
Data quality —

Part 117:

**Application of ISO 8000-115 to
identifiers in distributed ledgers
including blockchains**

STANDARDSISO.COM : Click to view the full PDF of ISO 8000-117:2023



STANDARDSISO.COM : Click to view the full PDF of ISO 8000-117:2023



COPYRIGHT PROTECTED DOCUMENT

© ISO 2023

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 Fundamental principles and assumptions.....	2
5 Requirements for on-ledger identifiers.....	2
6 Requirements for off-ledger data sets.....	3
Annex A (informative) Document identification.....	4
Annex B (informative) Example of generating an identifier for an off-ledger master data record.....	5
Bibliography.....	7

STANDARDSISO.COM : Click to view the full PDF of ISO 8000-117:2023

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 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 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/TC 184, *Automation systems and integration*, Subcommittee SC 4 *Industrial data*.

A list of all parts in the ISO 8000 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

0.1 Foundations of the ISO 8000 series

Digital data deliver value by enhancing all aspects of organizational performance including:

- operational effectiveness and efficiency;
- safety and security;
- reputation with customers and the wider public;
- compliance with statutory regulations;
- innovation;
- consumer costs, revenues and stock prices.

In addition, many organizations are now addressing these considerations with reference to the United Nations Sustainable Development Goals¹⁾.

The influence on performance originates from data being the formalized representation of information²⁾. This information enables organizations to make reliable decisions. This decision making can be performed by human beings directly and also by automated data processing including artificial intelligence systems.

Through widespread adoption of digital computing and associated communication technologies, organizations become dependent on digital data. This dependency amplifies the negative consequences of lack of quality in these data. These consequences are the decrease of organizational performance.

The biggest impact of digital data comes from two key factors:

- the data having a structure that reflects the nature of the subject matter;

EXAMPLE 1 A research scientist writes a report using a software application for word processing. This report includes a table that uses a clear, logical layout to show results from an experiment. These results indicate how material properties vary with temperature. The report is read by a designer, who uses the results to create a product that works in a range of different operating temperatures.

- the data being computer processable (machine readable) rather than just being for a person to read and understand.

EXAMPLE 2 A research scientist uses a database system to store the results of experiments on a material. This system controls the format of different values in the data set. The system generates an output file of digital data. This file is processed by a software application for engineering analysis. The application determines the optimum geometry when using the material to make a product.

ISO 9000 explains that quality is not an abstract concept of absolute perfection. Quality is actually the conformance of characteristics to requirements. This actuality means that any item of data can be of high quality for one purpose but not for a different purpose. The quality is different because the requirements are different between the two purposes.

EXAMPLE 3 Time data are processed by calendar applications and also by control systems for propulsion units on spacecraft. These data include start times for meetings in a calendar application and activation times in a control system. These start times require less precision than the activation times.

1) <https://sdgs.un.org/goals>

2) ISO 8000-2 defines information as “knowledge concerning objects, such as facts, events, things, processes, or ideas, including concepts, that within a certain context has a particular meaning”.

ISO 8000-117:2023(E)

The nature of digital data is fundamental to establishing requirements that are relevant to the specific decisions that are made by each organization.

EXAMPLE 4 ISO 8000-1 identifies that data have syntactic (format), semantic (meaning) and pragmatic (usefulness) characteristics.

To support the delivery of high-quality data, the ISO 8000 series addresses:

- data governance, data quality management and maturity assessment;

EXAMPLE 5 ISO 8000-61 specifies a process reference model for data quality management.

- creating and applying requirements for data and information;

EXAMPLE 6 ISO 8000-110 specifies how to exchange characteristic data that are master data.

- monitoring and measuring information and data quality;

EXAMPLE 7 ISO 8000-8 specifies approaches to measuring information and data quality.

- improving data and, consequently, information quality;

EXAMPLE 8 ISO/TS 8000-81 specifies an approach to data profiling, which identifies opportunities to improve data quality.

- issues that are specific to the type of content in a data set.

EXAMPLE 9 ISO/TS 8000-311 specifies how to address quality considerations for product shape data.

Data quality management covers all aspects of data processing, including creating, collecting, storing, maintaining, transferring, exploiting and presenting data to deliver information.

Effective data quality management is systemic and systematic, requiring an understanding of the root causes of data quality issues. This understanding is the basis for not just correcting existing nonconformities but for also implementing solutions that prevent future reoccurrence of those nonconformities.

EXAMPLE 10 If a data set includes dates in multiple formats including “yyyy-mm-dd”, “mm-dd-yy” and “dd-mm-yy”, then data cleansing can correct the consistency of the values. Such cleansing requires additional information, however, to resolve ambiguous entries (such as, “04-05-20”). The cleansing also cannot address any process issues and people issues, including training, that have caused the inconsistency.

