BRITISH STANDARD

BS EN
1923:1998

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

The European Standard EN 1923:1998 has the status of a British Standard

ICS 35.040

National foreword

This British Standard is the English language version of EN 1923:1998. It supersedes DD ENV 41503:1991, DD ENV 41505:1991 and DD ENV 41508:1991, which have been withdrawn.

The UK participation in its preparation was entrusted to Technical Committee IST/2, Character sets and information coding, which has the responsibility to:

- aid enquirers to understand the text;
- present to the responsible European committee any enquiries on the interpretation, or proposals for change, and keep the UK interests informed;
- monitor related international and European developments and promulgate them in the UK.

A list of organizations represented on this committee can be obtained on request to its secretary.

Cross-references

The British Standards which implement international or European publications referred to in this document may be found in the BSI Standards Catalogue under the section entitled "International Standards Correspondence Index", or by using the "Find" facility of the BSI Standards Electronic Catalogue.

A British Standard does not purport to include all the necessary provisions of a contract. Users of British Standards are responsible for their correct application.

Compliance with a British Standard does not of itself confer immunity from legal obligations.

Summary of pages

This document comprises a front cover, an inside front cover, the EN title page, pages 2 to 6, an inside back cover and a back cover.

This British Standard, having been prepared under the direction of the DISC Board, was published under the authority of the Standards Committee and comes into effect on 15 November 1998

Amendments issued since publication

Amd. No.	Date	Text affected

ISBN 0 580 30051 X

EUROPEAN STANDARD

NORME EUROPÉENNE

EUROPÄISCHE NORM

EN 1923

April 1998

ICS 35.040

Supersedes ENV 41503, ENV 41505, ENV 41508

Descriptors: data processing, information interchange, data transmission, character sets, coded character sets, graphic characters, directories, codification

English version

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

Europäische Zeichenvorräte und deren Codierungen — 8-Bit-Einzelbyte-Codierung

This European Standard was approved by CEN on 3 April 1998.

CEN members are bound to comply with the CEN/CENELEC Internal Regulations which stipulate the conditions for giving this European Standard the status of a national standard without any alteration. Up-to-date lists and bibliographical references concerning such national standards may be obtained on application to the Central Secretariat or to any CEN member.

This European Standard exists in three official versions (English, French, German). A version in any other language made by translation under the responsibility of a CEN member into its own language and notified to the Central Secretariat has the same status as the official versions.

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.

CEN

European Committee for Standardization
Comité Européen de Normalisation
Europäisches Komitee für Normung

Central Secretariat: rue de Stassart 36, B-1050 Brussels

Ref. No. EN 1923:1998 E

Foreword

This European Standard has been prepared by Technical Committee CEN/TC 304, Character set technology, the Secretariat of which is held by STRI.

This European Standard replaces ENV 41503, ENV 41505, ENV 41508 (drawn up by CEN/CENELEC/IT/WG-CSC).

This European Standard shall be given the status of a national standard, either by publication of an identical text or by endorsement, at the latest by October 1998, and conflicting national standards shall be withdrawn at the latest by October 1998.

This European Standard differs from the earlier version of ENV 41503 of December 1990 in the following main aspects.

— The base standard for the repertoires of this EN is now ISO/IEC 10646-1 (in place of ISO 646 and the parts of ISO 8859). The coding is based on the latest edition of ISO/IEC 4873.

— There are more combinations of character repertoires and only one coding method available in this European Standard.

— The symbols repertoire has been added to meet requirements expressed by users.

— The coding method of ISO 6937 is now excluded.

The standard is only available in English and German.

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

Contents

1 Scope

This European Standard 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 European Standard 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 European Standard 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 European Standard 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 European Standard only when incorporated 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 4873:1994, *Information technology — ISO 8-bit code for information interchange — Structure and rules for implementation.*

ISO/IEC 10367:1990, *Information technology — Standardized coded graphic character sets for use in 8-bit codes.*

ISO/IEC 10646-1:1993, *Information technology — Universal multiple-octet coded character set (UCS) — Part 1: Architecture and basic multilingual plane.*

3 Definitions

For the purposes of this standard, the definitions of ISO/IEC 10646-1:1993 and the following definitions apply.

