



ISO/TC 37/SC 2 "Terminology workflow and language coding"  
Secretariat: **SCC**  
Committee manager: **Benhoff Maryse Ms**



## ISO 12199 forward

Document type	Related content	Document date	Expected action
Project / Other	Meeting: <a href="#">Brussels (Belgium) 11 Jun 2023</a> Project: <a href="#">ISO 12199:2022</a>	2023-05-22	<b>INFO</b>



# ISO 12199

## Quo vadis?

2023-06  
Håvard Hjulstad

# ISO 12199 :2000 & :2022

## Alphabetical ordering of multilingual terminological and lexicographical data represented in the Latin alphabet

INTERNATIONAL  
STANDARD

ISO  
12199

First edition  
2000-08-01

Alphabetical ordering of multilingual  
terminological and lexicographical data  
represented in the Latin alphabet

*Mise en ordre alphabétique des données lexicographiques et  
terminologiques multilingues représentées dans l'alphabet latin*



Reference number  
ISO 12199:2000(E)

© ISO 2000



- Identical normative technical content
  - Informative content relating to Serbian language added
  - Minor updates of references
    - ... very minor ... :
- “1) In this minor revision of ISO 12199:2000, reference continues to be made to ISO/IEC 10646-1:1993. ISO/IEC 10646-1 and ISO/IEC 10646-2 have since been merged into ISO/IEC 10646:2020.”

INTERNATIONAL  
STANDARD

ISO  
12199

Second edition  
2022-06

Alphabetical ordering of multilingual  
terminological and lexicographical  
data represented in the Latin alphabet

*Mise en ordre alphabétique des données lexicographiques et  
terminologiques multilingues représentées dans l'alphabet latin*



Reference number  
ISO 12199:2022(E)

© ISO 2022



# Structure of ISO 12199

**Foreword**

**Introduction**

**1 Scope**

**2 Normative references**

**3 Terms and definitions**

**4 Preparatory procedures**

**5 First ordering level**

5.1 First-ordering-level values

5.2 First-ordering-level sequence

5.3 Equivalence between special Latin letters  
and basic letters

**6 Second ordering level**

6.1 Second-ordering-level values

6.2 Special Latin letters and letters with  
diacritical marks

**7 Third ordering level**

7.1 Third-ordering-level values

7.2 Ordering according to capitalization

**8 Fourth ordering level**

8.1 Fourth-ordering-level values

8.2 Ordering according to special characters

**Annex A (normative) Word-by-word ordering**

**Annex B (informative) Special rules for  
lexicographical and terminological ordering**

**Annex C (informative) Ordering rules for  
chemical names**

**Annex D (informative) Character repertoire of  
the Latin alphabet**

**Annex E (informative) Languages using the  
Latin alphabet**

**Annex F (informative) Alphabetical sequences  
and character repertoires**

**Annex G (informative) Formal description of  
the rules of the main body of this document**

**Bibliography**



# The basic rules of ordering levels 1 through 4

1. 0 1 2 3 4 5 6 7 8 9

a b c d e f g h i j k l m n o p q r s t u v w x y z þ

(mentioning the issue of Turkish i/I and i/İ)

2. Diacritical marks:

o ó ò õ ô ö ö ö ö ñ ò ı etc.

3. lowercase before UPPERCASE

4. Special characters according to the sequence of the default template of ISO/IEC 14651

## Seven annexes

- A. Word-by-word ordering
- B. Special rules for lexicographical and terminological ordering
- C. Ordering rules for chemical names
- D. Character repertoire of the Latin alphabet
- E. Languages using the Latin alphabet
- F. Alphabetical sequences and character repertoires
- G. Formal description of the rules of the main body

# Annex A (normative)

## Word-by-word ordering

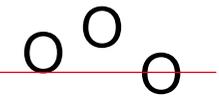
1. as multiple-key ordering
  - Space characters serve as key separators
2. Simple word-by-word ordering
  - Space character is first character on level 1

This is treated exactly the same way in EN 13710, in informative Annex B.

## Annex B (informative)

### Special rules for lexicographical and terminological ordering

#### 1. Position relative to baseline

- 

#### 2. Styles

- abc **abc** *abc* ***abc***

This feature is covered by informative Annex C in EN 13710.

“The features that are described in this annex cannot easily be described in the formalism given in ISO/IEC 14651.” (ISO 12199)

# Annex C (informative)

## Ordering rules for chemical names

- Three keys:
  - parent name; initial locants; other locants

This is not covered by the other documents. Are there other “special ordering rules” that could (or should) be included?

