

English version

European character repertoires and their coding - 8-bit single-byte coding

This Technical Specification (CEN/TS) was approved by CEN on 16 October 2002 for provisional application.

The period of validity of this CEN/TS is limited initially to three years. After two years the members of CEN will be requested to submit their comments, particularly on the question whether the CEN/TS can be converted into a European Standard.

CEN members are required to announce the existence of this CEN/TS in the same way as for an EN and to make the CEN/TS available. It is permissible to keep conflicting national standards in force (in parallel to the CEN/TS) until the final decision about the possible conversion of the CEN/TS into an EN is reached.

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



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

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

Contents

Foreword.....	3
1 Scope	4
2 Normative references	4
3 Terms and definitions.....	4
4 Conformance	5
4.1 Conformance for information interchange.....	5
4.2 Conformance of devices	5
4.2.1 General.....	5
4.2.2 Device description	5
4.2.3 Originating devices.....	5
4.2.4 Receiving devices.....	5
5 Scenario description	5
5.1 Repertoires	5
5.2 Combinations of repertoires and their coding.....	5
6 Repertoire descriptions.....	6
6.1 Latin script.....	6
6.2 Greek script	6
6.3 Cyrillic script	6
6.4 The symbols repertoire	6
7 Coding methods applicable	7
7.1 8-bit single-byte coding.....	7
7.2 Formation of G-sets	7
7.2.1 Invariant-Latin repertoire	7
7.2.2 Initial-Latin repertoire	7
7.2.3 Basic-Latin-a repertoire.....	7
7.2.4 Basic-Latin-b repertoire	7
7.2.5 Basic-Latin-c repertoire.....	7
7.2.6 Large-Latin-8-a repertoire	8
7.2.7 Large-Latin-8-b repertoire	8
7.2.8 Celtic repertoire.....	8
7.2.9 Romanian repertoire.....	8
7.2.10 Basic-Greek repertoire	8
7.2.11 Basic-Cyrillic repertoire	8
7.2.12 Symbols repertoire	8
8 Identification of options	8
Annex A (informative) Specifications of referenced ISO-IR code tables.....	10
Annex B (informative) CEN/TS 1923 options compared to ISO/IEC 7/8-bit standards.....	21
Annex C (informative) Code table illustrations	22

Foreword

This document (CEN/TS 1923:2003) has been prepared by Technical Committee CEN/TC 304, "Information and communications technology - European localization requirements", the secretariat of which is held by SIS.

According to the CEN/CENELEC Internal Regulations, the national standards organizations of the following countries are bound to announce this European Standard: Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Luxembourg, Malta, Netherlands, Norway, Portugal, Slovakia, Spain, Sweden, Switzerland and the United Kingdom.

This Technical Specification is a revision of the European Standard EN 1923:1998, which it cancels and replaces. The main purpose of the revision is to include, and thereby to publicize the availability of, 8-bit code tables developed after the publication of EN 1923:1998; in particular the code table of ISO/IEC 8859-15 and the tables of other additions to the ISO/IEC 8859 series. Although CEN/TC 304 decided that a revision of the contents of EN 1923:1998 was necessary, some uncertainty existed whether the standard as such is needed by the data community in the present-day direction towards multi-octet coding schemes. The committee therefore decided to classify the revised document as a Technical Specification. Its usefulness will thereby become evaluated.

The contents of this document differs from that of EN 1923:1998 in the following respects:

- Extensive editorial changes have been made to the text for conformance with present CEN/CENELEC drafting rules.
- Additional coding scheme options have been introduced, corresponding to ISO/IEC 8859 parts 14, 15 and 16 (Latin-8, Latin-9 and Latin-10), and also to ISO-IR 204 ("Latin-1 alternative with Euro").
- For consistency, the definitions of all options now refer to registrations according to ISO 2375:1985 in the ISO "International register of coded character sets to be used with escape sequences". Relationships to ISO/IEC 10646-1:2000 specifications are also given, to the extent applicable.
- An informative Annex A has been added, containing ISO/IEC 10646-1:2000 identifications for all characters in the options character sets.
- An informative Annex B has been added, listing relationships to ISO/IEC 7/8-bit coding standards.
- An informative Annex C has been added, illustrating the code tables for all options.

1 Scope

This Technical Specification specifies the graphic character repertoires and their single-byte coding, which are available for use for information interchange between information processing systems and for use within such systems, in the scripts that are commonly used by the members of CEN/CENELEC and the Institutions of the European Union and the European Free Trade Association.

This Technical Specification does not specify the interchange of information using a telematic service. The character repertoire and the coding used by a telematic service are defined by the specification of that service. The transmission of information based on the specifications of this Technical Specification using a telematic service may necessitate an adaptation of the number of characters of a repertoire (repertoire transformation function) or a change to the coding (code transformation function).

2 Normative references

This Technical Specification incorporates by dated or undated reference, provisions from other publications. These normative references are cited at the appropriate places in the text and the publications are listed hereafter. For dated references, subsequent amendments to or revisions of any of these publications apply to this Technical Specification only when incorporated in it by amendment or revision. For undated references the latest edition of the publication referred to applies.

ISO/IEC 2022:1994, *Information technology – Character code structure and extension techniques*.

ISO 2375:1985, *Data processing – Procedure for registration of escape sequences*

ISO/IEC 4873:1991, *Information technology – ISO 8-bit code for information interchange – Structure and rules for implementation*.

3 Terms and definitions

For the purposes of this Technical Specification, the following terms and definitions apply:

3.1 bit combination

ordered set of bits used for the representation of characters

3.2

byte

bit string that is operated upon as a unit

3.3

character

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

3.4

coded-character-data-element CC-data-element

element of interchanged information that is specified to consist of a sequence of coded representations of characters, in accordance with one or more identified standards for coded character sets

3.5

coded character set code

set of unambiguous rules that establishes a character set and the one-to-one relationship between the characters of the set and their bit combinations

3.6

code extension

techniques for the encoding of characters that are not included in the character set of a given code

3.7

code table

table showing the characters allocated to each bit combination in a code

3.8

control character

control function the coded representation of which consists of a single bit combination

3.9

control function

action that affects the recording, processing, transmission or interpretation of data, and that has a coded representation consisting of one or more bit combinations

3.10

to designate

to identify a set of characters that are to be represented, in some cases immediately and in others on the occurrence of a further control function, in a prescribed manner

3.11

device

component of information processing equipment which can transmit and/or receive coded information within CC-data-elements; it may be an input/output device in the conventional sense, or a process such as an application program or gateway function