



BSI Standards Publication

Geographic information — Training data markup language for artificial intelligence

Part 1: Conceptual model

National foreword

This British Standard is the UK implementation of ISO 19178-1:2025.

The UK participation in its preparation was entrusted to Technical Committee IST/36, Geographic information.

A list of organizations represented on this committee can be obtained on request to its committee manager.

Contractual and legal considerations

This publication has been prepared in good faith, however no representation, warranty, assurance or undertaking (express or implied) is or will be made, and no responsibility or liability is or will be accepted by BSI in relation to the adequacy, accuracy, completeness or reasonableness of this publication. All and any such responsibility and liability is expressly disclaimed to the full extent permitted by the law.

This publication is provided as is, and is to be used at the recipient's own risk.

The recipient is advised to consider seeking professional guidance with respect to its use of this publication.

This publication is not intended to constitute a contract. Users are responsible for its correct application.

© The British Standards Institution 2025
Published by BSI Standards Limited 2025

ISBN 978 0 539 30629 3

ICS 35.240.70

Compliance with a British Standard cannot confer immunity from legal obligations.

This British Standard was published under the authority of the Standards Policy and Strategy Committee on 30 June 2025.

Amendments/corrigenda issued since publication

Date	Text affected
------	---------------

English Version

Geographic information - Training data markup language
for artificial intelligence - Part 1: Conceptual model (ISO
19178-1:2025)

Information géographique - Langage de balisage des
données d'entraînement pour l'intelligence artificielle -
Partie 1: Modèle conceptuel (ISO 19178-1:2025)

This European Standard was approved by CEN on 3 June 2025.

CEN members are bound to comply with the CEN/CENELEC Internal Regulations which stipulate the conditions for giving this European Standard the status of a national standard without any alteration. Up-to-date lists and bibliographical references concerning such national standards may be obtained on application to the CEN-CENELEC Management Centre or to any CEN member.

This European Standard exists in three official versions (English, French, German). A version in any other language made by translation under the responsibility of a CEN member into its own language and notified to the CEN-CENELEC Management Centre has the same status as the official versions.

CEN members are the national standards bodies 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.



EUROPEAN COMMITTEE FOR STANDARDIZATION
COMITÉ EUROPÉEN DE NORMALISATION
EUROPÄISCHES KOMITEE FÜR NORMUNG

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

European foreword

This document (EN ISO 19178-1:2025) has been prepared by Technical Committee ISO/TC 211 "Geographic information/Geomatics" in collaboration with Technical Committee CEN/TC 287 "Geographic Information" the secretariat of which is held by BSI.

This European Standard shall be given the status of a national standard, either by publication of an identical text or by endorsement, at the latest by December 2025, and conflicting national standards shall be withdrawn at the latest by December 2025.

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/national committee. A complete listing of these bodies can be found on the CEN website.

According to the CEN-CENELEC Internal Regulations, the national standards organizations of the following countries are bound to implement this European Standard: 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 the United Kingdom.

Endorsement notice

The text of ISO 19178-1:2025 has been approved by CEN as EN ISO 19178-1:2025 without any modification.

Contents

Page

Foreword	v
Introduction	vi
1 Scope	1
2 Normative references	1
3 Terms, definitions and abbreviated terms	1
3.1 Terms and definitions	1
3.2 Abbreviated terms	3
4 Conventions	4
4.1 General	4
4.2 Identifiers	4
4.3 UML notation	4
5 Conformance	6
6 Overview	6
6.1 General	6
6.2 AI tasks for EO	6
6.3 Modularization	7
6.4 General modelling principles	8
6.4.1 Element modelling	8
6.4.2 Class hierarchy and inheritance of properties and relations	8
6.4.3 Definition of the semantics for all classes, properties and relations	9
6.4.4 Data integrity, authenticity and non-repudiation	9
6.5 Extending TrainingDML-AI	9
7 TrainingDML-AI UML model	9
7.1 General	9
7.2 ISO dependencies	9
7.3 Overview of the UML model	10
7.4 AI_TrainingDataset	12
7.4.1 General	12
7.4.2 Provisions	13
7.4.3 Class definitions	14
7.5 AI_TrainingData	15
7.5.1 General	15
7.5.2 Provisions	16
7.5.3 Class definitions	16
7.6 AI_Task	17
7.6.1 General	17
7.6.2 Provisions	18
7.6.3 Class definitions	18
7.7 AI_Label	18
7.7.1 General	18
7.7.2 Provisions	19
7.7.3 Class definitions	20
7.8 AI_Labeling	20
7.8.1 General	20
7.8.2 Provisions	21
7.8.3 Class definitions	22
7.9 AI_TDChangeset	22
7.9.1 General	22
7.9.2 Provisions	23
7.9.3 Class definitions	24
7.10 AI_DataQuality	24
7.10.1 General	24