3.1

CC-data-element

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

3.2

device

a component of information processing equipment which can transmit, and/or can 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)

3.3

user

a person or other entity that invokes the services provided by a device

(this entity may be a process such as an application program if the “device” is a code converter or a gateway function, for example)

3.4

G-set

the same as “coded graphic character set” in ISO/IEC 2022:1994

4 Abbreviations

The notation used for the character repertoires in clause **7** is as follows.

4.1

BMP

stands for “basic multilingual plane”, as defined in ISO/IEC 10646-1:1993

4.2

Rowxy

refers to Row xy of ISO/IEC 10646-1:1993

4.3

Tablexy

refers to Table xy of ISO/IEC 10646-1:1993

4.4

Positionab-to-cd

refers to a range of code positions from ab to cd (hex format) within the Table xy

5 Scenario description

5.1 Repertoires

There are four collections of graphic characters identified in this European Standard, comprising the characters needed for the:

— Latin script;

— Greek script;

— Cyrillic script;

— symbols repertoire.

These collections are further divided into repertoires as described in clause **7**.

5.2 Combinations of repertoires and their coding

This European Standard identifies combinations of character repertoires and their coding as options. An option identified in this European Standard defines only the minimum requirements, in terms of character repertoire and coding, applied to a conforming device. Additional capabilities of the originating or receiving device may be used, during the information interchange, subject to bilateral agreement.

8-bit single-byte coding shall be a version of ISO/IEC 4873:1994, clause **9**.

NOTE This European Standard is intended to be used with other standards specifying control functions, as needed by the base coding standards.

6 Conformance

6.1 Conformance for information interchange

A CC-data-element within coded information for interchange is in conformance with this standard if all the coded representations of graphic characters within that CC-data-element conform to the requirements of clauses **7** and **8**.

A claim of conformance shall identify the option adopted according to clause **9**.

6.2 Conformance of devices

6.2.1 *General*

A device is in conformance with this standard if it conforms to the requirements of **6.2.2**, and either or both of **6.2.3** and **6.2.4**. A claim of conformance shall identify the document which contains the description specified in **6.2.2**, and shall identify the option adopted.

6.2.2 *Device description*

A device that conforms to this standard shall be the subject of a description that identifies the means by which the user may supply characters to the device, or may recognize them when they are made available to him, as specified respectively in **6.2.3** and **6.2.4**.

6.2.3 *Originating devices*

An originating device shall allow its user to supply any sequence of graphic characters from the option adopted, and shall be capable of transmitting their coded representations within a CC-data-element.

6.2.4 *Receiving devices*

A receiving device shall be capable of receiving and interpreting any coded representations of graphic characters that are within a CC-data-element, and that conform to **6.1**, and shall make the corresponding characters available to the user in such a way that the user can identify them from among those conforming to the option adopted, and can distinguish them from each other.

7 Repertoire description

7.1 Latin script

Four subsets of this collection of graphic characters are identified, each with a subset/superset relation with the others. These subsets are the following.

7.1.1 The Invariant-Latin repertoire, containing 83 characters as specified in the BMP-Row00-Table01-Position20-to-22, 25-to-3F, 41-to-5A, 5F, 61-to-7A of ISO/IEC 10646-1:1993 (**Repertoire IVL**).

7.1.2 The Initial-Latin repertoire, containing 95 characters as specified in the BMP-Row00-Table01 of ISO/IEC 10646-1:1993 (**Repertoire IL**). It is a true superset of the Invariant-Latin repertoire (**Repertoire IVL**).

7.1.3 The Basic-Latin repertoire, comprising the Initial repertoire plus the repertoire of Latin-1 Supplement as specified in the BMP-Row00-Table02 of ISO/IEC 10646-1:1993. It is a true superset of the Initial repertoire (**Repertoire BL**).

7.1.4 The Large-Latin-8 repertoire for the 8-bit environment, comprising the union of the Basic-Latin repertoire with the repertoire consisting of the Latin characters coded in ISO/IEC 10367:1990. It is a true superset of the Basic-Latin repertoire (**Repertoire LL8**).

