurement scale

[SOURCE: ISO/IEC 14102:2008]

3.1.11
process
set of interrelated or interacting activities which transforms inputs into outputs

[SOURCE: ISO/IEC/IEEE 12207:2017]

3.1.12
product
result of a process

[SOURCE: ISO/IEC/IEEE 12207:2017]

3.1.13
property
property to quantify
property of a target entity that is related to a quality measure element and which can be quantified by a measurement method

[SOURCE: ISO/IEC 25021:2012, Figure 5, reported in Figure 12 of this document]

3.1.14
quality model
defined set of characteristics and of relationships between them, which provides a framework for specifying quality requirements and evaluating quality

[SOURCE: ISO/IEC 25000:2014]

3.1.15
system
combination of interacting elements organized to achieve one or more stated purposes

[SOURCE ISO/IEC 25000:2014]

3.2 Data governance

3.2.1
corporate governance
system by which corporations are directed and controlled

[SOURCE: ISO/IEC 38500:2024]

3.2.2
data governance
execution and enforcement of authority over the definition, production, and usage of data related assets

[SOURCE: IEEE 7005:2021]

3.2.3
data governance framework
strategy, policies, decision-making structures and accountabilities, through which the organization's governance arrangements operate on data

[SOURCE: ISO/IEC TR 38502:2017, modified – the data are specified]

3.2.4
governance
process for establishing and enforcing strategic goals and objectives, organizational policies, and performance parameters

[SOURCE: Software Extension to the PMBOK (R) Guide Fifth Edition) ISO/IEC/IEEE 21840:2019]

3.2.5
governing body
person or group of people who are accountable for the performance and conformance of the organization

[SOURCE: ISO/IEC 5259-5[3]; ISO/IEC 38500:2024]

3.2.6
management
system of controls and processes required to achieve the strategic objectives set by the organization's governing body

[SOURCE: ISO/IEC/IEEE 21840:2019]

3.2.7
strategy
organization's overall plan of development, describing the effective use of resources in support of the organization in its future activities. It involves setting objectives and proposing initiatives for action

[SOURCE: ISO/IEC/IEEE 24765]

3.2.8
process
predetermined course of events that occur during the execution of all or part of a program

[SOURCE: ISO/IEC 2382-1:2015]

3.2.9
personal data
<processing of personal data> any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person

[SOURCE: Regulation (EU) 2016/679 (GDPR) [28], Article 4 (1)]

---

[3] Under preparation. Current stage: ISO/IEC FDIS 5259-5:2024.

FprCEN/CLC/TR 18115:2024 (E)

3.2.10
processing
<processing of personal data> any operation or set of operations which is performed on personal data or on sets of personal, whether or not by automated means, such as collection, recording, organization, structuring, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction

[SOURCE: Regulation (EU) 2016/679 (GDPR) [28], Article 4 (2)]

3.2.11
product
result of a process

[SOURCE: ISO/IEC/IEEE 12207:2017, 3.1.36; ISO/IEC/IEEE 24748-1:2024, 3.34)

3.3 Data quality

3.3.1
analytics
composite concept consisting of data acquisition, validation, processing, including quantification, visualization and interpretation

[SOURCE: ISO/IEC 5259-1:—[4], modified; ISO/IEC 20546:2019, modified]

3.3.2
big data
extensive datasets, primarily in the data characteristics of volume, variety, velocity, and/or variability, that require a scalable technology for efficient storage, manipulation, management and analysis

[SOURCE: ISO/IEC 20546:2019]

3.3.3
data
reinterpretable 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 25012:2008, ISO/IEC 2382-1:2015]

Note 2 to entry: The reinterpretable representation is connected to the data attributes that enable to be read and interpreted by users (see ISO/IEC 25012:2008, 5.3.2.7).

---

[4] Under preparation. Current stage: ISO/IEC FDIS 5259-1:2024.