	7.10.2 Provisions.....	25
	7.10.3 Class definitions.....	26
8	TrainingDML-AI Data Dictionary	26
8.1	General.....	26
8.2	ISO Classes.....	26
8.2.1	Feature (from ISO 19101-1).....	26
8.2.2	MD_Band (from ISO 19115-1).....	26
8.2.3	MD_Scope (from ISO 19115-1).....	26
8.2.4	MD_ReferenceSystem (from ISO 19115-1).....	26
8.2.5	LI_Lineage (from ISO 19115-1).....	27
8.2.6	EX_Extent (from ISO 19115-1).....	27
8.2.7	CI_Citation (from ISO 19115-1).....	27
8.2.8	MD_Resolution (from ISO 19115-1).....	27
8.2.9	DataQuality (from ISO 19157-1).....	27
8.2.10	QualityElement (from ISO 19157-1).....	28
8.3	AI_TrainingDataset.....	28
8.3.1	Metadata.....	28
8.3.2	Classes.....	28
8.4	AI_TrainingData.....	30
8.4.1	Metadata.....	30
8.4.2	Classes.....	30
8.5	AI_Task.....	31
8.5.1	Metadata.....	31
8.5.2	Classes.....	32
8.6	AI_Label.....	32
8.6.1	Metadata.....	32
8.6.2	Classes.....	32
8.7	AI_Labeling.....	34
8.7.1	Metadata.....	34
8.7.2	Classes.....	34
8.8	AI_TDChangeset.....	35
8.8.1	Metadata.....	35
8.8.2	Classes.....	35
8.9	AI_DataQuality.....	36
8.9.1	Metadata.....	36
8.9.2	Classes.....	36
	Annex A (normative) Abstract test suite.....	37
	Annex B (informative) Examples.....	44
	Bibliography.....	47

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

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 ISO 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).

ISO draws attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO takes 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 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. ISO 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.

This document was prepared by Technical Committee ISO/TC 211, *Geographic information/Geomatics*, in collaboration with the European Committee for Standardization (CEN) Technical Committee CEN/TC 287, *Geographic Information*, in accordance with the Agreement on technical cooperation between ISO and CEN (Vienna Agreement) and in collaboration with the Open Geospatial Consortium (OGC).

A list of all parts in the ISO 19178 series can be found on the ISO website.

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

This document aims to develop the UML model and encodings for geospatial machine learning training data. Training data play a fundamental role in Earth Observation (EO) Artificial Intelligence Machine Learning (AI/ML), especially Deep Learning (DL). Training data are used to train, validate and test AI/ML models. This document defines a UML model and encodings consistent with the OGC Standards baseline to exchange and retrieve training data in the Web environment.

This document provides detailed metadata for formalizing the information model of training data. This includes, but is not limited to the following aspects:

- how the training data are prepared, such as provenance or quality;
- how to specify different metadata used for different ML tasks, such as scene/object/pixel levels;
- how to differentiate the high-level training data information model and extended information models specific to various ML applications;
- how to introduce external classification schemes and flexible means for representing labelling.