0.2 Understanding more about the ISO 8000 series

ISO 8000-1 provides a detailed explanation of the structure and scope of the whole ISO 8000 series.

ISO 8000-2³⁾ specifies the single, common vocabulary for the ISO 8000 series. This vocabulary is ideal reading material by which to understand the overall subject matter of data quality. ISO 8000-2 presents the vocabulary structured by a series of topic areas (for example, terms relating to quality and terms relating to data and information).

ISO has identified ISO 8000-1, ISO 8000-2 and ISO 8000-8 as horizontal deliverables⁴⁾.

0.3 Role of this document

As a contribution to the overall capability of the ISO 8000 series, this document specifies requirements to address the data quality of identifiers in distributed ledgers including blockchains^[12].

Organizations exchange data to support transactions in supply chains. These data can include descriptions of: the parties to the transaction; the goods or services involved; logistics information

3) The content is available on the ISO Online Browsing Platform. <https://www.iso.org/obp>

4) Deliverable dealing with a subject relevant to a number of committees or sectors or of crucial importance to ensure coherence across standardization deliverables.

relevant to the transaction (e.g. ship to, ship from and transportation conditions). These descriptions typically consist of identifiers that are references to other data sets.

Distributed ledgers offer improvements over existing mechanisms to manage data about transactions in supply chains. These improvements include: greater efficiency of transactions; validation of the identity of the parties, goods and services involved; security of the data. The improvements are, however, not achievable unless the identifiers in the ledger are of high quality, providing a reliable link to the data sets that are off the ledger.

This document builds on the requirements of ISO 8000-115 to achieve the necessary quality of the identifiers in distributed ledgers.

Organizations can use this document on its own or in conjunction with other parts in the ISO 8000 series.

This document supports activities that affect:

- one or more information systems;
- data flows within the organization and with external organizations;
- any phase of the data life cycle.

[Annex A](#) of this document contains an identifier that conforms to ISO/IEC 8824-1. The identifier unambiguously identifies this document in an open information system.

0.4 Benefits of the ISO 8000 series

By implementing parts in the ISO 8000 series to improve organizational performance, an organization achieves the following benefits:

- objective validation of the foundations for digital transformation of the organization;
- a sustainable basis for data in digital form becoming a fundamental asset class the organization relies on to deliver value;
- securing evidence-based trust from other parties (including supply chain partners and regulators) about the repeatability and reliability of data and information processing in the organization;
- portability of data with resulting protection against loss of intellectual property and reusability across the organization and applications;
- effective and efficient interoperability between all parties in a supply chain to achieve traceability of data back to original sources;
- readiness to acquire or supply services where the other party expects to work with common understanding of explicit data requirements.

STANDARDSISO.COM : Click to view the full PDF of ISO 8000-117:2023

Data quality —

Part 117:

Application of ISO 8000-115 to identifiers in distributed ledgers including blockchains

1 Scope

This document specifies requirements for using identifiers in distributed ledgers including blockchains. These requirements supplement those of ISO 8000-115.

The following are within the scope of this document:

- requirements for each identifier in distributed ledgers including blockchains, where the ledger enables supply chains to exchange transaction data;
- requirements for the off-ledger data set that is referenced by an identifier in a distributed ledger;
- off-ledger data sets that consist only of property-value tuples, where each tuple is the result of semantic encoding.

The following are outside the scope of this document:

- requirements for implementing distributed ledger technology including blockchains;
- requirements for technology to ensure immutability of data sets;
- requirements for the processes of creating identifiers;
- requirements for the processes of creating the content of off-ledger data sets;
- requirements for the processes of creating immutable data sets.

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 8000-2, *Data quality — Part 2: Vocabulary*

ISO 8000-115, *Data quality — Part 115: Master data: Exchange of quality identifiers: Syntactic, semantic and resolution requirements*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 8000-2 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 Fundamental principles and assumptions

Information systems typically rely on identifiers to establish a consistent basis on which to access a corresponding data set.

By using ISO 8000-115, organizations generate identifiers that:

- capture the authoritative owner of the corresponding data set;
- are suitable for exchange with other parties.

Distributed ledger technologies, including blockchains, provide an effective means to establish immutable data sets, which cannot be changed by either accidental or deliberate actions. This technology, however, has the constraint of only being efficient when the internal data is small. This constraint leads to the internal data generally including identifiers that correspond to external data sets (i.e. off-ledger data sets), which can be much larger but are not immutable.

As summarised in [Figure 1](#), this document specifies the requirements necessary to ensure an immutable identifier (on-ledger identifier, see [Clause 5](#)) and establishes the immutability of an external data set (off-ledger data set, see [Clause 6](#)).

