TECHNICAL SPECIFICATION

ISO/TS 22943

First edition
2022-06

Information and documentation — Principles of identification

Information et documentation — Principes d'identification

Reference number
ISO/TS 22943:2022(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

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 46, *Information and documentation, Subcommittee* SC 9, *Identification and description*.

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

ISO/TC46/SC9 is the ISO subcommittee on identification and description that creates standards for content identifiers, description and associated metadata and models for use in information organizations, including libraries, museums and archives, and the content industries, including publishing and other content producers and providers under the auspices of ISO Technical Committee 46 on Information and documentation. The committee is responsible for a suite of identifier standards, including the International Standard Book Number (ISBN), the International Standard Serial Number (ISSN), the International Standard Name Identifier (ISNI), the International Standard Recording Code (ISRC), the International Standard Audiovisual Number (ISAN), the Digital Object Identifier (DOI), the International Standard Link Identifier (ISLI), International Musical Work Code (ISWC), among several other standards. It also oversees the numerous registration authorities that manage these identifier systems.

Identification systems, as defined in this document, create concise codes ("identifiers") that can be used to refer unambiguously to something or someone (the "referent"). In this way, the identifier can be provided as a proxy for the referent itself, for instance in cataloguing it in a library, enquiring about its attributes, ordering a copy of it or (if a person or party) attributing authorship in something. To be useful an identification system should be built upon trust in a variety of its characteristics. These include the reliability of the binding between identifier and the referent, trust in the assignment of the identifier, the related metadata, the longevity of the identification system, and the extent of the system's use in its sector. This document describes the characteristics of identification systems that provide this trust to meet the needs of particular use cases and have other characteristics that make them useful in real world circumstances.

Having an identifier opens the possibility to access trusted information about the referent as well as to facilitate services that use the identifier. These related metadata can include information about the referent, such as rights or pricing information or connected information such as creators or derivatives. Services can include more efficient data exchange, order processing, or discovery. Services can even extend to deliver the referent itself via a supply chain if this is appropriate.

Identification systems are built upon trust among parties in the supply chain of information about a referent. Where the trust in the identifier has been generated by appropriate technologies, policies, and governance, then the value of systems which use it is increased as they enable the curation, discovery, commercial exploitation and preservation of the referents to which an identification system has been applied. This value comes about through increased efficiency for all participants in the ecosystem. There are a number of ways in which the binding between the identifier and the referent can be made trustworthy. This begins with the source of the data and it being recognized as a canonical source of data about the referent. This is further supported by consensus and engagement of the community of implementers and users of the identifier and its related data. The ability of the system to provide data reliably and in ways that users of the identification system require enhances this trust. The viability of the system and its ability to persist, both financially and technologically, buttress the community's willingness to rely on the system.

A core function of an identification system is to draw a distinction between referents. Referents may often be similar and share common attributes. What distinguishes one referent from another is often defined by the use cases that are described by the identifier standard. Clear assignment rules, i.e. the specifications for when a "thing" is the same as another "thing" and when they are different, are a critical component of any identification system. Any community of practice should carefully consider, and be appropriately cautious in adopting, any proposal that increases the number of identifiers used to deal with similar populations of referents. The costs of management escalate with the number of identifiers assigned and finer granularity may, unless clearly justified, increase costs both for the system and for users who need to deal with multiple identifiers to specify what they otherwise regard as a single referent.

The following are two examples of identifiers and some of their associated kernel metadata.

ISSN: ISSN 0261-0523

Sample Kernel Metadata

Key Title: Philosophical transactions of the Royal Society of London

Country: UK

ISSN Center responsible for the record:

ISSN National Centre for the UK

Medium: Print
Last modification date: 22/06/2020

ISAN: ISAN 0000-0000-1BF0-0000-C-0000-0000-1

Sample Kernel Metadata

Title: All The President's Men
Language: ENG

Type: Feature Film

Participants information

Types // First name / Last name

Director // Alan / Pakula (ISNI 0000 0000 8387 6497)

Actor // Dustin / Hoffman (ISNI 0000 0001 2148 4241)

The examples given adhere to many of the core principles of identification system described in this document. Both identifiers are globally unique in a sizable namespace that can cover more than the potential items in its conceivable universe. The identifier examples is in accordance with a variety of underlying standards about structure, and additional referencing standards. Outside of any relevant context, the identifier string itself is devoid of meaning. But within the context of a broader identification system, by being attached with a prefix, an appropriate display, or field heading, the string then becomes bound to a wealth of additional information that exists about the referent. It is not the string of characters, per se, that makes an identifier valuable, it is that associated metadata and the broader identification system that bind the identifier together in the ecosystem to create value. There is also a certain synergy if the attributes themselves are (or can be associated with) other identifiers – this contributes to the "web of identifiers" and enables further discovery (e.g. contributors associated with an ISBN or ISAN may themselves be identified via ISNI, such as noted in the ISAN example).