Geographic information — Training data markup language for artificial intelligence —

Part 1: Conceptual model

1 Scope

Within the context of training data for Earth Observation (EO) Artificial Intelligence Machine Learning (AI/ML), this document specifies a conceptual model that:

- establishes a UML model with a target of maximizing the interoperability and usability of EO imagery training data;
- specifies different AI/ML tasks and labels in EO in terms of supervised learning, including scene level, object level and pixel level tasks;
- describes the permanent identifier, version, licence, training data size, measurement or imagery used for annotation;
- specifies a description of quality (e.g. training data errors, training data representativeness, quality measures) and provenance (e.g. agents who perform the labelling, labelling procedure).

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 19101-1, *Geographic information — Reference model — Part 1: Fundamentals*

ISO 19103, *Geographic information — Conceptual schema language*

ISO 19115-1, *Geographic information — Metadata — Part 1: Fundamentals*

ISO 19156, *Geographic information — Observations, measurements and samples*

ISO 19157-1, *Geographic information — Data quality — Part 1: General requirements*

3 Terms, definitions and abbreviated terms

3.1 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/>

3.1.1

3D model reconstruction

<earth observation> task in which 3D objects and scenes are built from multi-view images

3.1.2

artificial intelligence

AI

branch of computer science devoted to developing data processing systems that perform functions normally associated with human intelligence, such as reasoning, learning and self-improvement

[SOURCE: ISO/IEC 2382:2015, 2121393, modified — Notes 1 and 2 to entry have been removed.]

3.1.3

change detection

<earth observation> recognition of changes between images acquired at different times

3.1.4

class

<classification> result of a classification process as part of a classification system which subdivides concepts within a given topic area

[SOURCE: ISO 19144-2:2023, 3.1.6]

3.1.5

dataset

identifiable collection of data

Note 1 to entry: A dataset can be a smaller grouping of data which, though limited by some constraint such as spatial extent or feature type, is located physically within a larger dataset. Theoretically, a dataset can be as small as a single feature or feature attribute contained within a larger dataset. A hardcopy map or chart can be considered a dataset.

[SOURCE: ISO 19115-1:2014, 4.3]

3.1.6

deep learning

DL

<artificial intelligence> approach to creating rich hierarchical representations through the training of neural networks with one or more hidden layers

Note 1 to entry: Deep learning uses multi-layered networks of simple computing units (or “neurons”). In these neural networks each unit combines a set of input values to produce an output value, which in turn is passed on to other neurons downstream.

[SOURCE: ISO/IEC TR 29119-11:2020, 3.1.26]

3.1.7

generative model

<artificial intelligence> method of large model training, which improves model performance through unsupervised pre-training

Note 1 to entry: In the fine-tuning phase, labelled data play a critical role in optimizing the model for specific vertical domains or tasks. By incorporating labelled data, the model can learn to accurately identify and extract relevant features, leading to better performance on specific downstream tasks. Overall, the combination of generative models and fine-tuning with labelled data can significantly improve the performance of large models in specialized domains or tasks.

3.1.8

label

<earth observation> known or expected results annotated as values of a dependent variable in training samples

Note 1 to entry: A training sample label is different from those on a geographical map, which are known as map labels or annotations.

3.1.9

machine learning

ML

<artificial intelligence> process of optimizing model parameters through computational techniques, such that the model's behaviour reflects the data or experience

Note 1 to entry: ML processes create models from training data by using a set of learning algorithms, and then can use these models to make predictions. Depending on whether the training data include labels, the learning algorithms can be divided into supervised and unsupervised learning.

[SOURCE: ISO/IEC 22989:2022, 3.3.5, modified — Note 1 has been added.]

3.1.10

object detection

<earth observation> recognition of objects from images

Note 1 to entry: The objects are often localized using bounding boxes.

3.1.11

provenance

organization or individual that created, accumulated, maintained and used records

Note 1 to entry: In this document provenance is a record of how training data were prepared.

[SOURCE: ISO 19115-1:2014, 4.16, modified — Note 1 to entry has been added.]

3.1.12

quality

degree to which a set of inherent characteristics of an object fulfils requirements

Note 1 to entry: Quality of training data (such as data imbalance and mislabelling) can impact the performance of artificial intelligence/machine learning (AI/ML) models.

