International Standard

ISO/IEC 5181

Information security, cybersecurity and privacy protection — Data provenance

Sécurité de l'information, cybersécurité et protection de la vie privée — Provenance des données

First edition
2026-10

Reference number
ISO/IEC 5181:2026(en)

© ISO/IEC 2026



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2026

All rights reserved.

This ISO publication is protected by copyright and is owned by ISO and/or its licensors.

The content of this ISO publication is provided under licence, not sold. Use is subject to the applicable licence terms issued by ISO, an ISO member body, or an authorized third-party distributor.

Except as required for implementation or expressly permitted by a separate licence, no part of this ISO publication may be reproduced, distributed, modified, used, or made available in any form or by any means – electronic or mechanical, including photocopying, scanning, recording, or posting on internal or external digital platforms.

Any use beyond the scope of the granted rights is prohibited and may result in legal action.

ISO copyright office
CP 401 • CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword ... **iv**
Introduction ... **v**
1 Scope ... **1**
2 Normative references ... **1**
3 Terms and definitions ... **1**
4 Abbreviated terms ... **5**
5 Data provenance methodology ... **6**
5.1 General assumptions ... 6
5.2 Verification and validation principles ... 7
5.3 Value stream semantics ... 8
5.3.1 Correspondence between LC value streams ... 8
5.3.2 Value stream behaviour ... 8
5.3.3 Representation of VS semantics ... 9
5.3.4 Continuity aspect of operating products and data ... 10
5.3.5 Formal language terms ... 10
5.3.6 Processing of trust ... 10
5.3.7 Quality of data provenance ... 11
5.4 Value stream description language ... 11
5.4.1 Using variables ... 11
5.4.2 Data provenance proposition ... 11
5.4.3 Catalogue of narrative of provenance artefacts ... 12
5.4.4 Generating a narrative of provenance by example ... 13
6 Data provenance information model ... **14**
6.1 General ... 14
6.2 Data flow control model ... 14
6.3 Data provenance complexity model ... 14
6.4 Data provenance reference architecture model ... 15
6.5 Data provenance security and privacy model ... 16
Annex A (informative) **DP artefacts of narrative of provenance** ... **18**
Annex B (informative) **Operational ontology of DP** ... **23**
Annex C (informative) **Dataspaces for data provenance** ... **25**
Annex D (informative) **Data provenance constraints derived from regulations** ... **27**
Bibliography ... **28**

© ISO/IEC 2026 – All rights reserved

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 27, *Information security, cybersecurity and privacy protection*.

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.

© ISO/IEC 2026 – All rights reserved

Introduction

This document provides guidelines and methods on how to derive data provenance information from a system of co-operating authenticated stakeholders. Data provenance is valid if, the data value stream represented by sequences of < data, metadata > tuples, in any life cycle phase of compound industrial data and production value streams, is always comparable to the value stream of production involving operational technology. Hence, the < data, metadata > tuples value stream is twinning the production value stream such that both data and production value streams are isomorphic to each other.

The use and processing of data performed by applications, systems, robots, individuals, agents or companies usually is subject of permissions from data owners as applied in ISO/IEC 62890.[23]

“Data owner” is a term that is applied to the rights of a data holder. Data holders have the rights to access, use, or manipulate their data. The data owner role is understood as an individual or a group of stakeholders who can have control over certain data attributes like data creation, accuracy or the names or identifiers of sources, etc.

Data provenance information is useful to all who want to use available data in other applications or want to decide to make data voluntarily available as a common good in a dataspace that can be subscribed by various application stakeholders. Identifying and deriving information from the history of metadata events provide details on the narratives of data provenance and data manipulations. The concept of narration is used because it visually describes sequences of activities or the state changes of a system. Graphically, a sequence of state changes or events is a graph trajectory.

© ISO/IEC 2026 – All rights reserved

Information security, cybersecurity and privacy protection — Data provenance

1 Scope

This document provides guidelines, methodologies and techniques for securely deriving information about data assets from multiple sources, intermediaries or users, which is denoted as provenance metadata.

2 Normative references

There are no normative references in this document.

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/

3.1
data
data asset
collection of data items that provide value to organizations playing roles like data provider, data intermediary and data user

Note 1 to entry: Data are a product, hence data are an asset that can be disseminated, shared, distributed, consumed, or copied by processes or humans involved in the workflow of an enterprise, factory, information and data infrastructure or human-machine cooperating communities like smart cities, smart factories etc.

3.2
data altruism
voluntary sharing of data without compensation

Note 1 to entry: Voluntary sharing of data is based on consent of certain data subjects to process their personal data and to get permission from data holders to use non-personal data.

Note 2 to entry: Data can be shared via a virtual data sharing platform,[11] such as a dataspace.

Note 3 to entry: A certification framework can provide organizations with voluntary authorization by governments compared to a framework with compulsory authorization.

Note 4 to entry: In this context, data altruism is interpreted according to Article 2(16) of the EU DGA.[26]

3.3
artefact
constructive, descriptive or semantic element of views on a thing of interest (ToI) as represented by the semiotic triangle with three views on the ToI

Note 1 to entry: The construction of things view obeys operational engineering constraints, the description view obeys requirements represented by an ontological language, and the semantics view is given in a notation that supports reasoning analysis of the past and predictive simulation-based analysis of future behaviour of a ToI.

