INTERNATIONAL STANDARD

ISO/IEC 8183

First edition
2023-07

Information technology — Artificial intelligence — Data life cycle framework

Reference number
ISO/IEC 8183:2023(E)

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

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 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 Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 42, *Artificial intelligence*.

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

Artificial intelligence (AI) systems are being adopted by organizations of all types, sizes and purposes. Data are essential to the development and operation of AI systems.

In the field of AI systems, there are many data life cycles in use and under consideration for different purposes (e.g. data quality, bias in data, data governance, development and use of AI systems). Without an overarching framework, these different data life cycles can be challenging to correctly interpret by those without previous knowledge, context and expertise. There is a risk that these multiple data life cycles will not be applied as intended.

This document provides a data life cycle overview in Clause 5, describes a data life cycle framework in Clause 6 and provides more information on the stages or processes of the data life cycle in Clause 7.

Information technology — Artificial intelligence — Data life cycle framework

1 Scope

This document defines the stages and identifies associated actions for data processing throughout the artificial intelligence (AI) system life cycle, including acquisition, creation, development, deployment, maintenance and decommissioning. This document does not define specific services, platforms or tools. This document is applicable to all organizations, regardless of type, size or nature, that use data in the development and use of AI systems.

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 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 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/

4 Symbols and abbreviated terms

AI	artificial intelligence
DPIA	data protection impact assessment
JSON	JavaScript object notation
ML	machine learning
OWL	web ontology language
PII	personally identifiable information
XML	extensible markup language

5 Data life cycle overview

The data life cycle for AI systems encompasses the processing of data from the earliest conception of a new AI system to the eventual decommissioning of the system and is separated into a number of distinct stages. Each stage will often, but not always, be part of a data life cycle for an AI system.

A data life cycle represents all the stages through which data can pass within any system that uses data of any kind. It is designed to support the achievement of objectives related to system governance, system utility, data quality and data security, by ensuring that data processing is given due consideration during the planning, development, use and decommissioning of the system.

The detailed purpose and timing of use of these stages throughout the life cycle are influenced by multiple factors, including societal, commercial, organizational and technical considerations, each of which can vary or at times be combined with other stages during the life of a system. This document describes the following 10 stages:

— stage 1 – idea conception;

— stage 2 – business requirements;

— stage 3 – data planning;

— stage 4 – data acquisition;

— stage 5 – data preparation;

— stage 6 – building model;

— stage 7 – system deployment;

— stage 8 – system operation;

— stage 9 – data decommissioning;

— stage 10 – system decommissioning.

For information about a data life cycle for data usage, see ISO/IEC 5212: —[1)].