[SOURCE: ISO 9000:2015, 3.6.2, modified — Notes 1 and 2 to entry have been removed, and a new Note 1 to entry has been added.]

3.1.13

scene classification

<earth observation> task of identifying scene categories of images, on the basis of a training set of images whose scene categories are known

3.1.14

semantic segmentation

<earth observation> task of assigning class labels to pixels of images or points of point clouds

3.1.15

training dataset

<earth observation> collection of samples, often labelled with known or expected values for supervised learning

Note 1 to entry: A training dataset can be divided into training, validation and test sets. "Training samples" referred to in this document are different from "samples" referred to in ISO 19156. They are often collected in purposive ways that deviate from purely probability sampling, with known or expected results labelled as values of a dependent variable for generating a trained predictive model.

3.2 Abbreviated terms

In this document, the following abbreviated terms and acronyms are used or introduced:

ATS	abstract test suite
DML	Data Markup Language
EO	earth observation
ISO	International Organization for Standardization
JSON	JavaScript Object Notation
LC	land cover
LU	land use
OGC	Open Geospatial Consortium
RS	remote sensing
SAR	synthetic aperture radar
TD	training data
UML	Unified Modelling Language
URL	Uniform Resource Locator
URI	Uniform Resource Identifier
XML	Extensible Markup Language

4 Conventions

4.1 General

This clause provides details and examples for any conventions used in the document. Examples of conventions are symbols, abbreviations, use of XML schema, or special notes regarding how to read the document.

4.2 Identifiers

The requirements in this specification are denoted by the URI:

<http://www.opengis.net/spec/TrainingDML-AI-1/1.0>

All requirements and conformance tests that appear in this document are denoted by partial URIs which are relative to this base.

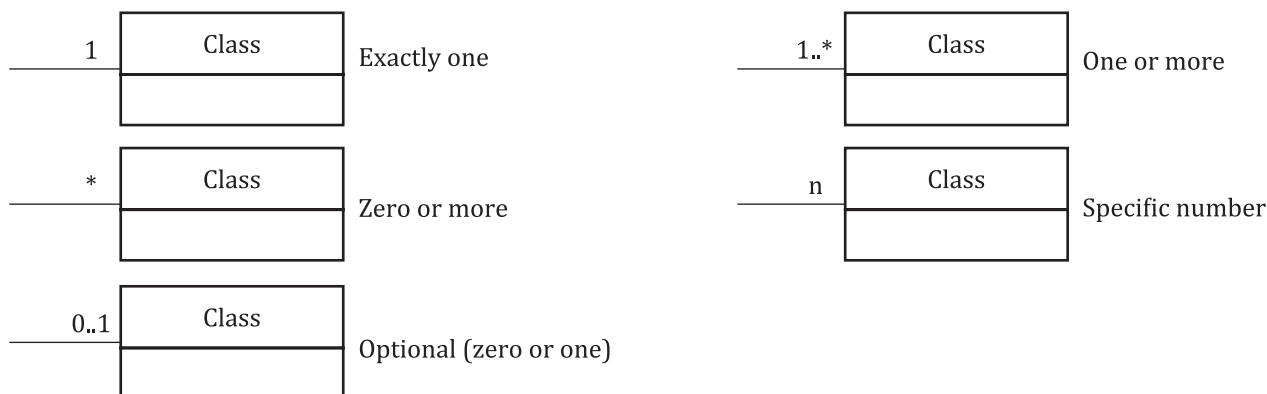
4.3 UML notation

The conceptual model is presented in this document through diagrams using the Unified Modelling Language (UML) static structure diagram. The UML notations used in this document are described in the diagram in [Figure 1](#).

Association between classes



Association cardinality



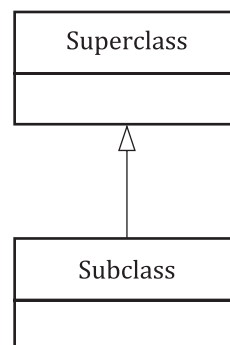
Aggregation between classes



Composition between classes



Class inheritance



NOTE For further information on the UML notation, see ISO 19103.