© ISO/IEC 2026 – All rights reserved

Note 2 to entry: The ToI can be modelled by an urban digital twin (UDT) which feeds back information from the performed analysis.

Note 3 to entry: The term artefact is preferred in formal systems because it means an intangible concept in contrast to a physical object.

3.4
computation
graph computation
sequence of directed edges of a graph (trajectory) that is enabled by annotated rules conditioning the execution of sequences of edges of the graph

Note 1 to entry: A data graph is an asset that is applied and manipulated along an executable life cycle value stream (LC VS) trajectory.

Note 2 to entry: A process graph is defined by three things i.e. a set of vertices (i.e. process states), a vertice-disjunct set of edges (i.e. transitions) and an incidence mapping that maps an edge either into an undirected pair of vertices (i.e. action), or that maps an edge into a pair of directed vertices (i.e. event). For short, the latter is called directed edge and the former undirected edge.

Note 3 to entry: Data or process assets both, are modelled by computable graphs, i.e. data graphs or process graphs.

3.5
dataspace
data governance framework and supporting services to build trustworthiness and enable data sharing through an agreed set of policies, semantic models, protocols and processes

Note 1 to entry: More formally, a dataspace can be considered as a set of data assets annotated with metadata such as data provenance information, which is organized as a publish-subscribe dataspace that manages many-sorted data type items as < *n*-tuples > , whereby < *n*-tuples > are defined for $n \geq 2$ elements of respective sorts.[5]

Note 2 to entry: *sort* is a term from applied mathematics that represents a variable i.e. a set of terms of a considered type.

[SOURCE: ISO/IEC 20151-1:—[1)], 3.1, modified — definition has been modified to align with the context of this document; notes 1 and 2 to entry have been added.]

3.6
data graph
graphical representation of a data structure using graph artefacts such as vertices, edges and an incidence mapping

Note 1 to entry: In the data provenance methodology of subclause 5.2, components of a data structure are graphically represented by tree-like graphs i.e. data graphs.

Note 2 to entry: The graph semantic artefacts allow to represent either data structures or knowledge but also behaviour of processes or stakeholders that communicate via the dataspace.

Note 3 to entry: A graph is defined by three elements i.e. a set of vertices, a vertices-disjunct set of edges and an incidence mapping that maps an edge either into an undirected pair of vertices, or that maps an edge into a pair of directed vertices.

3.7
process graph
graphical representation of a process using graph artefacts such as vertices, edges and an incidence mapping

Note 1 to entry: The elements of a process are graphically represented by a mesh-network-like graph.

Note 2 to entry: The graph semantic artefacts allow to represent behaviour of processes or stakeholders that communicate via the dataspace, but also data structures or knowledge.

1) Under development. Stage at the time of publication: ISO/IEC FDIS 20151-1:2026.

© ISO/IEC 2026 – All rights reserved

Note 3 to entry: A graph is defined by three elements i.e. a set of vertices, a vertices-disjunct set of edges and an incidence mapping that maps an edge either into an undirected pair of vertices, or that maps an edge into a pair of directed vertices.

3.8
data governance
process focused on managing the quality, consistency, usability, security, and availability of data

Note 1 to entry: This process is closely linked to the notions of data ownership and stewardship.

[SOURCE: ISO/TR 14872:2025,[27] 3.3, modified — "information" has been replaced by "data".]

3.9
data leakage
drain of data that makes the operation phenomenon, including the phenomenon's data, untrustworthy

Note 1 to entry: The longer the life cycle value stream (LC VS) of operation, the higher the complexity of production. Higher complexity creates a higher risk of data leakage or other non-authorized manipulations.

3.10
metadata
data that defines and describes other data

Note 1 to entry: In the ISO/IEC 11179 series and ISO/IEC 19763 series of standards, the objects being described are instances of the subclasses of item, e.g. concepts, data elements, services.

[SOURCE: ISO/IEC 11179-3:2023,[19] 3.2.30, modified — note 1 to entry has been added from ISO/IEC 11179-3:2023/Amd 1:2026[20]]

3.11
provenance metadata
metadata about the provenance of data

Note 1 to entry: Metadata comprises in practice various categories of information about the state of data that is observable during the life cycle value stream (LC VS).

Note 2 to entry: Metadata derived during the LC VS of processing assets has the purpose to provide trust in the incorporation of data subscribed from a dataspace.

Note 3 to entry: Metadata annotated to data that are generated and derived during the operation of a physical machinery (e.g. energy distribution, automated production, self-driving cars) is called the narrative of provenance (NoP) of the associated data of interest.

Note 4 to entry: By inspecting the NoP, the data of interest can be applied by stakeholders, processes, or individuals in contexts different from the source contexts. Based on the NoP, interested parties can decide which of their data subscribed from the dataspace, they want to use as a common good and which not.