Ordered according to Annex C	Ordered according to general rules
Bromine fluoride (BrF <sub>3</sub> )	1-Butanone
Bromine fluoride (BrF <sub>5</sub> )	1-Butanone, 1-phenyl-
2-Butanol	2-Butanol
2-Butanol, ( <i>R</i> )-	2-Butanol, 1-chloro-
2-Butanol, ( <i>S</i> )-	2-Butanol, 4-(trimethylstannyl)-
2-Butanol, sodium salt, ( <i>S</i> )-	2-Butanol, ( <i>R</i> )-
2-Butanol, 1-chloro-	2-Butanol, ( <i>S</i> )-
2-Butanol, 4-(trimethylstannyl)-	2-Butanol, sodium salt, ( <i>S</i> )-
1-Butanone	2-Butanone
1-Butanone, 1-phenyl-	2-Butanone, 1-(dimethylamino)-3,3-dimethyl-
2-Butanone	2-Butanone-1,1,1- <i>d</i> <sub>3</sub>
2-Butanone, <i>O</i> -methyloxime	2-Butanone-1,1,1- <i>d</i> <sub>3</sub> , 3,3-dimethyl-
2-Butanone, oxime	2-Butanone, 3-(4-acetylphenyl)-
2-Butanone, polymer with formaldehyde	2-Butanone, 3-ethoxy-1,1-dihydroxy-
2-Butanone, 3-(4-acetylphenyl)-	2-Butanone, <i>O</i> -methyloxime
2-Butanone, 1-(dimethylamino)-3,3-dimethyl-	2-Butanone, oxime
2-Butanone, 3-ethoxy-1,1-dihydroxy-	2-Butanone, polymer with formaldehyde
2-Butanone-1,1,1- <i>d</i> <sub>3</sub>	Bromine fluoride (BrF <sub>3</sub> )
2-Butanone-1,1,1- <i>d</i> <sub>3</sub> , 3,3-dimethyl-	Bromine fluoride (BrF <sub>5</sub> )
Butanoyl chloride	Butanoyl chloride

## Annex D (informative)

### Character repertoire of the Latin alphabet

Name	Type	Position for lowercase/uppercase in ISO/IEC 10646-1		Languages used
LATIN LETTER A	b	U0061	U0041	+
WITH ACUTE	d	U00E1	U00C1	af ca cs cy da es fo fur ga gd gl hu is kl nl pt qal sk smi ss vi [Cyr] [ar]
WITH GRAVE	d	U00E0	U00C0	ca cy de fr fur fy gd it nl no pt qal rm vi [Cyr]
WITH BREVE	d	U0103	U0102	mo ro vi [Cyr]
WITH BREVE AND ACUTE	d	U1EAF	U1EAE	vi
WITH BREVE AND GRAVE	d	U1EB1	U1EB0	vi

This is not covered by the other documents, but it may be derived from information in the Unicode CLDR (to be explained later in this presentation).

## Annex E (informative)

### Languages using the Latin alphabet

This may also be derived from the Unicode CLDR.

kr	kau	Kanuri	kanouri
ki	kik	Kikuyu; Gikuyu	kikuyu
rw	kin	Kinyarwanda; Rwanda	kinyarwanda; rwan
rn	run	Kirundi; Rundi	kirundi; rundi
kg	kon	Kongo; Kikongo	kongo
ku	kur	Kurdish	kurde
kj	kua	Kwanyama	kwanyama
–	qal	Ladin	ladin
–	lad	Ladino	ladino; judéo-espagnol
la	lat	Latin	latin
lv	lav	Latvian; Lettish	letton; lettonien
ln	lin	Lingala	lingala
lt	lit	Lithuanian	lituanien
lu	lub	Luba; Luba-Katanga	tchiluba; luba-katanga
lg	lug	Luganda; Ganda	louganda; ganda
lb	ltz	Luxemburgish	luxembourgeois

# Annex F (informative)

## Alphabetical sequences and character repertoires

- Language-specific ordering rules for a number of languages (but far from all).

This may also be derived from the Unicode CLDR.

Language	Variants; References	Deviations and other information on alphabet (① first level; ② other levels; ③ native letters; ④ non-native)
Afrikaans af / afr		① - ② - ③ á U00E1    é U00E9    è U00E8    ê U00EA    ë U00EB    í U00ED î U00EE    ï U00EF    ó U00F3    ô U00F4    ö U00F6    ú U00FA û U00FB    ý U00FD    (à) U00E0    (ä) U00E4    (ç) U00E7 (i) U00EC (ò) U00F2    (ü) U00F9 ④ þ
Albanian sq / sqi		① c + ç    d + dh    e + ë    g + gj    l + ll    n + nj    r + rr s + sh    t + th    x + xh    z + zh ② - ③ ç U00E7    ë U00EB ④ þ
Basque eu / eus		① l + ll    n + ñ    t + tt ② - ③ ñ U00F1    (ç) U00E7    (ü) U00FC ④ þ

# Annex G (informative)

## Formal description of the rules of the main body of this document

This is what is done in the main body of ISO/IEC 14651 and EN 13710. In ISO 12199 this is included “for information”.