These systems require a great deal of infrastructure and investment to manage the information, including the assignment process, the curation efforts, the technical systems, and the maintenance of the entire system as community needs grow over time. The principles described in this document help support these systems.

In some cases, identifier systems might not include all the attributes or characteristics described in this document. Not all these attributes need apply in every circumstance. Furthermore, it is understood that there are some elements of these principles which may not be operable in existing or future systems and a lack of conformance with these principles should be evaluated on its merits within that system and its own context. For example, there are a variety of distributed identifier systems that have wide adoption, such as URN or URI, which do not involve centralized management of the identification systems.

Information and documentation — Principles of identification

1 Scope

This document defines the philosophy of why identifiers exist and why they are valuable for trade and information management. It establishes a core set of relevant characteristics and expectations for identifiers and the general business case of guidelines for identifiers. This document explains the reason identifiers are structured the way they are and for what purpose, while acknowledging other communities define identifiers differently.

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 terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1
attributes
characteristics of a *referent* (3.16) that describe and disambiguate it from any other *referent* (3.16) within a given *identifier system* (3.9)

Note 1 to entry: An attribute may be common to other *referents* (3.16) and relate them for a specific purpose.

Note 2 to entry: A particular *referent's* (3.16) attributes, and, in particular, their distinction from those attributes should make the *referent* (3.16) globally unique within that *identifier system* (3.9).

3.2
binding
association between an *identifier* (3.5) and a *referent* (3.16)

Note 1 to entry: This binding may be represented by *metadata* (3.12) recording the relevant *attributes* (3.1) of the *referent* (3.16), or another surrogate.

Note 2 to entry: The binding mechanism may be explicit (recorded in an authoritative database) or implicit (derived from the referent by a deterministic procedure).

3.3
governance
system that manages the overall assignment and curation of *identifiers* (3.5)

Note 1 to entry: This can include bringing together stakeholders involved in the use of the *identifier* (3.5) for the purposes of establishing rules for the assignment of *identifiers* (3.5) and requirements for associated *metadata* (3.12), overseeing the process of assignment, monitoring the system for issues, managing conflict resolution, and suggesting changes to the system.

3.4
granularity
ability to distinguish or cluster within an *identifier system* (3.9)

Note 1 to entry: The granularity of an *identifier system* (3.9) determines when entities are regarded as the same *referent* (3.16) and when they are recognized as distinct from each other.

EXAMPLE 1 A library managing subscriptions to a journal needs to identify a title in a particular format (such as "The New England Journal of Medicine") with an ISSN encompassing the entire corpus of articles, letters, etc., published in that journal, whereas a researcher seeking a scholarly paper needs a more granular *identifier* (3.5), such as an article-level DOI, to cite a single article within it.

EXAMPLE 2 A reprint of a book may carry the same ISBN as the corresponding edition, while a revised edition should have a different ISBN.

3.5
identifier
sequence of characters that uniquely denotes a *referent* (3.16)

Note 1 to entry: Identifiers may be used to specify the *referent* (3.16). In some cases, they may be substituted for the *referent* (3.16) or may be used to retrieve the *referent* (3.16) or its *metadata* (3.12).

3.6
identifier metadata
metadata (3.12) that describes the administration of the *identifier* (3.5), not the *referent* (3.16)

Note 1 to entry: An *identifier* (3.5) may be associated with administrative metadata such as its date of assignment, or the source of mandatory attributes. Identifier metadata should include information about correction of errors, such as a preferred *identifier* (3.5) in instances where a duplicate *identifier* (3.5) has been assigned in error, or when two *referents* (3.16) have been identified as the same thing after *identifiers* (3.5) had been assigned.

3.7
identifier-referent metadata schema
specification of a format and structure for recording the *attributes* (3.1) of a referent that comprise its *metadata* (3.12)

3.8
identifier standard
document specifying an *identifier system* (3.9)

Note 1 to entry: Identifier standard may specify, for example, syntax, scope, *namespace* (3.14), and *kernel metadata* (3.11) of the *identifier system* (3.9).

3.9
identifier system
identification system
structures that support the assignment, management and governance of *identifiers* (3.5) within a community

Note 1 to entry: This document cites some of the identification systems specified in ISO TC 46/SC 9, however, the principles may be useful to other communities developing such systems, whether in ISO or elsewhere.

Note 2 to entry: The identification system may comprise a combination of the following:

— the *identifier standard* (3.8) specifying an *identifier* (3.5), its associated metadata, and assignment mechanism,

— the management systems for collecting, storing and sharing the *identifiers* (3.5) and associated metadata,

— a mechanism for resolving an *identifier* (3.5) to retrieve its *metadata* (3.12) and, in some cases, a *referent* (3.16),

— its *governance* (3.3) systems,

— the social network that supports the use of the *identifier* (3.5).

