



DPC: 25/30469641 DC

BSI Group Headquarters

389 Chiswick High Road London W4 4AL

Tel: +44 (0)20 8996 9000

Fax: +44 (0)20 8996 7400

www.bsigroup.com

Date: 04 December 2025

Origin: International

Latest date for receipt of comments: 27 January 2026

Project No. 2023/00322

Responsible committee: IST/33/4 Security Controls and Services

Interested committees:

Title: Draft BS ISO/IEC 5181 Information technology — Security and privacy — Data provenance.

Please notify admin.start@bsigroup.com if you are aware of any keywords that might assist in clarifying or identifying the standard or of the content of this standard:

- i) has any issues related to 3rd party IPR, patent or copyright
- ii) affects other national standard(s)
- iii) requires additional national guidance or information

WARNING: THIS IS A DRAFT AND MUST NOT BE REGARDED OR USED AS A BRITISH STANDARD. THIS DRAFT IS NOT CURRENT BEYOND 27 January 2026

This draft is issued to allow comments from interested parties; all comments will be given consideration prior to publication. No acknowledgement will normally be sent. **See overleaf for information on the submission of comments.**

No copying is allowed, in any form, without prior written permission from BSI except as permitted under the Copyright, Designs and Patent Act 1988 or for circulation within a nominating organization for briefing purposes. Electronic circulation is limited to dissemination by e-mail within such an organization by committee members.

Further copies of this draft may be purchased from BSI Shop <http://shop.bsigroup.com> or from BSI Customer Services, Tel: +44(0) 20 8996 9001 or email cservices@bsigroup.com. British, International and foreign standards are also available from BSI Customer Services.

Information on the co-operating organizations represented on the committees referenced above may be obtained from <https://standardsdevelopment.bsigroup.com/>.

E-mail: admin.start@bsigroup.com

Introduction

This draft standard is based on national and international discussions. Your comments on this draft are invited and will assist in the preparation of the consequent standard.

For international standards, comments will be reviewed by the relevant UK national committee before sending the consensus UK vote and comments to the international committee, which will then decide appropriate action. If the international standard is approved, it is usual for the text to be published as a British Standard.

For national standards, comments will be reviewed by the relevant UK national committee and the resulting standards published as a British Standard.

UK Vote

Please indicate whether you consider the UK should submit a negative (with supporting technical reasons) or positive vote on this draft. Please indicate if you are aware of any reason why this draft standard should not be published as a British Standard.

Submission of Comments

- **Annotated drafts are not acceptable and will be rejected.**
- All comments should be submitted online at <https://standardsdevelopment.bsigroup.com/>. You will need to register in order to comment.

Template for comments and secretariat observations

Date: xx/xx/20xx	Document: ISO/DIS xxxx
------------------	------------------------

1	2	(3)	4	5	(6)	7
MB	Clause No./ Subclause No./Annex (e.g. 3.1)	Paragraph/ Figure/Table/Note	Type of comment	Comment (justification for change) by the MB	Proposed change by the MB	Secretariat observations on each comment submitted
	3.1	Definition 1	ed	Definition is ambiguous and needs clarifying.	Amend to read '...so that the mains connector to which no connection...'	
	6.4	Paragraph 2	te	The use of the UV photometer as an alternative cannot be supported as serious problems have been encountered in its use in the UK.	Delete reference to UV photometer.	



DRAFT International Standard

ISO/IEC DIS 5181

Information technology — Security and privacy — Data provenance

ICS: 35.030

ISO/IEC JTC 1/SC 27

Secretariat: **DIN**

Voting begins on:
2025-12-04

Voting terminates on:
2026-02-26

This document is circulated as received from the committee secretariat.

THIS DOCUMENT IS A DRAFT CIRCULATED FOR COMMENTS AND APPROVAL. IT IS THEREFORE SUBJECT TO CHANGE AND MAY NOT BE REFERRED TO AS AN INTERNATIONAL STANDARD UNTIL PUBLISHED AS SUCH.