3.12
data provenance
ownership and usage history of data of a system, consisting of information that is given by sets of metadata describing origins, sources of data, parties involved in generating, manipulating, and managing any data

Note 1 to entry: A similar definition of data provenance is given in VDE SPEC 90009 v1.0[10] that defines data provenance (DP) as the origin of the metadata values comprising person, organization, software agent, file or unknown. VDE SPEC 90009 v1.0[10] is based on IEC 61360-1[13] that defines the attributes for the specification of meta-metadata.[10]

Note 2 to entry: In this document and in contrast to the VDE SPEC 90009 v1.0[10], metadata (values) of data provenance information are interpreted semantically which means metadata (values) are derived during the life cycle value streams in the same activity or phase of production from the data and from the production.

Note 3 to entry: A similar definition of data provenance is given in ISO/IEC 11179-33[21] that defines provenance as information on place and time of origin, derivation or generation of a data set, proof of authenticity of the data set, or a data set of past and present ownership of that data set.

© ISO/IEC 2026 – All rights reserved

3.13
provability
verifiability
degree to which something can be reliably demonstrated or formally proven

Note 1 to entry: Defining the data provenance involves specifying provenance evaluation criteria and requirements that can demonstrate the data provenance of a system in a provable way.

Note 2 to entry: Provability is an important element of the data provenance methodology.

3.14
quality of data
degree to which the characteristics of data satisfy stated and implied needs when used under specified conditions

Note 1 to entry: Quality of data (QoD), e.g. of a blueprint of car production, is coupled with the quality of the operation phenomenon of a technical production of a car.

Note 2 to entry: Data are an informational asset that digitally mirrors the behaviour of an operational asset. The coupling is understood as the technical phenomena of a system's machinery, e.g. car production, that is tightly coupled with the mirroring asset of data capturing through sensors and data analysis provisioning to actuators on a high level of quality of data.

[SOURCE: ISO/IEC 25012:2008, 4.3, modified — preferred term has been made "quality of data" (QoD); Notes to entry have been added.]

3.15
security
resistance to intentional unauthorized acts designed to cause harm or damage to a digital data asset or to personally identifiable information captured by a system

Note 1 to entry: Security is the freedom from misuse of data.

Note 2 to entry: In the context of data provenance, "system" includes data and their associated metadata.

[SOURCE: ISO/IEC 23643:2020, 3.16, modified — in the definition, "system" has been replaced by "digital data asset or to personally identifiable information captured by a system"; Notes to entry have been added.]

3.16
privacy
right of individuals to control or influence information related to them that may be collected and stored and by whom and to whom that information may be disclosed

Note 1 to entry: In the context of data provenance, privacy refers to control of the data and the associated metadata.

[SOURCE: ISO 7498-2:1989, 3.3.43,[28] modified — Note to entry has been replaced.]

3.17
semantics
mapping between elements of a language and the real world

Note 1 to entry: The semantics of data manipulation can be described in computational terms such as processes, variables and many-sorted data types.

Note 2 to entry: The semantics of production process based on state changes during life cycle value streaming can be represented by graphs comprising events and states.

[SOURCE: ISO/IEC TR 24800-1:2012,[29] 2.5, modified — Notes to entry have been added.]

3.18
semiotic triangle
model that relates linguistic symbols to semantic concepts and to the objects they represent

Note 1 to entry: Linguistic symbols can be terms or signs, possibly from an ontology.

© ISO/IEC 2026 – All rights reserved

Note 2 to entry: Semantic concepts can be represented by graph vertices and edges.

3.19
sort
data type
variable
formal notion from algebraic data type theory, which describes sub-data types that can be extended to many-sorted (i.e. complex) data types comprising multiple sets of data

Note 1 to entry: Sorts are algebraic terms that represent sets of data in a dataspace.

Note 2 to entry: The terms data type and variable are synonymous terms for a sort i.e. a set of data items.

3.20
trust
recognize an asset (something, someone) as being trustworthy in a defined context of use that is of value to the user

3.21
trustworthiness
ability to meet stakeholders' expectations in a verifiable way

Note 1 to entry: Depending on the context or sector, and on the specific product or service, data, technology and process used, different characteristics apply and need verification to ensure stakeholders' expectations are met.

[SOURCE: ISO/IEC TS 5723:2022,[7] 3.1.1, modified — Notes 2, 3 and 4 to entry have been deleted.]

3.22
validity
proof of a data provenance proposition in certain circumstances or system states that become valid or invalid

Note 1 to entry: Semiotic mappings is achieved by defining pairs of semiotic artefacts from the three semiotic domains. By this view, validity is a kind of "weak correctness proof" based on propositions about relationships between semantic artefacts (declarative), ontological artefacts (descriptive) and physical artefacts (constructive, phenomenological).

4 Abbreviated terms

AAS	asset administration shell
CDDL	concise data definition language
DP	data provenance
DPIM	data provenance information model
DS	data secrecy
DTw	digital twin
HMI	man-machine interface
GDPR	general data protection regulation
IoT	internet of things
MD	metadata
MMI	machine-machine interface
ML	machine learning

© ISO/IEC 2026 – All rights reserved