Figure 1 — UML notation

All associations between model elements in the TrainingDML-AI conceptual model are uni-directional. Thus, associations in the model are navigable in only one direction. The direction of navigation is depicted by an arrowhead. In general, the context an element takes within the association is indicated by its role. The role is displayed near the target of the association. But, if the graphical representation is ambiguous, the position of the role has to be drawn to the element to which the association points.

The following stereotypes are used in this model.

- «DataType» defines a set of properties that lack identity. A data type is a classifier with no operations, whose primary purpose is to hold information.
- «CodeList» enumerates the valid attribute values. In contrast to Enumeration, the list of values is open and, thus, not given inline in the TrainingDML-AI UML Model. The allowed values can be provided within an external code list.

5 Conformance

This document defines a conceptual model that is independent of any encoding or formatting technologies. The standardization target for this document is:

— TrainingDML-AI conceptual model

Conformance with this document shall be checked using all the relevant tests specified in [Annex A](#) of this document. The framework, concepts and methodology for testing, and the criteria to be achieved to claim conformance are specified in the OGC Compliance Testing Policies and Procedures and the OGC Compliance Testing web site.^[9]

All requirements-classes and conformance-classes described in this document are owned by the standard identified.

6 Overview

6.1 General

This document defines how to represent and exchange ML training data. The conceptual model includes the most relevant training data entities from datasets, to instances (i.e. individual training samples), to labels. The conceptual schema specifies how and into which parts the training data should be decomposed and classified.

This document strategically addresses geospatial requirements by providing a modular and extensible framework tailored to EO applications. The content and format of training datasets differ depending on the EO ML scenarios they were collected for (e.g. scene/object/pixel levels). This document defines a UML model and encodings consistent with the OGC/ISO baseline standards to exchange and retrieve geospatial training data. Existing geospatial standards (e.g. ISO 19101-1, ISO 19115-1, ISO 19157-1) can be reused when defining geospatial requirements on source RS images, label geometry, metadata and quality. While some general geospatial information such as the spatial extent and reference system information are defined for training data at the high level, other EO-specific information, such as the size of each sample image, spatial resolution, and bands, can be extended in a subclass at the low level. With a hierarchical and extensible structure, the training data model accommodates diverse geospatial data characteristics, ensuring flexibility and interoperability.

The training data model defined in this document facilitates interoperability by enabling heterogeneous training datasets to conform to a unified representation and exchange form. It ensures that training data from different vendors can be consistently shared and interpreted, improving the accessibility and promoting the integration of geospatial AI/ML resources.

The TrainingDML-AI conceptual model ([Clause 7](#)) is formally specified using UML class diagrams, complemented by a data dictionary ([Clause 8](#)) providing the definitions and explanations of the object classes and attributes. This conceptual model provides the basis for specifying encoding implemented in languages such as JSON, or XML. [Annex B](#) provides a series of encoding examples, including representations for TrainingDataset, DataQuality, and TDChangeset encoding.

6.2 AI tasks for EO

In recent years AI/ML has been increasingly used in the EO domain. The new AI/ML algorithms frequently require large training datasets as benchmarks. AI/ML TD have been used in many EO applications to calibrate the performance of AI/ML models. Many efforts have been made to produce training datasets to make accurate predictions. As a result, a number of training datasets are publicly available, with new datasets being constantly released. In the EO domain, examples of AI/ML training datasets have been developed in various tasks including the following typical scenarios.

— Scene classification: These algorithms determine image categories from numerous pictures (e.g. agricultural, forest and beach scenes). The training samples are a series of labelled pictures. The data

can be either from satellite, drones or aircrafts. The metadata of the datasets often includes the number of training samples, the number of classes and the image size.