IN ADDITION TO THEIR EVALUATION AS BEING ACCEPTABLE FOR INDUSTRIAL, TECHNOLOGICAL, COMMERCIAL AND USER PURPOSES, DRAFT INTERNATIONAL STANDARDS MAY ON OCCASION HAVE TO BE CONSIDERED IN THE LIGHT OF THEIR POTENTIAL TO BECOME STANDARDS TO WHICH REFERENCE MAY BE MADE IN NATIONAL REGULATIONS.

RECIPIENTS OF THIS DRAFT ARE INVITED TO SUBMIT, WITH THEIR COMMENTS, NOTIFICATION OF ANY RELEVANT PATENT RIGHTS OF WHICH THEY ARE AWARE AND TO PROVIDE SUPPORTING DOCUMENTATION.

Reference number
ISO/IEC DIS 5181:2025(en)

Full document details: 25/30469641 DC



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2025

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

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	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 Introduction	6
5.2 Data provenance methodology	8
5.3 Data provenance semantics	9
5.3.1 Life cycle value stream behaviour	9
5.3.2 Description of data provenance semantics	10
5.3.3 Validity of data provenance semantics	10
5.3.4 Quality of data provenance	11
5.4 Data provenance description language	11
5.4.1 Data provenance description	11
5.4.2 Data provenance proposition	12
5.5 Data provenance language artefact mappings (morphisms)	12
5.5.1 Catalogue of data provenance narrative artefacts	12
5.5.2 Data provenance narrative	13
5.5.3 Data provenance representation and computation	14
5.5.4 Data provenance data and metadata	14
6 Data provenance information model	15
6.1 General Requirements	15
6.2 Data flow control model	15
6.3 Data provenance reference architecture model	16
6.4 Data provenance security and privacy	18
Annex A (informative) DP narrative artefacts	20
Annex B (informative) An operational ontology of DP	24
Annex C (informative) Dataspaces for data provenance	27
Annex D (informative) Data provenance constraints derived from EU Acts	29
Bibliography	30

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 documents can 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 shall 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 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/IEC JTC1, Joint Technical Committee 1, Subcommittee SC 27/WG 4, Information security, cybersecurity and privacy protection - Security controls and services.

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 has been developed to define safety, security and privacy needs for a system adopting the communication capabilities of a dataspace¹⁾. Dataspaces can manage data assets annotated with metadata based on semantic principles.

This document provides guidelines and methods on how to derive authentic information from a system of co-operating authenticated stakeholders. Authenticity in industrial systems means that the *<data, metadata>-tuples* value stream (cf. data provenance (DP) information model of 6) in any phase of the data life cycle value stream (LC VS) is authentic to the value stream of the *production* life cycle (cf. [Figure 1](#)).

The purpose of this document is to provide data provenance guidelines in relation to data provenance related standards that ought to be applied with international treaty conventions and with GDPR²⁾. Hence, the document provides a common set of roles and rules providing authority for processing data assets by means of a dataspace.

The security and privacy framework addressed in this document is intended to help stakeholders in identifying their privacy safeguarding needs, related to personally identifiable information (PII)³⁾ within an enterprise or organizational environment:

- specifying a common privacy terminology;
- defining actors and their roles in processing PII;
- describing privacy safeguarding requirements;
- referencing known privacy policy principles.

The use and processing of data performed by applications, systems, robots, individuals, agents or companies usually is subject of permissions from data owners⁴⁾.^[22] The data owner role is understood as an individual or a group of stakeholders who can have control over certain data asset attributes like data creation, accuracy, data or, the names or identifiers of sources, etc. However, a data owner can give allowance for the use of 'owned' data within known application domains. Alternatively, authenticated data owners i.e. legal holders of data, can publish their data without expecting any remuneration to domains to which authenticated users can subscribe from the same or from other application domains.

Another kind of control in the use of data are the possible changes of global circumstances like a pandemic or malfunctioning of a critical infrastructure. For the purpose of performing adaptations or repairs effectively, data from trusted third parties are validated with information derived from data provenance histories.

Hence provenance *<data, metadata>* pairs provisioning can be a market enabler for dataspaces supporting data provenance (DP). DP supporting dataspaces are needed by smart applications such as smart manufacturing, smart city, smart grid, smart agriculture, robotics, logistics, mobility, health care etc.

