
**Information and documentation —
Codes for written language conversion
systems**

*Information et documentation — Codes pour les systèmes de
conversion des langues écrites*

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO 24229:2022

<https://standards.iteh.ai/catalog/standards/sist/f41ef8db-8b75-4ba6-9f79-13373be75abc/iso-24229-2022>



iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO 24229:2022

<https://standards.iteh.ai/catalog/standards/sist/f41ef8db-8b75-4ba6-9f79-13373be75abc/iso-24229-2022>



COPYRIGHT PROTECTED DOCUMENT

© ISO 2022

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	v
Introduction	vi
1 Scope	1
2 Normative references	1
3 Terms and definitions	1
4 Conversion system codes	3
4.1 Structure of conversion system codes.....	3
4.1.1 General.....	3
4.1.2 Construction of the conversion system code.....	3
4.1.3 Titular segment.....	3
4.1.4 Source spelling system segment.....	4
4.1.5 Target spelling system segment.....	4
4.1.6 Identifying segment.....	4
4.2 Requirements for new conversion system codes.....	4
4.3 Deprecation of conversion system codes.....	5
4.4 User assigned conversion system codes.....	5
4.5 Capitalization of conversion system codes.....	5
4.6 Abbreviated conversion system codes.....	5
4.7 Examples of conversion system codes.....	5
5 Conversion system authority	6
5.1 General.....	6
5.2 Requirements.....	6
5.2.1 General.....	6
5.2.2 Inactive authorities.....	6
5.2.3 Varia authorities.....	6
5.2.4 Competency.....	6
5.3 Registration.....	7
5.4 Conversion system authority identifiers.....	7
5.4.1 Principles for construction of identifiers.....	7
5.4.2 Examples of conversion system authority identifiers.....	8
6 Data model and attributes	8
6.1 Common data model and attributes.....	8
6.1.1 General.....	8
6.1.2 Data models.....	8
6.1.3 Usage of ISO 15924 code elements.....	8
6.1.4 Usage of ISO 639 code elements.....	8
6.1.5 Usage of ISO 3166 code elements.....	8
6.1.6 Usage of ISO 8601 expressions.....	8
6.2 System authority data model and attributes.....	9
6.2.1 Diagram.....	9
6.2.2 Conversion system authority.....	9
6.2.3 Authority identifier.....	9
6.3 Conversion system data model and attributes.....	10
6.3.1 Diagram.....	10
6.3.2 Written language conversion system.....	10
6.3.3 Spelling system.....	11
6.3.4 Conversion system relation.....	11
6.3.5 Conversion system code status.....	11
6.3.6 Conversion system status.....	11
6.3.7 Conversion system relation type.....	12
Annex A (normative) Registration authority	13

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO 24229:2022

<https://standards.iteh.ai/catalog/standards/sist/f41ef8db-8b75-4ba6-9f79-13373be75abc/iso-24229-2022>

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*.

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

A number of international applications require the identification of written language conversion systems, including for terminology, lexicography, bibliography, and linguistics, especially for reverse transliteration, computational linguistics and machine pronunciation.

This document sets out the necessary procedures to maintain the registry of written language conversion systems.

The chosen term “written language conversion” is intended to refer to all types of conversions, i.e. transformations of written texts from one spelling system to another. It thus includes both script conversion (change of script: transliteration, transcription) and conversion of texts without changing the script (e.g. transcription of foreign names or words using the alphabet of a target language, change of the orthography in a language, etc.). For the sake of compactness of expression, “written language conversion” has been shortened to “conversion” in this document where it does not cause ambiguity.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO 24229:2022

<https://standards.iteh.ai/catalog/standards/sist/f41ef8db-8b75-4ba6-9f79-13373be75abc/iso-24229-2022>

Information and documentation — Codes for written language conversion systems

1 Scope

This document provides principles for establishing codes for the representation of written language conversion systems.

The codes are devised for usage in any application requiring the expression of written language conversion systems, including transliteration and romanization systems, in coded form.