- Object detection: These algorithms detect and localize different objects (e.g. airplanes, cars and buildings) in a single image. The image can be optical or non-optical, such as synthetic aperture radar (SAR). Recent work also suggests an increasing focus on object detection from street view imagery. Objects can be labelled using either polygons or bounding boxes. The bounding boxes can be either oriented vertically or horizontally. The geometry of a bounding box can be expressed using top-left/bottom-right coordinates, coordinates of four corners, or centre coordinates along with the length and width of the box.
- Semantic segmentation: In terms of land cover (LC) and land use (LU) classification, this process assigns a LC/LU class label to a pixel (or groups of pixels) of RS imagery. In the context of semantic segmentation of 3D point clouds, it classifies points of a 3D point cloud into categories. TDs are usually composed of RS images/point clouds, and the corresponding labelled value of each pixel/point recording its class.
- Change detection: These algorithms identify the difference between images acquired over the same geographical area but taken at different times. The TD comprise a set of pre-change and post-change RS images, with the corresponding reference map labelled for changed and unchanged pixels. The image can be optical or SAR images.
- 3D model reconstruction: These algorithms infer the 3D geometry and structure of objects and scenes, mainly realized from the dense matching of multi-view images. The TD are usually composed of two-view or multi-view images, with the corresponding disparity map or depth maps as reference data respectively.

6.3 Modularization

The TrainingDML-AI conceptual model provides models for the most important elements within TD. These elements have been identified to be either required or important in many different AI/ML tasks. However, implementations are not required to support the complete TrainingDML-AI model in order to be conformant to this document. Implementations may employ a subset of constructs according to their specific information needs. For this purpose, modularization is applied to the TrainingDML-AI model.

As shown in [Figure 2](#), the TrainingDML-AI conceptual model is thematically decomposed into a Basic module, a Provenance module, a Quality module and a Changeset module. The Basic module comprises the basic concepts and elements, including AI_TrainingDataset, AI_TrainingData, AI_Label and AI_Task, of the TrainingDML-AI, and thus, shall be implemented by any conformant system. The Provenance module provides a comprehensive definition of provenance by AI_Labeling, AI_Labeler and AI_Labeling Procedure. The Quality module offers quality description of TD with AI_DataQuality elements. The Changeset module defines AI_TDChangeset between versions of datasets.

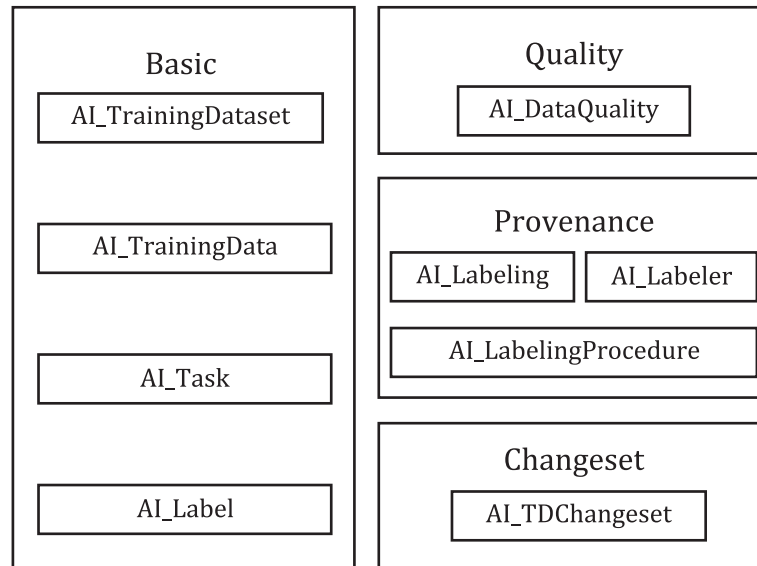


Figure 2 — TrainingDML-AI module overview

6.4 General modelling principles

6.4.1 Element modelling

The modelling of all elements in the TrainingDML-AI conceptual model has the following principles^[15].

