

English version

**Information technology –  
Multilingual European Subsets in ISO/IEC 10646-1**

**Technologies de l'information –  
Jeux partiels européens multilingues  
dans l'ISO/CEI 10646-1**

**Informationstechnologie –  
Mehrsprachige europäische Untermengen  
in ISO/IEC 10646-1**

This CEN Workshop Agreement has been drafted and approved by a Workshop of representatives of interested parties, whose names and affiliations can be obtained from the CEN/ISSS Secretariat.

The formal process followed by the Workshop in the development of this Workshop Agreement has been endorsed by the National Members of CEN, but neither the National Members of CEN nor the CEN Central Secretariat can be held accountable for the technical content of this CEN Workshop Agreement or for possible conflicts with standards or legislation.

This CEN Workshop Agreement can in no way be held as being an official standard developed by CEN and its Members.

This CEN Workshop Agreement is publicly available, as a reference document, from the CEN Members National Standard Bodies.

CEN Members are the National Standards Bodies of Austria, Belgium, the Czech Republic, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom.

CEN

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

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

## Contents

<b>Foreword</b>	<b>3</b>
<b>Introduction</b>	<b>4</b>
<b>1. Scope</b>	<b>5</b>
<b>2. Normative references</b>	<b>7</b>
<b>3. Definitions and abbreviations</b>	<b>7</b>
<b>4. The Multilingual European Subset No. 1 (MES-1)</b>	<b>9</b>
<b>5. The Multilingual European Subset No. 2 (MES-2)</b>	<b>10</b>
<b>6. The Multilingual European Subset No. 3A (MES-3A)</b>	<b>11</b>
<b>7. The Multilingual European Subset No. 3B (MES-3B)</b>	<b>12</b>
<b>Annex A (Informative) Coded character set repertoires covered by MES-2</b>	<b>14</b>
<b>Annex B (Informative) List of languages covered by MES-1</b>	<b>16</b>
<b>Annex C (Informative) List of languages covered by MES-2</b>	<b>16</b>
<b>Annex D (Informative) Rationales for inclusion of characters in this CWA</b>	<b>17</b>
<b>Annex E (Informative) Bibliography</b>	<b>21</b>
<b>Annex F (Informative) Comments accompanying Negative ballots</b>	<b>23</b>

## **Foreword**

The production of this document which describes a number of “Multilingual European Subsets in ISO/IEC 10646-1” was agreed by the CEN/ISSS Workshop on Multilingual European Subsets (WS-MES) in the Workshop’s Kick-Off Meeting on 1998-06-11.

The document has been developed through the collaboration of a number of contributing partners in WS-MES. WS-MES representation gathers a wide mix of interests, coming from academia, public administrations, IT-suppliers, and other interested experts. A list of company experts who have supported the document’s contents may be obtained from the CEN/ISSS Secretariat.

It was not possible to produce a document that addressed the concerns of all contributors, as some of these contributions were made when the consensus building process was closed already. In these late contributions, concerns were raised with regard to the cost to Industry to implement the subsets beyond MES-1, and with regard to interoperability problems when implementing the MESes, that would lead to further costs. In addition, it was requested that the entire document be rewritten both for clarity and better human reviewability, as well as consistency of actual use of characters in ISO/IEC 10646, arguing that the present document was not user-friendly for reviewers, implementors, procurers, and interested users.

The final review/endorsement round for this CWA was started on 1999-10-14 and was successfully closed on 1999-12-01. The final text of this CWA was submitted to CEN for publication on 2000-03-02.

## Introduction

There is a growing need for IT-communication across the national boundaries of Europe, with public administrations cooperating in large data systems, and with commerce and trade between countries increasingly using IT-techniques. In addition, legal requirements regarding the spelling of personal names for individuals in Europe must be considered. This leads to extensive requirements on the character set repertoires used in European IT-equipment.

In Europe there is a large number of official languages and officially-recognized indigenous minority languages. These employ a large number of letters. In addition, a large number of other graphic characters is required for day-to-day computer use in Europe.

Prior to the development of the Universal Character Set (UCS), there was no character set standard which included all the characters and scripts needed by Europe. The possibilities of combining several standardized character sets – coded in 7 bits or 8 bits – in the same IT-systems with existing code extension techniques have proven to be impractical and insufficient.