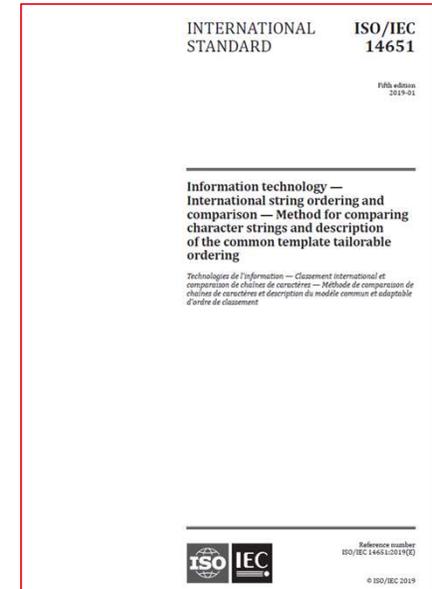
```
<U010F> <d>; <CARON>; <LOWERCASE>; <U010F> % LATIN SMALL LETTER D WITH CARON
<U010E> <d>; <CARON>; <UPPERCASE>; <U010E> % LATIN CAPITAL LETTER D WITH CARON
<U1E0B> <d>; <DOT_ABOVE>; <LOWERCASE>; <U1E0B> % LATIN SMALL LETTER D WITH DOT ABOVE
<U1E0A> <d>; <DOT_ABOVE>; <UPPERCASE>; <U1E0A> % LATIN CAPITAL LETTER D WITH DOT ABOVE
<U1E0D> <d>; <DOT_BELOW>; <LOWERCASE>; <U1E0D> % LATIN SMALL LETTER D WITH DOT BELOW
<U1E0C> <d>; <DOT_BELOW>; <UPPERCASE>; <U1E0C> % LATIN CAPITAL LETTER D WITH DOT BELOW
<U1E0F> <d>; <MACRON_BELOW>; <LOWERCASE>; <U1E0F> % LATIN SMALL LETTER D WITH LINE BELOW
<U1E0E> <d>; <MACRON_BELOW>; <UPPERCASE>; <U1E0E> % LATIN CAPITAL LETTER D WITH LINE BELOW
<U0111> <d>; <MODIFIED1>; <LOWERCASE>; <U0111> % LATIN SMALL LETTER D WITH STROKE
<U0110> <d>; <MODIFIED1>; <UPPERCASE>; <U0110> % LATIN CAPITAL LETTER D WITH STROKE
<U0257> <d>; <MODIFIED2>; <LOWERCASE>; <U0257> % LATIN SMALL LETTER D WITH HOOK
<U018A> <d>; <MODIFIED2>; <UPPERCASE>; <U018A> % LATIN CAPITAL LETTER D WITH HOOK
<U00F0> <d>; <MODIFIED3>; <LOWERCASE>; <U00F0> % LATIN SMALL LETTER ETH
<U00D0> <d>; <MODIFIED3>; <UPPERCASE>; <U00D0> % LATIN CAPITAL LETTER ETH
<U02A3> "<d><z>"; <BLANK><BLANK>; <LOWERCASE><LOWERCASE>; <U0064><U007A> % LATIN SMALL
LETTER DZ DIGRAPH
<U01F3> "<d><z>"; <BLANK><BLANK>; <LOWERCASE><LOWERCASE>; <U0064><U007A> % LATIN SMALL
LETTER DZ
<U01C6> "<d><z>"; <BLANK><CARON>; <LOWERCASE><LOWERCASE>; <U0064><U017E> % LATIN SMALL
LETTER DZ WITH CARON
<U01F2> "<d><z>"; <BLANK><BLANK>; <UPPERCASE><LOWERCASE>; <U0044><U007A> % LATIN CAPITAL
LETTER D WITH SMALL LETTER Z
<U01F1> "<d><z>"; <BLANK><BLANK>; <UPPERCASE><UPPERCASE>; <U0044><U005A> % LATIN CAPITAL
LETTER DZ
<U01C5> "<d><z>"; <BLANK><CARON>; <UPPERCASE><LOWERCASE>; <U0044><U017E> % LATIN CAPITAL
LETTER D WITH SMALL LETTER Z WITH CARON
<U01C4> "<d><z>"; <BLANK><CARON>; <UPPERCASE><UPPERCASE>; <U0044><U017D> % LATIN CAPITAL
LETTER DZ WITH CARON
<U0065> <e>; <BLANK>; <LOWERCASE>; <U0065> % LATIN SMALL LETTER E
```



# ISO/IEC 14651 :2019

*Information technology — International string ordering and comparison — Method for comparing character strings and description of the common template tailorable ordering*

