



National Standards Authority of Ireland

IRISH STANDARD

I.S. CR 14400:2002

ICS 01.080.01

**EUROPEAN ORDERING RULES - ORDERING
FOR LATIN, GREEK, CYRILLIC, GEORGIAN
AND ARMENIAN SCRIPTS**

National Standards
Authority of Ireland
Dublin 9
Ireland

Tel: (01) 807 3800
Tel: (01) 807 3838

*This Irish Standard was
published under the
authority of the National
Standards Authority of
Ireland
and comes into effect on:
March 1, 2002*

**NO COPYING WITHOUT NSAI
PERMISSION EXCEPT AS
PERMITTED BY COPYRIGHT
LAW**

© NSAI 2002

Price Code J

Údarás um Chaighdeán Náisiúnta na hÉireann

CEN REPORT

CR 14400

RAPPORT CEN

CEN BERICHT

December 2001

ICS

English version

European ordering rules - Ordering for Latin, Greek, Cyrillic, Georgian and Armenian scripts

This CEN Report was approved by CEN on 12 December 2001. It has been drawn up by the Technical Committee CEN/TC 304.

CEN members are the national standards bodies of Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and United Kingdom.



EUROPEAN COMMITTEE FOR STANDARDIZATION
COMITÉ EUROPÉEN DE NORMALISATION
EUROPAISCHES KOMITEE FÜR NORMUNG

Management Centre: rue de Stassart, 36 B-1050 Brussels

CR 14400:2001 (E)

Foreword

This CEN report is intended to facilitate cross border communications and data exchange and to ensure that European cultural requirements are safeguarded in the increasingly interconnected world of today. It provides rules for ordering multilingual European texts and data into a single sequence. These rules come into effect if data from different languages must be brought into a predictable order that makes it easy for users to find information, which is often the case in pan-European applications.

This CEN report extends the repertoire which is specified in ENV 13710:2000 *European Ordering Rules – Ordering of characters from the Latin, Greek and Cyrillic scripts*.

This CEN report does not intend to influence, let alone substitute itself for, national standards or customs in this field. Nevertheless, national standards have the opportunity to adapt this CEN report by declaring a formalized set of deviation rules ("delta") if they so wish.

Sorting assists users by presenting information in a structured way. This may include the subdivision of information by subject matters, e. g. by having several registers in a book, by splitting a phone book into several sections, one for each town that falls into its purview or by having multiple indices in a library. *Ordering* — the arrangement of information in alphabetical sequence — is in most circumstances an integral part of this procedure.

This CEN report must cater for two mutually exclusive demands: Implementers need clear guidelines and data which can readily be used in existing and future ordering applications. This can best be done by defining a European default ordering table in the syntax of the ordering standard ISO/IEC 14651:2001, of which the present document is a "profile". Users with no specific IT-background, however, need an explanation of the principles in a form more in line with existing national ordering standards or relevant practice. As tailoring tables in the syntax of ISO/IEC 14651 can be difficult to read for human readers, an explanation of the principles behind that table is given in the informative annexes. They are written in a more general style and users not familiar with the formal syntax of the tailoring table are advised to consult those annexes first. A web site on this subject is hosted by the Icelandic Standards Organization STADLAR for further reference.¹

1 Scope

This CEN report specifies the sequence to be established by alphabetical ordering of multilingual data composed of characters comprised in the *Multilingual European Subset Number 3* or subsets thereof. This collection is defined in CWA 13783.

NOTE The *Multilingual European Subset Number 3* is usually termed MES-3. A predecessor was known as the *Extended European Subset* (EES). Cf. ENV 1973:1995. MES-3 covers the Latin, Greek, Cyrillic, Armenian, and Georgian letters needed in European data interchange as well as symbols which are needed in Europe. MES-3 comes in two versions: MES-3A is an open collection whereas the fixed collection MES-3B is a snapshot of MES-3A against the repertoire of ISO/IEC 10646-1:1993 with amendments 1 to 31. A CEN workshop agreement on the *Multilingual European Subsets of ISO/IEC 10646* has been published as CEN ISS CWA 13873.

The ordering rules given here are only intended for data in more than one European language. They are not meant to influence, let alone replace existing national standards or practices.