Identifying and deriving information captured by metadata collections on the history of events telling the narratives of the provenance of data including their manipulations. Data provenance information is useful to individuals and companies that want to use available data assets in other applications or want to decide to make data voluntarily available as a common good in *dataspaces* ([3.4](#)) subscribed by various application stakeholders.

1) Here a dataspace is understood as a medium that interconnects the various stakeholders of a community (like smart city, I4.0, enterprises, energy networks etc.) that need to share data for their required co-operation. In the scope of this document it is referred to the definition of dataspace ([3.4](#))

2) General Data Protection Regulation (GDPR) ISO/IEC 27701^[2] or ISO/IEC 29100:2024^[3]

3) PII is defined in ISO 27701, ISO 27018 etc.

4) 'Data owner'(ISO/IEC 62890[22]) is a term that is captured by the EU data service act with respect to the rights of a data holder. Data holders have the rights to access, use, or manipulate their data.

From a data asset user's point of view there exist three information categories of personal data provenance that are to be handled, e.g.

- legal data (e.g. a person's name, birthday, personal identifiers) that can authenticate a person;
- PII unless there is private data that includes or excludes PII (e.g. assertions on the current geographic position, mobility profiles) that can be authorized for use by a user or customer of data acquired from a person;
- collective data (e.g. social media data, proximity warning application, encrypted data) that can be made publicly available and can be proven to be ethically compliant to adopted compliance rules.

From an industrial operational technology (OT) asset's point of view of a smart factory for example, there exist various categories of data assets related to human-machine mixed provenances:

- human-machine interfacing (HMI) data (e.g. machine control commands, recognition of light panels);
- machine-machine interfacing (MMI) data (e.g. machine-readable work piece descriptions, digital twin work piece processing models);
- machine-learning data (e.g. work piece data models to be learned and enriched by robots).

International and national organizations or institutions developing standards (SDO) or governmental acts work on projects that are related to data provenance e.g.^[8] on cards and security devices^[9], on data space concepts on distributed platforms^[12], on the interchange of provenance information. There are more contributions from other SDOs as well that are in relation to the issue of DP:

- Projects from the above-mentioned standardization committees address data provenance related challenges of using dataspace technology based on digital data chain technologies on distributed platforms [09, 11, 14, 16, 17], trustworthiness vocabulary,^[7] data credibility, multi-sourced data processing provenance^[10] of metadata or, interchange of provenance information^[12] etc.
- ISO/IEC 30141:04-2022^[1] describes the IoT reference architecture from the trustworthiness viewpoint regarding physical entity and digital entity worlds. The trustworthiness viewpoint considers system structures where trust matters in a normative fashion. A trust model^[16] comprising the trustworthiness viewpoint identifies operational aspects of systems under failure mode.
- ITU-T Y.3602^[17] describes recommendations on big data i.e. functional requirements for data provenance identifying factors for the determination of reliability of data and big data ecosystems.
- Regionally the EU has adopted the view on 'data sharing' by their regulatory acts i.e. 'Data Governance Act (DGA)' and 'Digital Services Act (DSA)' (see [Annex D](#)). The DGA promotes the *availability* of data in certain communities or industries by building trustworthy environments. In those environments data is shared free of charge to support research and the creation of new services and products.

In principle, DP metadata are the means of providing trustworthy information to data users.

This acquired information (DP metadata-based narratives on e.g. production data assets) allows data users to decide on intended applications comprising data assets. The information providing trust on using data, is to be validated from the history of metadata of the data of interest. For example, the metadata history can be inspected with respect to the following possible questions:

- What is the method applied for data classification?
- Which stakeholders have access to metadata labelling?
- Who has access to sensitive data?
- What is the method for so-called dark data or dark metadata checking? – In big data sets much data that exist cannot be used because the sources of dark data and metadata ⁵⁾ are not authenticated.