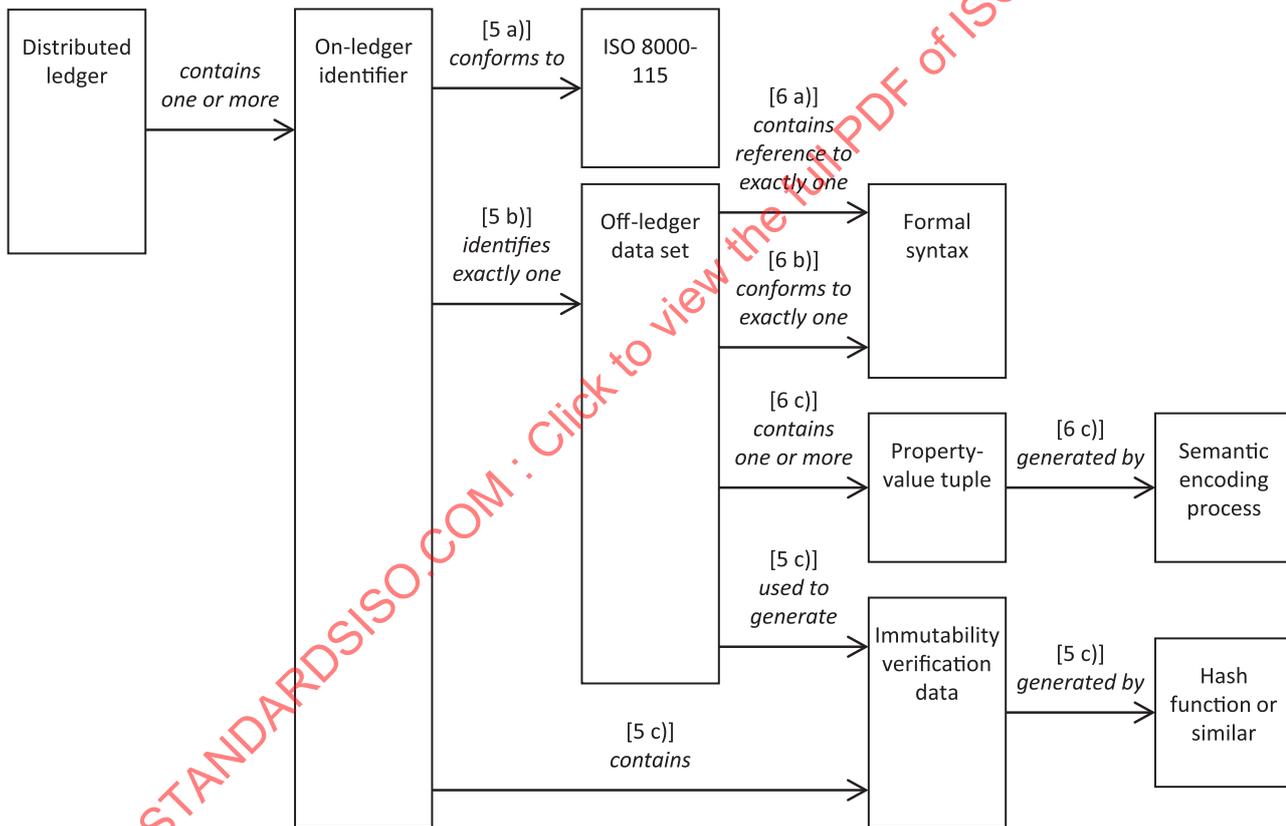


Figure 1 — Key elements of distributed ledgers and related off-ledger data sets

5 Requirements for on-ledger identifiers

- a) Within the distributed ledger, all on-ledger identifiers shall conform to ISO 8000-115.
- b) Each on-ledger identifier shall identify a single, off-ledger data set.
- c) Each on-ledger identifier shall include data that can be used to verify the off-ledger data set has not been changed since creation of the corresponding on-ledger identifier [(see [5 b\)](#)].

NOTE There are many different approaches to establishing the immutability of a data set^[13].

EXAMPLE See [Annex B](#) for an example using the CRC-32 algorithm specified by ISO/IEC 13239.

6 Requirements for off-ledger data sets

- a) The off-ledger data set shall reference a formal syntax.
- b) The off-ledger data set shall conform to the formal syntax [(see [6 a\)](#)].
- c) The off-ledger data set shall consist of property-value tuples, where each tuple is the result of semantic encoding.
- d) The contents of the data set shall be the data that was the basis for generating the on-ledger identifier that identifies the data set.

NOTE [5 c\)](#) specifies the requirement for on-ledger identifiers to include data that can be used to verify the integrity of the corresponding off-ledger data set.

STANDARDSISO.COM : Click to view the full PDF of ISO 8000-117:2023

Annex A
(informative)

Document identification

To provide for unambiguous identification of an information object in an open system, the following object identifier is assigned to this document. The meaning of this value is defined in ISO 10303-1.

{ iso standard 8000 part(117) version(1) }

STANDARDSISO.COM : Click to view the full PDF of ISO 8000-117:2023