- Granularity: Two levels of granularity are differentiated in the conceptual model: The Training Dataset is used to refer to the collection level, and the Training Data are used to refer to the individual level.
- Label semantics: The training dataset will not be limited to one classification scheme. External classification schemes should be allowed to be linked into the Training Dataset to accommodate different cases in practice. The development of external classification schemes can follow the ISO 19144 series.
- Light-weight design: The lightweight designed conceptual model has a minimum set of metadata elements, provenance, or quality measures at the collection level instead of at the individual level. This is to facilitate the understanding of the dataset and improve the scalability for communicating large training datasets.
- Alignment: The modelling of elements in TDs can leverage existing efforts for wide adoption, such as for ISO 19109, ISO 19115-1, ISO 19157-1, and the OGC Geography Markup Language (GML) Standard. The conceptual model can be aligned with these existing standards and leverage capabilities fulfilled in part by other standards.
- Quality, bias, and ethics: Elements related to quality, or more specifically, bias that can be used to reduce the errors when using AI/ML. For example, any knowledge of the TD imbalance and mislabelling can be stored in TD quality. In addition, data ethics aims to safeguard the responsible use of TD, and it can be addressed by using the licence property in the TD.
- Changeset: This will be an optional module in TD modelling. Changeset addresses how to capture changes in TD datasets. The change model considers the trend in TD collections to use the crowdsourcing platforms and borrow the change representation from the platforms such as OpenStreetMap.

6.4.2 Class hierarchy and inheritance of properties and relations

In the TrainingDML-AI conceptual model, the specific elements such as EO training datasets, EO training data, scene label, object label and pixel label are defined as subclasses of more general higher-level classes. Hence, elements build a hierarchy along specialization/generalization relations where more specialized

elements inherit the properties and relations of all their super classes along the entire generalization path to the topmost element.

6.4.3 Definition of the semantics for all classes, properties and relations

The meanings of all elements defined in the TrainingDML-AI conceptual model are normatively specified in the data dictionary in [Clause 8](#).

6.4.4 Data integrity, authenticity and non-repudiation

Sometimes training datasets can be downloaded, disseminated, and changed by anyone. The data integrity, authenticity and non-repudiation are important to avoid unexpected bias propagation and distorted results. Currently the standard focuses on the information modelling, while data dissemination can be enriched with strategies from the general information domain by publishing hashes (e.g. MD5) and public-keys (e.g. RSA) after signing and encrypting.

6.5 Extending TrainingDML-AI

The TrainingDML-AI conceptual model is designed as a universal information model that defines elements and attributes which are useful for a broad range of AI/ML applications in the geospatial domain. In practical AI/ML applications, the elements within specific TDs will most likely contain attributes which are not explicitly modelled in TrainingDML-AI. Moreover, there can be TD elements which are not covered by the TrainingDML-AI thematic classes.

The model provides an abstract class-based method to support the exchange of such data. Elements not represented by the predefined thematic classes of the model may be modelled and exchanged by extending abstract class.

7 TrainingDML-AI UML model

7.1 General

The TrainingDML-AI UML model is a series of UML classes designed based on the TrainingDML-AI conceptual model. The tables and figures in this clause are software generated from the UML model. As such, this clause provides a concrete representation of the TrainingDML-AI conceptual model.

7.2 ISO dependencies

TrainingDML-AI builds on the related documents, ISO 19101-1, ISO 19103, ISO 19115-1, ISO 19156 and ISO 19157-1, as identified in [Figure 3](#). Data dictionaries are included for all the ISO-defined classes explicitly referenced in the TrainingDML-AI UML model. These data dictionaries are provided for the convenience of the user. The ISO documents are the source documents.

ISO 19178-1:2025(en)