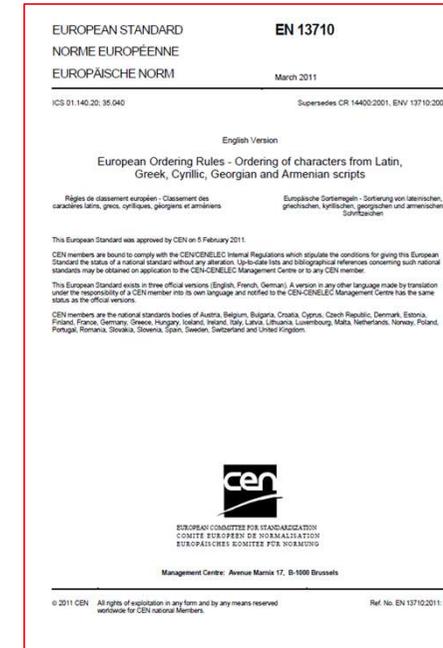
- The document specifies a formalism for ordering and comparing character strings of any script or combination of scripts.
- All characters that are included in ISO/IEC 10646 are covered, but not features that are not explicit in the character encoding (for instance style).
- The ordering template is tailorable according to language-specific ordering rules.



# EN 13710 :2011

## European Ordering Rules — Ordering of characters from Latin, Greek, Cyrillic, Georgian and Armenian scripts

- The document may be seen as a subset of ISO/IEC 14651 covering the languages and scripts used in Europe.
- It is expressed as a delta of (deviation rules from) ISO/IEC 14651.
- It also specifies rules for language-specific deltas based on EN 13710.



# Unicode CLDR

## The Unicode Common Locale Data Repository

<https://cldr.unicode.org/>

It includes:

- **Locale-specific patterns for formatting and parsing:** dates, times, timezones, numbers and currency values, measurement units, ...
- **Translations of names:** languages, scripts, countries and regions, currencies, eras, months, weekdays, day periods, time zones, cities, and time units, emoji characters and sequences (and search keywords), ...
- **Language & script information:** characters used; plural cases; gender of lists; capitalization; rules for sorting & searching; writing direction; transliteration rules; rules for spelling out numbers; rules for segmenting text into graphemes, words, and sentences; keyboard layouts ...
- **Country information:** language usage, currency information, calendar preference, week conventions, ...
- **Validity:** Definitions, aliases, and validity information for Unicode locales, languages, scripts, regions, and extensions, ...



## Does ISO 12199 add anything?

- The normative parts of ISO 12199 do not add specifications that are not covered by ISO/IEC 14651, except that the specifications are expressed in “human language” in ISO 12199.
- Annex B (“Special rules for lexicographical and terminological ordering”), some of which is also included in EN 13710, includes specifications that are not covered by ISO/IEC 14651.
- Annex C (“Ordering rules for chemical names”) is unique to ISO 12199.
- Annexes D (“Character repertoire of the Latin alphabet”), E (“Languages using the Latin alphabet”), and F (“Alphabetical sequences and character repertoires”) do not add much, but may for some purposes serve as useful collections.
- ISO 12199 is developed by the “language committee” of ISO.



# Possible ways forward

Formally, ISO 12199:2023 is a valid International Standard until July 2027, unless ISO/TC 37/SC 2 in the meantime decides to

1. extend its validity for another five years, a process that may be repeated an indefinite number of times
2. revise the entire document
3. withdraw ISO 12199 as such, but develop a new document based on parts of the current document, in particular informative parts
4. withdraw ISO 12199 in its entirety with no replacement

It is, in particular, options 2 and 3 that we need to study



## Why might option 4 or 3 be a good idea?

That would mean withdrawing all of ISO 12199 or the normative parts of the document.

- For all practical purposes the rules of alphabetical ordering are maintained by ISO/IEC JTC 1/SC 2 in ISO/IEC 14651.
- The maintenance of the normative parts of ISO 12199 would be to ensure alignment with ISO/IEC 14651.
- In the case of lack of alignment, ISO/IEC 14651 would be “right”, and ISO 12199 would be “wrong”.

## What kind of project would option 2 give?

A total revision of ISO 12199, including the informative parts (and possibly adding more information).

- Ensuring total alignment with ISO/IEC 14651.
- Adding character set information and alphabetical ordering information for a large number of languages.
- More or less continuous maintenance in parallel with ISO/IEC 14651 and the Unicode CLDR.

## What kind of document would option 3 give?

Informative parts of ISO 12199: current annexes B through F.

- It would be possible to maintain that as a Technical Report.
- Maintenance work would be almost as extensive as for a total revision.
- The Technical Report could be a useful reference document for lexicography and terminology.





# Thank you.

Making lives *easier*, *safer* and *better*.

Håvard Hjulstad

Standards Norway

[hhj@standard.no](mailto:hhj@standard.no) & [havard@hjulstad.com](mailto:havard@hjulstad.com)

[iso.org](https://www.iso.org)