

TECHNICAL REPORT

CEN/CLC/TR 18115

RAPPORT TECHNIQUE

TECHNISCHER REPORT

November 2024

ICS 35.240.01

English version

Data governance and quality for AI within the European context

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

Datenmanagement und -qualität für KI im europäischen Kontext

This Technical Report was approved by CEN on 30 September 2024. 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.



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

Contents	Page
European foreword.....	4
Introduction	5
1 Scope	8
2 Normative references	8
3 Terms and definitions.....	8
3.1 General	8
3.2 Data governance	10
3.3 Data quality.....	12
4 Abbreviations.....	14
5 JRC research and data-related standards on AI.....	15
5.1 General	15
5.2 Research: Data quality requirements for inclusive, non-biased and trustworthy AI.....	16
5.3 Data-related standards on AI for data governance and data quality.....	18
5.3.1 General	18
5.3.2 A short description of the standards mentioned in Figure 4 (taken from www.iso.org).	19
6 Data governance	24
7 Data quality.....	35
8 Elements for data, data sets, information for testing and evaluation	45
9 Data governance and data quality for large European contexts	50
9.1 General	50
9.2 Italian government: Strategy program on Artificial Intelligence.....	50
9.3 Italian agency application of data quality model for public administrations	51
9.4 Spanish experience on data Governance: Data Office	52
9.5 European governance relating to the Directive on inclusivity and accessibility	53
10 General considerations on innovative technology: Ethics, Governance, AI Act.....	54
11 Potential challenges.....	56
11.1 General.....	56
11.2 Stakeholders' engagement.....	56
11.3 Contextualization	56
11.4 Critical infrastructures.....	57
11.5 Ethics and regulatory challenges.....	57
11.6 Interoperability	58
11.7 Big volume of data.....	59
12 Best practices from organizations, industries and research activities.....	59
12.1 General	59

12.2	AI in healthcare: the MES-CoBraD approach	59
12.3	Overview of industries that stand out for their approach to data governance	60
	Bibliography.....	62
Figures		
Figure 1	— Connections of Legislations, Standards, Guidelines & Monitoring specifications.....	7
Figure 2	— Active organizations mentioned in JRC.....	17
Figure 3	— Standards and Technical reports mentioned in JRC	18
Figure 4	— Clusters of standards, TS, TR data-related	18
Figure 5	— Example of relationships among quality aspects of ISO/IEC 5259-2, ISO/IEC 25059 and AI Act , eliciting new requirements to be harmonized	23
Figure 6	— European legal references and ISO standards for AI on data quality (or complementary)	26
Figure 7	— Data governance framework.....	27
Figure 8	— Data governance flow at European level.....	30
Figure 9	— Data managing integration and synthesis of experiences	32
Figure 10	— Data Governance summary	34
Figure 11	— Data Quality Measures and Data Life Cycle Model	36
Figure 12	— Relationship among quality models, characteristics, QM, QME, property, target entity	38
Figure 13	— Data life cycle framework	44
Figure 14	— Example of conceptual perspective visualization of data testing and evaluation...	47
Figure 15	— Visualization of elements for governance of data, data sets, testing.....	47
Figure 16	— Example of ontological contextual schema of elements resulting in the conference online held in October 2020 with 100 speakers [13].....	49
Tables		
Table 1	— Main documents considered for data governance framework.....	24
Table 2	— Type of governance and multi-level point of view	28
Table 3	— Characteristics of the data quality model adapted from ISO/IEC 25012	37
Table 4	— Characteristics of data quality models from ISO 8000, ISO/IEC 25012 and ISO/IEC 5259-2	41

European foreword

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

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. CEN shall not be held responsible for identifying any or all such patent rights.

Any feedback and questions on this document should be directed to the users’ national standards body. A complete listing of these bodies can be found on the CEN website.

This document is a preview generated by EVS

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), 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 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.

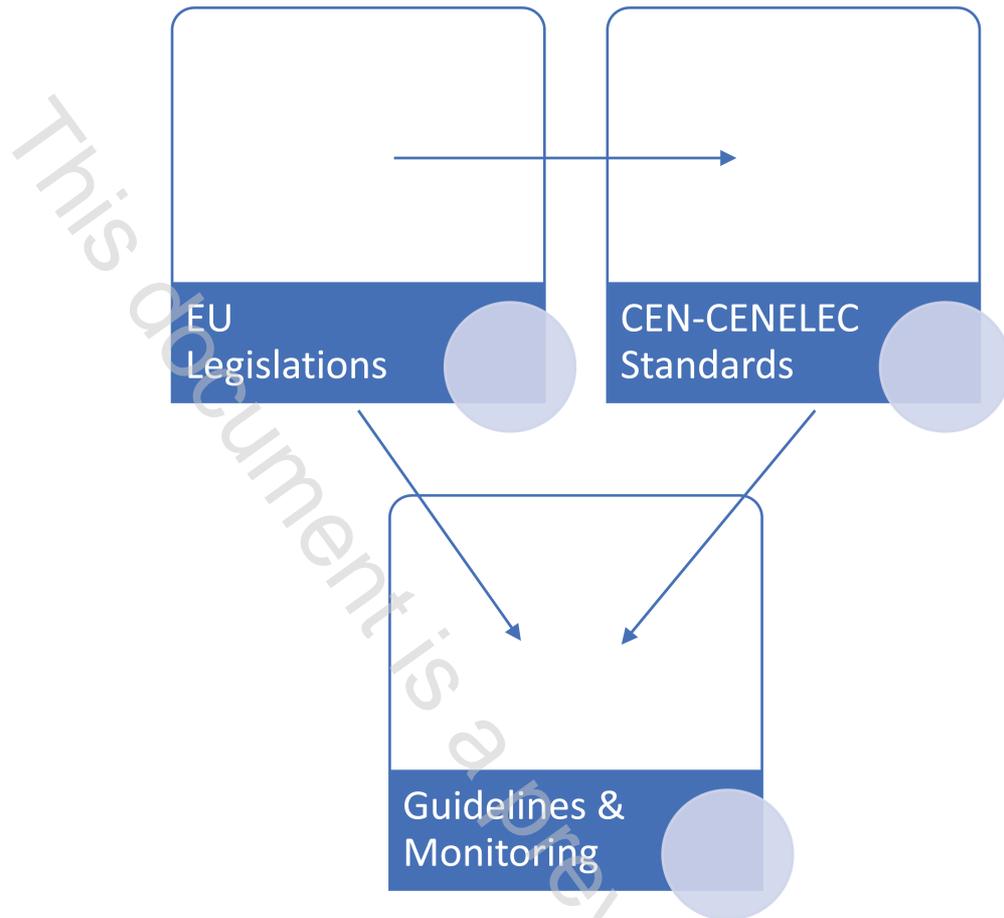


Figure 1 — Connections of Legislations, Standards, Guidelines & Monitoring specifications

EU AI Act and CEN-CENELEC JTC21 are harmonizing legislations and Standards. Guidelines & Monitoring can be developed by Member States / Companies: examples are quoted in Clause 9 and 12 of this TR. Following these perspectives, the goal of this document is promoting a complement to the overview of a 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 and mitigating risks. 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.

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 mechanisms 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 22989:2022¹]

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