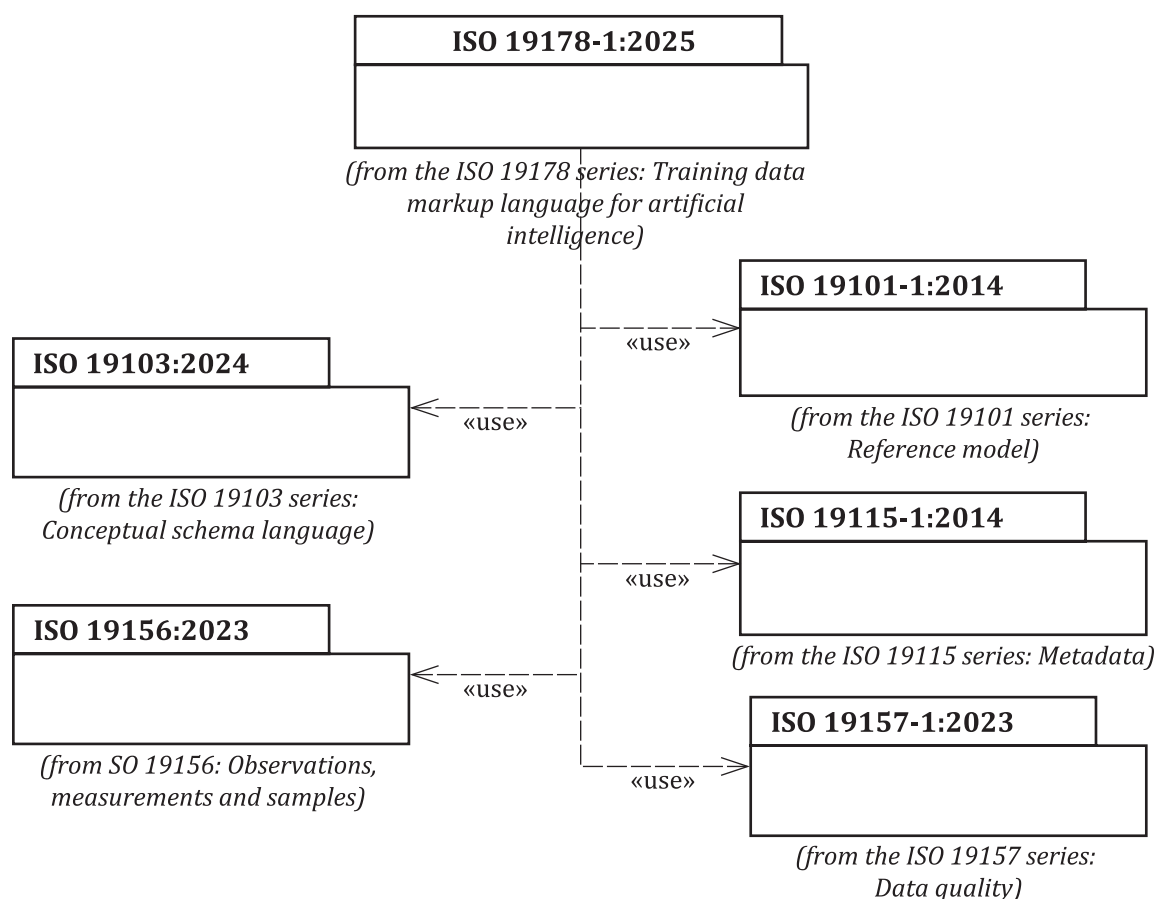


Figure 3 — Use of ISO documents in TrainingDML-AI

The ISO classes explicitly used in the TrainingDML-AI UML model are introduced in [Table 1](#). Further details about these classes can be found in the data dictionary within [Clause 8](#).

Table 1 — ISO classes used in TrainingDML-AI

Class name	Description
Feature	Abstraction of real-world phenomena.
MD_Band	Range of wavelengths in the electromagnetic spectrum.
MD_Scope	The target resource and physical extent for which information is reported.
MD_ReferenceSystem	Information about the reference system.
LI_Lineage	Information about the events or source data used in constructing the data specified by the scope or lack of knowledge about lineage.
EX_Extent	Extent of the resource.
CI_Citation	Standardized resource reference.
MD_Resolution	Level of detail expressed as a scale factor, a distance or an angle.
DataQuality	Quality information for the data specified by a data quality scope.
QualityElement	Aspect of quantitative quality information.

7.3 Overview of the UML model

The UML model is presented with core concepts in [Figure 4](#), followed by the concrete classes in [Figure 5](#). The following describes the core concepts.

- **AI_TrainingDataset**: This concept represents a collection of training samples, i.e. a training dataset.