7.2 Greek script

In the 8-bit environment, only one Greek repertoire is defined, which is:

7.2.1 The Basic-Greek repertoire, comprising the characters defined in the BMP-Row03-Table09 of ISO/IEC 10646-1:1993 (**Repertoire BG**).

7.3 Cyrillic script

In the 8-bit environment, only one Cyrillic repertoire is defined, which is:

7.3.1 The Basic-Cyrillic repertoire, comprising the characters defined in the BMP-Row04-Table11-Position01-to-5F of ISO/IEC 10646-1:1993 (**Repertoire BC**).

7.4 The symbols repertoire

This repertoire shall comprise the characters defined in Registration 155 of ISO 2375:1985. These characters are derived from BMP-Row25-Table45 and BMP-Row25-Table46 of ISO/IEC 10646-1:1993 (**Repertoire BS**).

8 Coding methods applicable

8.1 Eight-bit single-byte coding

8.1.1 Each character shall be coded by the use of a single byte. No control function shall be used that would cause characters within a repertoire to be combined to represent any other character.

8.1.2 The various repertoires shall form G-sets, according to the relevant provisions of ISO/IEC 2022:1994.

8.1.3 When code extension techniques are applied, then the provisions of ISO/IEC 4873:1994 shall be followed. The application should always conform to a certain level of ISO/IEC 4873:1994.

8.1.4 When code extension techniques are applied, then all the necessary control functions shall exist, coded as specified in ISO/IEC 4873:1994.

8.2 Formation of G-sets

The characters belonging to the repertoires defined in clause **7** shall be arranged to the code table positions and shall form G-sets as follows.

8.2.1 The **IVL** repertoire shall always form a G0 code element in a version of ISO/IEC 4873:1994.

The characters shall be arranged in the code table as specified in BMP-Row00-Table01-Position20-to-22, 25-to-3F, 41-to-5A, 5F, 61-to-7A of ISO/IEC 10646-1:1993. The Row octet will be omitted and each character will be coded by the use of the Cell octet only.

The escape sequence to designate this set will be:

ESC 02/08 02/01 04/02.

8.2.2 The **IL** repertoire shall always form a G0 code element in a version of ISO/IEC 4873:1994.

The characters shall be arranged in the code table as specified in BMP-Row00-Table01 of ISO/IEC 10646-1:1993. The Row octet will be omitted and each character will be coded by the use of the Cell octet only.

The escape sequence to designate this set will be:

ESC 02/08 04/02.

8.2.3 The **BL** repertoire shall form two G-sets in a version of ISO/IEC 4873:1994.

One G-set shall contain the IL repertoire and shall be coded according to **8.2.2**.

The Latin-1 Supplement repertoire shall form either a G1 or a G2 or a G3 set in a version of ISO/IEC 4873:1994. The characters shall be arranged in the code table as specified in BMP-Row00-Table02 of ISO/IEC 10646-1:1993. The Row octet will be omitted and each character will be coded by the use of the Cell octet only.

The escape sequences to designate this set will be:

ESC 02/13 04/01 as G1
ESC 02/14 04/01 as G2
ESC 02/15 04/01 as G3

8.2.4 The **LL8** repertoire shall form four G-sets in a version of ISO/IEC 4873:1994.

Two G-sets will contain the BL repertoire and shall be coded according to **8.2.3**. The rest of the repertoire shall be arranged in code table positions as in ISO 2375:1985 registrations 101 and 154, thus forming two G-sets that can be used as G1 or G2 or G3 sets in a version of ISO/IEC 4873:1994. All the additional characters contained in ISO 2375:1985 registrations 101 and 154 shall be retained.

The escape sequences to designate these sets will be:

for registration 101	ESC 02/13 04/02 as G1
	ESC 02/14 04/02 as G2
	ESC 02/15 04/02 as G3
for registration 154	ESC 02/13 05/00 as G1
	ESC 02/14 05/00 as G2
	ESC 02/15 05/00 as G3

8.2.5 The **BG** repertoire shall form one G-set in a version of ISO/IEC 4873:1994.

The repertoire shall be arranged in code table positions as in BMP-Row03-Table09 of ISO/IEC 10646-1:1993, as a G1 or G2 or G3 set. All the additional characters contained in ISO 2375:1985 registration 126 shall be retained.