3.10
identify
use an identifier string to represent a *referent* (3.16)

3.11
kernel metadata
minimum set of referent *attributes* (3.1) that provides sufficient information to uniquely specify the *referent* (3.16), and that is thus key to its disambiguation

Note 1 to entry: Common synonyms used to describe kernel metadata include "minimum metadata" or "reference metadata".

3.12
metadata
data about other data, documents or records that describes their content, context, structure, data format, provenance, and/or rights attached to them

[SOURCE: ISO 5127:2017, 3.1.10.26.01]

3.13
minting system
subsystem of an *identifier system* (3.9) by which new *identifiers* (3.5) are created

3.14
namespace
set of *identifiers* (3.5) with a defined scope of usage, and within which a single *identifier* (3.5) is unique

3.15
persistence
survival over time of key elements of an *identifier system* (3.9)

Note 1 to entry: Key elements include *identifiers* (3.5), their relationship to the *referents* (3.16), records of the *binding* (3.2) between each *referent* (3.16) and its *identifier* (3.5), *identifier metadata* (3.6) and *referent metadata* (3.17).

3.16
referent
entity that is referenced by an *identifier* (3.5)

Note 1 to entry: Referents may be physical, digital, abstract, names, legal entities, items, content, components of items, persons, or classes of these, etc.

3.17
referent metadata
all *metadata* (3.12) associated with the *referent* (3.16)

Note 1 to entry: This includes the *kernel metadata* (3.11) and any other *metadata* (3.12) associated with the *referent* (3.16). Referent metadata may include *identifiers* (3.5) from other *identifier systems* (3.9) to enable further discovery about the *referent* (3.16). For example, a contributor to a *referent* (3.16) with an ISBN may be identified through an ISNI, thus providing access to information about the contributor without incurring a maintenance overhead within the referent metadata. Some referent metadata might not be made publicly available, for business, privacy, or other reasons.

3.18
Registration Authority
organization delegated to manage the implementation, maintenance, and promotion of an *identifier standard* (3.8)

3.19
Registration Agency (of an *identifier system* (3.9))
entity entrusted by a Registration Authority to perform certain tasks or services as defined within an *identifier standard* (3.8)

Note 1 to entry: There may be multiple registration agencies serving particular regions, market segments, etc.

3.20
resolution
<of an identifier> utilisation of the *identifier* (3.5), *metadata* (3.12) associated with the *referent* (3.16) and/or the *referent* (3.16) itself

3.21
sector
group of users or potential users whose needs are sufficiently aligned that they can use the *identifier system* (3.9)

Note 1 to entry: There are efficiency benefits to minimising the number of *identifier systems* (3.9) required to meet sector needs.

4 Optimal attributes of identifiers and identifier systems

4.1 Uniqueness

Identifiers should be globally unique within the context of their own identifier system. No two different referents should be assigned the same identifier within a namespace, nor should the same referent within a namespace be assigned two different identifiers. Uniqueness is related to the granularity of the identifier system as described in 4.3.

4.2 Persistence

In order to provide value over time and across institutions, the identifier system should provide stability and trust in the various elements in the system. These various aspects of an identifier system should have appropriate levels of persistence for its intended application. For example:

— The persistence of the identifier, to include:
  — identifiers should persist separately from the persistence of the registration system or minting system;
  — a registration system should make provision for data preservation of the kernel metadata and the binding;
— The persistence of the binding that connects the referent and its identifier, to include:
  — the relationship between the referent and the identifier shall be maintained through the curation of metadata over time;
  — the persistence of a representation of the referent bound to the identifier. This can include the metadata, the identifier and the binding;
— The persistence of the identifier systems that support the trust in the identifier, such as registration authority, a delegated registration agency, resolution systems, minting systems, a method for updating the metadata, etc., including technical, financial and/or legal sustainability.

4.3 Granularity of the identifier to suit the needs of the sector

Identifiers should, subject to the need for sustainability, be assigned at the granularity that meets the agreed needs of the sector. Within the context of granularity, identifiers may be assigned at a single, fine-grain, or fundamental level. Alternatively, a hierarchical system may involve the aggregation of

multiple fine-grain identities into a single identifier whose level of detail suits particular needs. A hierarchical identifier may combine numerous granular identifiers into a single referent.

In determining these needs, a system should recognise the interests of a sector when there is a need to distinguish referents or when combinations are sufficient.

If different parts of the sector have requirements so distinct that they cannot be reconciled, separate systems (or possibly a single system with separate schemes) are necessary.