Part 1 of ISO/IEC 10646-1, the BMP, provides a good base for European character coding, since it defines fixed code positions for almost all presently known letters and a very large number of symbols and other characters for non-specialized use.

It is likely that various compression schemes and transformation formats will be used for data storage and transmission of UCS encoded data. Consequently, required storage space and communication time will not double compared with single-octet codes.

Implementation of ISO/IEC 10646-1 characters requires resources for supporting, for example:

- font definitions
- tables for code table translation, ordering, matching, and upper/lower casing functions

The resources needed are material or human, for example:

- system memory space for definitions, tables, and functions
- licence costs for fonts
- documentation
- working time for program development
- learning time
- convenient and standardized keyboard input methods

Some of the implementation problems discussed above could be solved by subsets of the UCS. This work on European subsets of the UCS is particularly aimed at solving the problems of outputting the full character set of the UCS.

It is estimated that implementing the full character set of the UCS may be costly in the first stages of UCS use, and that many manufacturers will implement in subset-stages. To ensure that a common subset usable to the vast majority of European users be available for a reasonable price, and as a guide to manufacturers, it will be helpful to specify, to users and procurers of systems, European subsets of the UCS encompassing the characters for use in European languages as well as other frequently used and specialist characters. Such subsets may also be useful with regard to further standardization work (for example, on ordering specifications), so that the work is reasonably limited and still useful in a European environment.

## 1. Scope

**1.1** This CEN Workshop Agreement specifies the following European repertoires in the form of Subsets of ISO/IEC 10646-1:

- MES-1: A Latin repertoire based on ISO/IEC 6937:1994 (a limited subset, fixed collection);
- MES-2: A Latin, Greek, Cyrillic repertoire based on ENV 1973:1996 (a limited subset, fixed collection);
- MES-3: A repertoire needed to write the languages of Europe and transliterate between them (a selected subset (non-fixed collection) and a limited subset (fixed collection)).

These repertoires are intended for, for example:

- governmental administrations in transborder transactions;
- the European Union and EFTA for use within the Schengen area and within the European Economic Area;
- CEN (the European Committee for Standardization);
- support of the Council of Europe member country ratifications of the European Charter for Regional and Minority Languages.

NOTE: This CWA does *not* specify any encoding of the European Subsets.

Procurement guidance and conformance requirements are outside the scope of this CWA. It should be noted that the subsets defined herein can be used in combination with other collections or characters in ISO/IEC 10646 for specific applications. For example, MES-1 might be combined with UCS Collection 39 (Mathematical Operators) for educational needs.

NOTE: It is the intention of this CWA that, after ratification of this document, CEN TC304 shall submit the subsets to ISO/IEC JTC1/SC2/WG2 as collections to be added to annex A of ISO/IEC 10646-1.

## 1.2 MES-1

### 1.2.1 Coverage

MES-1 is a Latin repertoire of 335 characters, based on ISO/IEC 6937:1994, but identical to the LL8 repertoire of EN 1923:1998 (the two differ in that the latter contains 00D0 LATIN CAPITAL LETTER ETH and the former does not) with the addition of 20AC EURO SIGN. It is believed to cover at least the 44 languages listed in annex B. It is a superset of the Latin 1 through Latin 5 character sets and similar vendor coded character sets (not listed here).

### 1.2.2 Field of application

This repertoire is mainly intended for use by information systems which are restricted by users and/or technology to Latin characters. Example: name registries in countries using the Latin script.

## 1.3 MES-2

### 1.3.1 Coverage

MES-2 is a Latin, Greek, and Cyrillic repertoire of 1062 characters. It is believed to cover at least the 128 languages listed in annex C and is a superset of MES-1, all repertoires of EN 1923:1998, and a large number of commonly used coded character sets (see examples in annex A).

The Cyrillic repertoire includes all Cyrillic characters apart from those of Old Church Slavonic, thus the language coverage is more than the minimum required by the scope, but no particular benefit was seen to be gained by the exclusion of any of those languages.