The escape sequences to designate this set will be:

ESC 02/13 04/06 as G1
ESC 02/14 04/06 as G2
ESC 02/15 04/06 as G3

8.2.6 The **BC** repertoire shall form one G-set in a version of ISO/IEC 4873:1994.

The repertoire shall be arranged in code table positions as in BMP-Row04-Table11 of ISO/IEC 10646-1:1993, as a G1 or G2 or G3 set.

All the additional characters contained in ISO 2375:1985 registration 144 shall be retained.

The escape sequences to designate this set will be:

ESC 02/13 04/12 as G1
ESC 02/14 04/12 as G2
ESC 02/15 04/12 as G3

8.2.7 The **BS** repertoire shall form one G-set in a version of ISO/IEC 4873:1994.

The repertoire shall be arranged in code table positions as in ISO 2375:1985 registration 155, as a G1 or G2 or G3 code element.

The escape sequences to designate this set will be:

ESC 02/13 05/01 as G1
ESC 02/14 05/01 as G2
ESC 02/15 05/01 as G3

9 Identification of options

If a reference to this European Standard is made in another European Standard, the option adopted shall be clearly identified.

Table 1 summarizes the options that conform to the requirements of this European Standard.

NOTE Announcement of the version that is currently in use should always be done according to the provisions laid down in clause **10** of ISO/IEC 4873:1994.

Table 1 — 8-bit coding (ISO/IEC 4873)

Option	Repertoire	G-set used	ISO/IEC-4873 level
A	IVL	G0	1 or 2 or 3
B	IL	G0	1 or 2 or 3
C	BL	G0/G1	1 or 2 or 3
D	LL8	G0/G1/G2/G3	2 or 3
E	BG	G1/G2/G3	1 or 2 or 3
F	BC	G1/G2/G3	1 or 2 or 3
G	BS	G1/G2/G3	1 or 2 or 3
BE	IL+BG	G0/G1	1 or 2 or 3
CE	BL+BG	G0/G1/G2	2 or 3
BEG	IL+BG+BS	G0/G1/G2	2 or 3
CEG	BL+BG+BS	G0/G1/G2/G3	2 or 3
BF	IL+BC	G0/G1	1 or 2 or 3
CF	BL+BC	G0/G1/G2	2 or 3
BFG	IL+BC+BS	G0/G1/G2	2 or 3
CFG	BL+BC+BS	G0/G1/G2/G3	2 or 3
BEF	IL+BG+BC	G0/G1/G2	2 or 3
BEFG	IL+BG+BC+BS	G0/G1/G2/G3	2 or 3
CEF	BL+BG+BC	G0/G1/G2/G3	2 or 3

blank

BSI — British Standards Institution

BSI is the independent national body responsible for preparing British Standards. It presents the UK view on standards in Europe and at the international level. It is incorporated by Royal Charter.

Revisions

British Standards are updated by amendment or revision. Users of British Standards should make sure that they possess the latest amendments or editions.

It is the constant aim of BSI to improve the quality of our products and services. We would be grateful if anyone finding an inaccuracy or ambiguity while using this British Standard would inform the Secretary of the technical committee responsible, the identity of which can be found on the inside front cover. Tel: 020 8996 9000. Fax: 020 8996 7400.

BSI offers members an individual updating service called PLUS which ensures that subscribers automatically receive the latest editions of standards.

Buying standards

Orders for all BSI, international and foreign standards publications should be addressed to Customer Services. Tel: 020 8996 9001. Fax: 020 8996 7001.

In response to orders for international standards, it is BSI policy to supply the BSI implementation of those that have been published as British Standards, unless otherwise requested.

Information on standards

BSI provides a wide range of information on national, European and international standards through its Library and its Technical Help to Exporters Service. Various BSI electronic information services are also available which give details on all its products and services. Contact the Information Centre. Tel: 020 8996 7111. Fax: 020 8996 7048.

Subscribing members of BSI are kept up to date with standards developments and receive substantial discounts on the purchase price of standards. For details of these and other benefits contact Membership Administration. Tel: 020 8996 7002. Fax: 020 8996 7001.

Copyright

BSI
389 Chiswick High Road
London
W4 4AL