5) dark data and metadata are data that cannot be seen by data users. One of the reasons is that they do not exist in the (hi)stories of data provenance. Another reason lays in the policies of the data owner or data creator which put restrictions of use onto <data metadata> pairs or sets

- What are the basic and advanced security control techniques applied for dynamic dataspace distributions?
– In such environments are PII effectively be protected?
- How can users be protected against identification or classification errors?
- Can risks from distributed environments of the dataspace being assessed based on metadata by users?
- How is the actual quality of identity checks – including credentials, key management etc.?
- Can information on hidden redundancy of data representations sufficiently be derived from metadata descriptions?
- Is there information to identify faked data?
- Can data be identified that have been manipulated by malicious code or poisoning of data?
- Is there information about the LC phases allowing decommissioning of data and metadata tuples?

The sets of tuples <data, metadata> of a data asset history, captured by a dataspace are an asset that represents digitally e.g. the production process of a product, e. However, at the end of the LC VS, the collected tuples are to be decommissioned similar to the related product.

Information technology — Security and privacy — Data provenance

1 Scope

This document provides guidelines, methodology and techniques for deriving securely information from real activities denoted to as provenance metadata about data assets from multiple sources, intermediaries or users.

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 altruism

voluntary sharing of data⁶⁾ without compensation

Note 1 to entry: voluntary sharing of data⁷⁾ is based on consent of certain data subjects to process their personal data and to get permissions 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 voluntary certification framework can provide organizations with low intensity regulation by government compared to a framework with compulsory authorization.

3.2

artefact

constructive, descriptive or semantic element of views on a thing of interest (ToI) as represented by the *semiotic triangle* (3.17) with three views on the ToI

Note 1 to entry: Construction of things view obeys operational engineering constraints, 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.

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

Note 3 to entry: The term *artefact* (3.2) is preferred because in formal systems it means an intangible concept in contrast to an 'ancient artefact' meaning a physical object.

6) as defined in Article 2(16) of the EU DGA.

7) as defined in Article 2(16) of the EU DGA.

3.3

computation

graph computation

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

Note 1 to entry: A *data graph* (3.5) is an asset that is applied and manipulated along an executable LC VS trajectory.

Note 2 to entry: A *process graph* (3.6) is defined by three things i.e. a set of vertices (i.e. process states), a vertex-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 modeled by *computable graphs* (3.3), i.e. *data graphs* (3.5) or *process graphs* (3.6).

3.4

dataspace

data governance (3.7) framework and supporting services to build *trustworthiness* (3.19) 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 $\langle n\text{-tuples} \rangle$, whereby $\langle n\text{-tuples} \rangle$ are defined for $n \geq 2$ elements respective sorts

[SOURCE: ISO/IEC DIS 20151,3.1, modified – added “Note 1 to entry”]

3.5

data graph

graphical representation of a data structure using graph *artefacts* (3.2) such as vertices, edges and an incidence mapping

Note 1 to entry: The components of a data structure are graphically represented e.g. by a tree-like graph.

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.

3.6

process graph

graphical representation of a process using graph *artefacts* (3.2) 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.

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

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: TR 14872:2019, 3.2, modified – “information” replaced by “data” and Note 2 to entry deleted.]

3.8

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 of operation, the higher the complexity of production. Higher complexity creates a higher risk of data leakage or other non-authorized manipulations.

3.9

metadata

descriptive data about an object

Note 1 to entry: An object is anything perceivable or conceivable, including data, and data asset.

[SOURCE: ISO/IEC 11179-3:2023/DAmD1:2025(en), 3.2.30, modified – Note 1 to entry replaced]

3.10

provenance metadata

metadata (3.9) about the provenance of data assets

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.

Note 2 to entry: Metadata derived during the life cycle value streams (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 assets 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.

3.11

data provenance

ownership and usage history of data assets of a system

Note 1 to entry: Provenance of data is information that is given by sets of metadata describing origins, sources of data, parties involved in generating, manipulating, and managing any data.

Note 2 to entry: The national VDE SPEC 90009 v1.0^[10] defines 'data provenance' as the origin of the metadata values comprising 'person, organization, software agent, file or unknown'. The 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 3 to entry: In this document and in contrast to the VDE SPEC 90009 v1.0^[10] metadata (values) of data asset provenance information are interpreted semantically which means, metadata (values) are derived during the LC value streams in the same step or phase of production from the data asset and from the production asset.

Note 4 to entry: ISO/IEC 11179-33:2023^[20] 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'.

3.12

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 provenance of data assets of a system in a verifiable way.