EXAMPLE The ISO identifier for sound recordings and music video recordings, an International Standard Recording Code (ISRC) focuses on rights management and content discovery, so it is assigned at a suitable granularity to distinguish studio and live versions of the same song (which have different identifiers), whereas an International Standard Musical Work Code (ISWC), which focuses on abstract musical compositions for copyright administration, intrinsically combines all technical encodings (MP3, FLAC, etc.) or versions of the same recorded song.

4.4 Stability of kernel metadata

While the binding between the referent and the identifier should remain stable and be persistent, attributes of the referent can change and, therefore, its metadata can change over time. The kernel metadata associated with an entity and that bind it to its identifier should not normally change over time so that the binding is robust and secure. Any changes to kernel metadata (to correct errors for instance – see 4.11) should be made carefully and with full understanding of the effect on binding that will be produced. In addition, if errors are found or if changes become required, they can be corrected by addressing those errors in the attributes described in the metadata. Other metadata should be updated throughout the referent lifecycle to ensure its accuracy.

4.5 No ambiguity within its own namespace

Kernel metadata should be comprehensive enough to determine uniqueness. Based on the granularity of the identifier system and the kernel metadata about the referent, the identifier should be uniquely bound to the referent. The sector and the managers of the identification system should determine which characteristics demand granular identifiers and appropriate metadata to uniquely describe the referent.

4.6 Access

While identification systems can be useful for internal processes within a single organization, impact is driven by the network effects of the use of the system. As such, public identifiers, along with the kernel metadata, should be available openly, ideally via automated machine interfaces to facilitate interoperability. Any system user should be able to disambiguate the referent and connect the identifier to the referent. That kernel data is made publicly available does not automatically imply that such data should be available free of charge, although it should be available under commercially reasonable and non-discriminatory terms.

Additional metadata may be stored in the identifier system which is not made available publicly for either business, privacy, or other reasons.

Although external rich metadata (other than kernel metadata) is not formally part of an identification system, enabling access to such data is usually a main reason to establish those management systems (for example, a library catalogue or retail inventory system).

As with access to rich metadata, access to the referent itself may be available under certain circumstances from other systems. Accurately specifying the referent is enabled by an effective identification system and appropriate parties providing such access should also be considered as stakeholders.

Cooperation between providers of identification systems and providers of services will tend to increase the utility and increase the sustainability of both.

4.7 Scope

It is essential to know from the outset what class of things an identifier identifies (e.g., that ISBNs identify monographic publications in their supply chains, that ISWCs identify musical compositions, etc.). When the identifier deals with a particular level of abstraction, it is important to define this abstraction clearly in its scope. For example, an abstract textual work, the typesetting source file expressing this work and a physical book available for sale are at different levels of abstraction and cannot normally share an identification system. Examples of hierarchies of abstraction are defined in the systems established under Functional Requirements for Bibliographic Records (FRBR)[5], Indecs[11], and others.

This scope may be specified through the identifier standard, its governance structure, and/or the relevant assignment processes. The defined scope guides the kernel metadata related to the identifier.

4.8 Absence of semantics in identifier strings

The identifier should not encode any of the attributes of the referent. While names of things can be used as identifiers, names and other semantic information should not be included in a formal identifier.

The identifier itself should have one or a small number of standardised "presentations" [e.g. as a plain string or number, as a Uniform Resource Identifier (URI)] sufficient to connect the identifier to its namespace and enable resolution (see 4.9).

The representation of the identifier should normally require the inclusion of information to achieve recognition and resolution, which might be a prefix to the identifier, but this is distinct from the semantics of the identifier discussed here. The context in which an identifier is used (or may, in future, be used), will determine the specific needs for such identifier system semantics.

In some cases, organizations may use attribute information in the process of creating (encoding or assigning) an identifier string in their minting systems. However, after the information has been incorporated in an identifier, it should no longer be construed as maintaining meaning itself, although the referent metadata may contain similar data.

4.9 Resolution

Identification systems should resolve the identifier to the kernel metadata, thereby allowing the verification of the binding between the identifier and the referent. Where possible, an identifier should be actionable, such that it provides access to the resolution service, the identifier's metadata or to the referent.

A resolution service that provides this shared access to kernel metadata should be available to all parties that are using the system so that they have trust in the system.

Further information may also be available through the resolution process and this may create additional value for the identification system. Additional information may be accessed from third party systems by using the identifier as a look-up key.

4.10 Timing of assignment

The origination of a referent and the assignment of an identifier are not always contemporaneous. However, this process should be coordinated through the use of identifiers and identification systems. Therefore, identifiers should be assigned as early as possible. When an identifier is assigned early in the creation process, then care should be taken to ensure the metadata is updated to reflect any changes in the referent in the development process.

4.11 Resilience in the face of errors

Every system will experience occasional errors. Ideally, identification systems will have, and make public, clearly defined processes to promptly deal with errors of assignment when they are discovered – for example, merge/split procedures, deprecation of "old" and promotion of canonical identifiers, or