The main part of this CEN report specifies letter-by-letter ordering of character strings. Informative Annex A presents equivalent information in a more readily accessible way. Informative Annex B deals with word-by-word ordering as a special form of ordering with multiple keys. Informative Annex C explains the use of further ordering

¹At present STADLAR can be accessed under <http://www.stadlar.is>

criteria. Informative Annex D presents a widely used alternative to the main part, namely the amalgamation of several scripts in one index via implicit transliteration. Informative Annex F, finally, presents the information inherent in section 6 of the body of this CEN report in a formally equivalent, though condensed, form.

Following the practice of ISO/IEC 14651 characters are referenced as UXXXX where *X* stands for any hexadecimal digit and refers to the value of that character in ISO/IEC 10646-1:2000. This convention is used throughout this CEN report.

2 Normative references

This CEN report incorporates by dated or undated reference provisions from other publications. These normative references are quoted at the appropriate places in the text, and the publications are listed hereafter.

All standards are subject to revision. Dated references do not always refer to subsequent amendments of the publication in question. Undated references always refer to the latest edition.

ISO/IEC 10646-1:2000, Information Technology — Universal Multi-Octet Coded Character set (UCS). Second edition.

ISO 12199:2000, Alphabetical ordering of multilingual terminological and lexicographical data represented in the Latin alphabet.

ISO/IEC 14651:2001, International string ordering and comparison — Method for comparing character strings and description of the common template tailorable ordering.

NOTE An amendment to ISO/IEC 14651 is currently under preparation. It will extend the repertoire which is covered in the common template table of ISO/IEC 14651 to the full repertoire of ISO/IEC 10646-1:2000. It is expected that the table of the amendment will be a true superset of the table in ISO/IEC 14651:2001.

ENV 13710:2000, European Ordering Rules – Ordering of characters from the Latin, Greek and Cyrillic scripts.

3 Definitions

For the purpose of this CEN report the following definitions of ISO/IEC 10646-1 and of ISO/IEC 14651 apply:

3.1

character

A member of a set of elements used for the organization, control, or representation of data. [ISO/IEC 10646-1]

NOTE For the purpose of this CEN report a character is always a member of the MES-3.

3.2

character string

A sequence of characters. [ISO/IEC 14651]

3.3

delta

Differences from a given collation table. The given collation table, together with a given delta, forms a new collation table. Unless otherwise specified in this CEN report, the term “delta” always refers to differences from the Common Template Table as defined in ISO/IEC 14651. [ISO/IEC 14651]

CR 14400:2001 (E)**3.4****ordering**

A process by which two strings are determined to be in exactly one of the relationships of less than, greater than or equal to another. [ISO/IEC 14651]

4 Conformance

In order to be conformant to this CEN report an application shall meet the requirements prescribed in section 6 of ISO/IEC 14651 and use the default table of section 6 or an equivalent description of the information contained therein.

5 Tailorability

The European Ordering Rules defined in this CEN report can be taken as a template which can be tailored to the needs of any European country in the manner specified by ISO/IEC 14651.

6 Default Table

NOTE For the syntax of the table please consult ISO/IEC 14651:2001.'

NOTE The repertoire on which this delta table is based is the intersection of MES-3 and the repertoire of table 1 of ISO/IEC 14651:2001 with the addition of the following characters which are already in MES-2 and ENV 13710:2000:

- the modifier letter double apostrophe (U02EE);
- the Greek small letters digamma (U03DD), stigma (U03DB), koppa (U03DF) and sampi (U03E1);
- the Greek kai symbol (U03D7);
- the Cyrillic letters IE with grave (U0400, U0450) and I with grave (U040D, U045D).

```
%% EOR's EORDeltaTable
%
%% European Ordering Rules.
%
% EOR delta for MES-3 from ISO/IEC 14651:2000's CTT (ISO14651_2000_TABLE1).
%
% This delta gives only the actual changes from the first edition of the CTT.
%
```

```
reorder-after <BASE> % Introduce the LIG weight.
collating-symbol <LIG>
<BASE>
<LIG>
```

```
reorder-after <VRNT3> %Introduce more variants
collating-symbol <VRNT4>
collating-symbol <VRNT5>
collating-symbol <VRNT6>
<VRNT4>
<VRNT5>
<VRNT6>
```

```
collating-element <U000D_U000A> from "<U000D><U000A>"
```

```
reorder-after <SFFFF> % The only place where we can put the order_start line.
```

This is a free preview. Purchase the entire publication at the link below:

[Product Page](#)

-
- Looking for additional Standards? Visit Intertek Inform Infostore
 - Learn about LexConnect, All Jurisdictions, Standards referenced in Australian legislation
-