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 639-2, *Codes for the representation of names of languages — Part 2: Alpha-3 code*

ISO 639-3, *Codes for the representation of names of languages — Part 3: Alpha-3 code for comprehensive coverage of languages*

ISO 639-5, *Codes for the representation of names of languages — Part 5: Alpha-3 code for language families and groups*

ISO 3166-1, *Codes for the representation of names of countries and their subdivisions — Part 1: Country code*

ISO 5127, *Information and documentation — Foundation and vocabulary*

ISO 8601 (all parts), *Date and time — Representations for information interchange*

ISO 15924, *Information and documentation — Codes for the representation of names of scripts*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 5127 and the following 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 <http://www.electropedia.org>

3.1

script

particular graphic representation or class of representations of a set of characters used to write one or more languages

[SOURCE: ISO 5127:2017, 3.1.6.02]

**3.2
spelling system**

set of rules governing the orthography of a language

Note 1 to entry: Typically, a spelling system defines how the spoken form of a language is represented in writing. Several languages have undergone orthographic reforms which means they have had different spelling systems.

**3.3
natural language**

language which is or was in active use in a community of people, and the rules of which are mainly deduced from the usage

[SOURCE: ISO 5127:2017, 3.1.5.02]

**3.4
character**

member of a set of elements that is used for the representation, organization, or control of data

[SOURCE: ISO 5127:2017, 3.1.4.02]

**3.5
written language**

natural language (3.3) realized through the writing of *characters* (3.4)

[SOURCE: ISO 5127:2017, 3.1.5.04]

**3.6
written language conversion**

process whereby one *spelling system* (3.2) is converted into another spelling system

Note 1 to entry: This is a general term that includes script conversion but also, e.g. cases when a language changes its orthography without changing the script.

**3.7
transliteration**

process which consists of representing the characters of an alphabetical or syllabic system of writing by the characters of a conversion alphabet

**3.8
transcription**

process whereby the sounds of a given language are noted by the system of signs of a conversion language

**3.9
romanization**

script conversion from non-Roman to Roman *script* (3.1) by means of *transliteration* (3.7), *transcription* (3.8) or both

[SOURCE: ISO 5127:2017, 3.1.6.14]

**3.10
written language conversion system**

set of rules for *written language conversion* (3.6)

**3.11
language code**

combination of characters used to represent the name of a *language* or languages

[SOURCE: ISO 5127:2017, 3.2.5.14]

3.12**script code**

combination of characters used to represent the name of a *script* (3.1)

[SOURCE: ISO 15924:2004, 3.8]

3.13**conversion system code**

combination of characters used in a structured way to represent a *written language conversion system* (3.10)

4 Conversion system codes**4.1 Structure of conversion system codes****4.1.1 General**

A conversion system code shall consist of four segments:

- titular segment;
- source spelling system segment;
- target spelling system segment;
- identifying segment.

Each segment shall consist of one or more elements.

4.1.2 Construction of the conversion system code

The following rules are to be adhered to for the construction of a conversion system code:

- The codes shall consist of elements from the following Unicode ranges:
 - DIGIT ZERO through DIGIT NINE (U+0030 — U+0039)
 - LATIN CAPITAL LETTER A through LATIN CAPITAL LETTER Z (U+0041 — U+005A)
 - LATIN SMALL LETTER A through LATIN SMALL LETTER Z (U+0061 — U+007A)
- Segments shall be separated by a single “COLON” (“:”, Unicode U+003A).
- Elements within a segment shall be separated by a single “HYPHEN-MINUS” (“-”, Unicode U+002D).
- “HYPHEN-MINUS” (“-”, Unicode U+002D) within an element (e.g. 233-3) will also be accepted.
- Other characters in the elements not covered by the above should be omitted or substituted.

4.1.3 Titular segment

This part will contain a reference to the conversion system authority or authorities by using identifiers, the list of which is maintained by ISO 24229/RA (see A.1). If an authority cannot be identified but the conversion system has a national character and/or is used by the government, the 2-letter country code from ISO 3166-1 should be used as the conversion system authority. If no conversion system authorities can be identified or its identification is not relevant, “Var” (varia) is used as the titular segment. See [Clause 5](#) for more details.

4.1.4 Source spelling system segment

Except as specified in 4.6, a script code is a mandatory element. Language-specific spelling systems also have language codes. In order to cover more specific needs, the following four elements in the order given shall be used:

- language code (3-letter code from ISO 639-2 or ISO 639-3 with preference to terminological codes. If a synonym is used from ISO 639-2, the ISO 639-2/T associated code should be used. ISO 639-2/T codes are intended to be used for terminology applications.);
- script code (4-letter code from ISO 15924);
- country code (2-letter code from ISO 3166-1);
- spelling system extension (an ad hoc string to refer to a non-default spelling system of a language, such as old orthography).