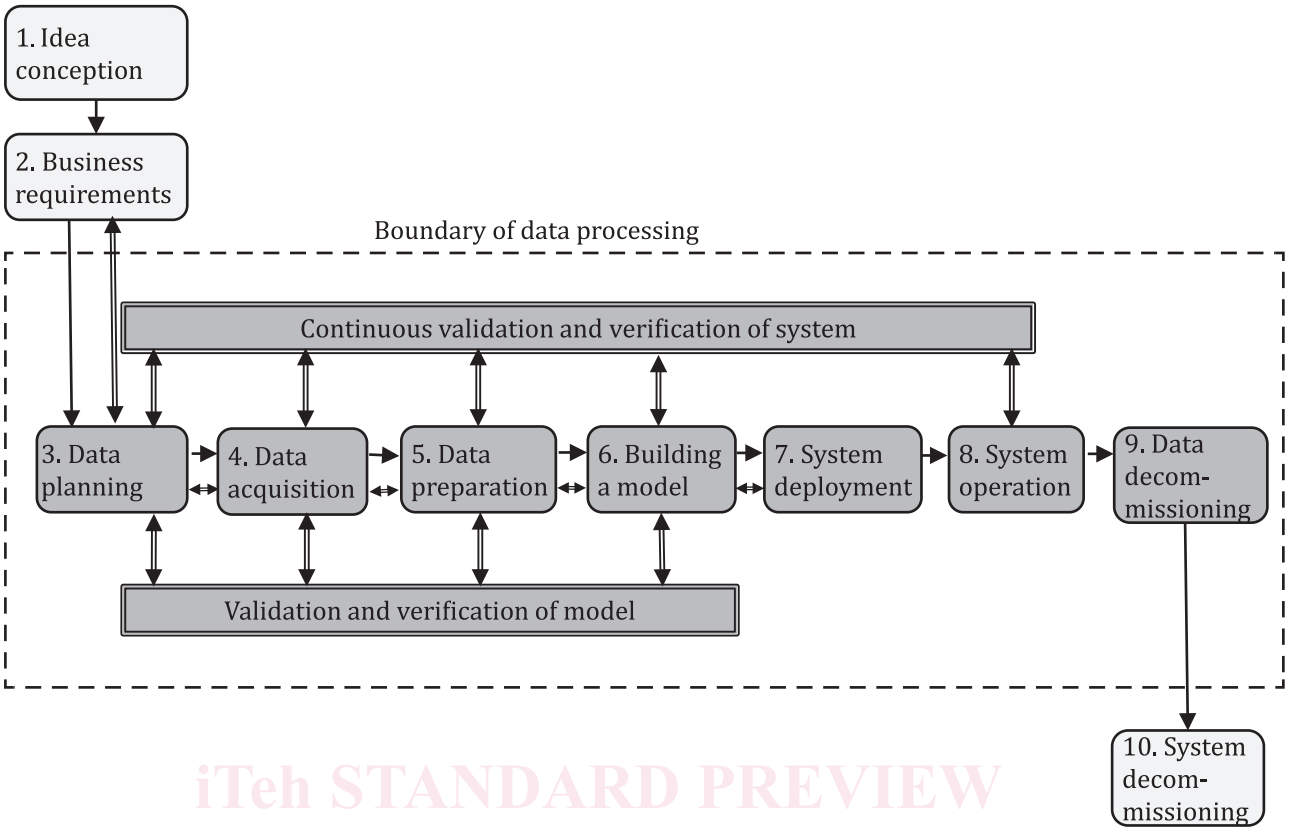
6 Data life cycle framework

6.1 General

The data life cycle framework, shown in Figure 1, identifies a set of conceptually distinct stages that data used in an AI system go through from data planning to data decommissioning. Figure 1 also includes idea conception, business requirements and system decommissioning, which are system-level life cycle stages. For information regarding data sets, refer to ISO/IEC 23053:2022, 6.5. Life cycle processes appropriate to a defined task can be assigned to each stage. Life cycle processes describe the actions taken on the data within the life cycle stage.

Stage 9 (data decommissioning) and stage 10 (system decommissioning) both pertain to decommissioning but stage 9 specifically covers what happens to the data (e.g. secure deletion, archiving, repurposing) while stage 10 covers what happens to the system irrespective of what happens to the data that is being processed.

1) Under preparation. Stage at the time of publication: ISO/IEC DIS 5212:2023.



Key

→ primary development pathway

↔ feedback pathway

stage with data processing

stage outside the data processing boundary

NOTE 1 The single-headed arrows depict a linear path through the life cycle stages, while the double-headed arrows show feedback paths between life cycle stages.

NOTE 2 The verification and validation of the model refers to the internal development process, whose output is a model. The validation and verification of the system refers to the system as a whole, extending through its entire period of operation.

Figure 1 — Data life cycle framework

6.2 Stage 1: Idea conception

Idea conception is when a need or requirement for a new or revised AI system is recognized. The AI system can be used as a partial or complete solution to an existing or potential problem or opportunity faced by the organization.

Idea conception can also be driven by broader organizational context needs (e.g. economic, technical, strategic, market or legal requirements). Ultimately, this idea can be expressed as one or more questions that the AI system can answer. The questions to which the AI system provides answers should be mapped to and aligned with business objectives and metrics.

6.3 Stage 2: Business requirements

The business requirements stage can include one or more stakeholders with appropriate authority or influence deciding to investigate whether the idea can be turned into a functioning system and deciding whether to invest further in the idea. This stage involves:

— determining the ambition of the project (e.g. vision, goals and strategy);

— determining assets, including those available and those that need to be acquired;

— specifying the data requirements, a key element for AI systems, based on the business goals and end user requirements;

— identifying enablers for the project, including in-house skills and knowledge, organizational architecture, technology and external resources;

— ensuring the project can be developed in line with organizational policies and procedures (or processes), including:

  — compliance (e.g. privacy requirements);

  — ethics (e.g. fairness of outcomes);

  — culture;

  — leadership;

  — governance processes.

The business requirements stage can conclude with a determination of whether or not the project is feasible.

NOTE The business requirements stage does not include processing of the data.

6.4 Stage 3: Data planning

The data planning stage involves deciding upon the scope of the data required to address the questions identified in the business requirements stage. The primary data factors for consideration at this stage include:

— whether the necessary data exist, are available for reuse, need to be acquired, collected, transformed, authored or curated, or a combination of some or all of these;

— amount of data required;

— source of the data;

— whether synthetic, i.e. artificial, data can be created to augment available data;

— what data outputs are created and how the system will deal with them;

— format of the data;

— what the data represent;

— properties of the data that can affect the choice of algorithm for model development;

— data licensing requirements;

— data security, privacy and resilience requirements;

— data acquisition requirements, such as data collection requirements;

— data safety requirements;