DRAFT INTERNATIONAL STANDARD

ISO/IEC DIS 8183

ISO/IEC JTC **1**/SC **42**

Secretariat: **ANSI**

Voting begins on:
2022-07-26

Voting terminates on:
2022-10-18

Information technology — Artificial intelligence — Data life cycle framework

ICS: 35.020

This document is circulated as received from the committee secretariat.

Reference number
ISO/IEC DIS 8183:2022(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

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. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

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).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC will not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on 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 the following URL: www.iso.org/iso/foreword.html.

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.

Introduction

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

In the field of AI systems, there are any number of 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.

The purpose of this document is to provide a data life cycle framework, including terms and concepts, that can be referenced by specialized data life cycles. The aim is to make it easier for users in different roles to understand and correlate specialized data life cycles and how they apply to their organization's needs by describing a set of high-level data life cycle stages.

Information technology — Artificial intelligence — Data life cycle framework

1 Scope

This document is applicable to the data processing throughout the AI system life cycle including the acquisition, creation, development, deployment, maintenance and decommissioning. This document is applicable to the acquisition, creation, development, deployment, maintenance and decommissioning of data in AI systems. This document does not define specific services, platforms or tools. This document is applicable to all organizations, regardless of type, sizes and 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:-1), *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 DIS 22989:— apply.

ISO and IEC maintain terminological 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/

4 Symbols and abbreviated terms

AI	artificial intelligence
DPIA	data protection impact assessment
JSON	JavaScript object notation
ML	machine learning

1) Under preparation. Stage at the time of publication: ISO/IEC FDIS 22989:2022

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 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.

6 Data life cycle framework

6.1 General

The data life cycle framework, shown in Figure 1, identifies a set of 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 datasets refer to FDIS ISO/IEC 23053:-2) [3]. 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. deletion, destruction, return) while stage 10 covers what happens to the system irrespective of what happens to the data.

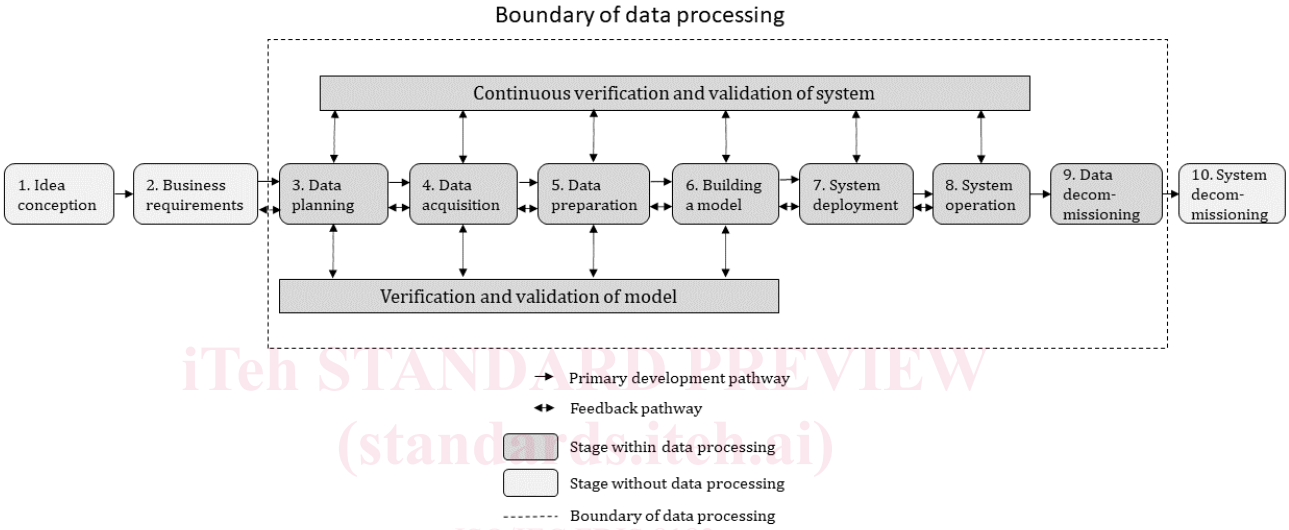

Figure 1 — Data life cycle framework

NOTE 1 The single-headed arrows in Figure 1 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.

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 a complete solution to an existing or potential problem or opportunity faced by the organization.

2) Under preparation. Stage at the time of publication: ISO/IEC FDIS 23053:—

Idea conception can also be driven by broader organizational context needs (e.g. economic, technical, strategic, market or legal requirements). Ultimately, this idea should be expressed as one or more questions that the AI system can answer.

6.3 Stage 2: Business requirements

The business requirements stage involves one or more stakeholders, with appropriate authority or influence deciding 1) to investigate whether the idea can be turned into a functioning system and 2) 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 can be defined at this stage;

— 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 concludes with a determination of whether or not the project is feasible.

NOTE No data is processed at the business requirements stage.

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. The primary data factors for consideration at this stage include:

— data existence: whether the necessary data exists, is available for reuse, needs to be acquired, transformed, authored, curated, etc. or a combination of some or all of these;

— amount of data required;

— source of data;

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