EXAMPLE 1 ind-Latn-pre1972 (Indonesian language using the pre-1972 orthography).

EXAMPLE 2 bos-Arab (Bosnian language using Arabic script).

EXAMPLE 3 uzb-Arab-AF (Uzbek language as used in Afghanistan).

4.1.5 Target spelling system segment

This part may have the same four elements as listed in 4.1.4.

4.1.6 Identifying segment

This part will serve to distinguish by version, year of issue, etc. conversion systems that otherwise have the same scope. It may also contain elements necessary for the recognition of the system itself if the system has some kind of identification element. All in all, the following elements may occur (in the order given):

- identifying numbers, letters or else (such as standard number, e.g. 843);
- version number (e.g. v6, v4-1);
- year of adoption;
- year of issue;
- method identifier (if a standard devises more than one method of conversion, this optional ad hoc identifier can be used for distinction).

If there are cases when no elements can be used for this part, “na” (not applicable) will be the substitute.

EXAMPLE 2017 is the identifying segment of the system coded as UN:ara-Arab:Latn:2017.

4.2 Requirements for new conversion system codes

Additions to the list of conversion system codes shall be made on the basis of information from upon the request of a member of ISO 24229/AG (see A.2) or the conversion system authority that manages this system.

The ISO 24229/AG decides upon the addition, on the basis of the justification given for the actual requirements for international interchange. Code elements will be allocated accordingly.

A written language conversion system is eligible for a conversion system code assignment if it fulfils one of the following criteria.

- The system has been approved for official use at some level of government.

- The system has been developed and used by educational/scientific institutions, published in a peer reviewed scientific publication.
- The system has been in substantial usage.

Assigning of a conversion system code also requires demonstration of one of the following usage factors:

- necessity of identification of the system in interchange.
- necessity of identification of the system in data encoding.

Systems that are used in isolation or only for temporary usage do not need to have assigned codes.

4.3 Deprecation of conversion system codes

Deprecation of conversion system codes shall be made upon request of a member of ISO 24229/AG or the conversion system authority that manages the system.

The ISO 24229/AG will decide upon the marking of deprecation, on the basis of the information received. The corresponding code is reserved for backwards-compatibility.

NOTE Deprecation only applies to the code representation of the written languages conversion system, and not the system itself. For example, deprecation can be necessary when the authority undergoes a rename.

4.4 User assigned conversion system codes

If users need codes to represent conversion systems not included in the conversion system registry, the code prefix of `zz` can be used, which shall be placed at the beginning of the conversion system code, in the titular segment, and followed by a “HYPHEN MINUS” character (“-”, Unicode U+002D).

NOTE Users are advised that the above series of codes are not universally used, those code elements are not compatible between different entities.

<https://standards.iteh.ai/catalog/standards/sist/f41ef8db-8b75-4ba6-9f79-13373be75abc/iso-24229-2022>

4.5 Capitalization of conversion system codes

Conversion system codes will use capitalization according to the relevant standards but this does not have any distinctive meaning. For example, an all lower case code will be an equally valid code.

4.6 Abbreviated conversion system codes

In case of user demand, abbreviated conversion system codes may additionally be registered whereby in identifying language-specific spelling systems script codes are omitted if they can be considered as default scripts for the languages concerned. Examples are given in 4.7. Sources, such as Common Locale Data Repository (CLDR) of the Unicode Consortium, should be consulted when determining default scripts for languages.

4.7 Examples of conversion system codes

The examples given here are only indicative and do not guarantee that such codes will be actually registered.

EXAMPLE 1 UN:ara-Arab:Latn:2017 (possible abbreviation — UN:ara:Latn:2017; United Nations system for the romanization of Arabic, approved 2017)

EXAMPLE 2 UN:mon-Mong-CN:Latn:1977 (possible abbreviation — UN:mon-CN:Latn:1977; United Nations system for the romanization of Mongolian in China, approved 1977)

EXAMPLE 3 BGN-PCGN:chn-Hans:Latn:1979 (BGN/PCGN 1979 Agreement — Romanization of Chinese)

EXAMPLE 4 ALA-LC:mal-Mlym:Latn:2012 (possible abbreviation — ALA-LC:mal:Latn:2012; ALA-LC romanization system that transliterates the Malayam language from Malayam script characters into Latin script)