### 1.3.2 Field of application

This repertoire is intended for:

- general purpose data and text processing applications in typical office environments in Europe;
- governmental administrations in transborder transactions;
- intergovernmental organizations, such as the European Union and EFTA for use within the Schengen area and within the European Economic Area;
- CEN (the European Committee for Standardization);

## 1.4 MES-3

### 1.4.1 Coverage

MES-3 is a repertoire covering all the characters belonging to European scripts. This coverage is guaranteed by defining MES-3A as a script-based and non-fixed collection. It is, therefore, a superset of any Latin, Greek, Cyrillic, Armenian, and Georgian UCS repertoire, including MES-1 and MES-2. The MES-3 includes whole scripts, thus the language coverage is more than the minimum required by the scope, but no particular benefit was seen to be gained by the exclusion of any of those languages.

### 1.4.2 Field of application

In addition to the field of application for MES-2 described in 1.3.2 above, this repertoire is needed in support of the Council of Europe member country ratifications of the European Charter for Regional and Minority Languages. It is intended to be used in information systems which allow writing the languages of Europe and transliterating between them. It also aims to provide for specialist administrative, scientific, technical, educational, and publishing requirements in Europe.

NOTE: This repertoire is based on selected *scripts* of European origin with an inherent left-to-right directionality. Although the repertoire thus includes some characters that are not necessarily used in European languages, it does not include the characters needed for all the indigenous languages of Europe: the Karaim, Ladino, Tati, and Yiddish languages, for instance, are indigenous European languages using the Hebrew script. As this script has right-to-left directionality, complete multilingual support for all the indigenous languages of Europe (either official languages or those with a recognized status as minority languages) would require bi-directional support, which requirement could be seen excessive for the majority of multilingual European IT systems. The requirement to support Yiddish, however, is a real one, not just in principle, but also in a multilingual environment. Collections 12, 13, and 63 in ISO/IEC 10646-1 should be used to provide full support for these languages, though they are not a part of the MES-3 specification itself.

MES-3 does not broadly correspond to any single existing character set: it does correspond to the sum of many existing character sets. It is a collection-based subset of the UCS.

MES-3 will help developers worldwide to implement all characters belonging to European scripts. The requirement for support of European scripts is not intended to imply that European users have no need for non-European scripts; but it is logical to specify a subset defining the collections in ISO/IEC 10646-1 which contain European scripts. Non-European scripts are being provided for by their primary users.

MES-3 is also intended to provide guidance to product developers to facilitate cost-effective provision of fonts for rendering devices, for instance.

## 2. Normative references

This CWA 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 herebelow. For dated references, subsequent amendments to, or revisions of, any of these publications apply to this CWA only when incorporated in it by amendment or revision. For undated references the latest edition of the publication referred to applies.

**EN 1923:1998** *European character repertoires and their coding – 8 bit single byte coding.*

**ENV 1973:1996** *Information technology – European subsets of ISO/IEC 10646-1.*

**ISO/IEC 6937:1994** *Information technology – Coded graphic character set for text communication – Latin alphabet.*

**ISO/IEC 10646-1:1993** (and its **Amendments 1–31**) *Information technology – Universal Multiple-Octet Coded Character Set (UCS) – Part 1: Architecture and Basic Multilingual Plane.*

## 3. Definitions and abbreviations

### 3.1 Definitions

For the purposes of this CEN Workshop Agreement the following basic definitions from ISO/IEC 10646-1 apply:

*character*: A member of a set of elements used for the organization, control, or representation of data.

*coded character*: A character together with its coded representation.

*coded character set*: A set of unambiguous rules that establishes a character set and the relationship between the characters of the set and their coded representation.

*combining character*: A member of an identified subset of the coded character set of ISO/IEC 10646 intended for combination with the preceding non-combining graphic character, or with a sequence of combining characters preceded by a non-combining character.

*limited subset*: A limited subset consists of a list of graphic characters in the specified subset. This specification allows applications and devices that were developed using other codes to interwork with this coded character set. A claim of conformance referring to a limited subset shall list the graphic characters in the subset by the names of graphic characters or code positions as defined in ISO/IEC 10646.

*script*: A set of graphic characters used for the written form of one or more languages.

*selected subset*: A selected subset consists of a list of collections of graphic characters as defined in ISO/IEC 10646. The collections from which the selection may be made are listed in annex A of each part of ISO/IEC 10646. A selected subset shall always automatically include the Cells 20 to

7E of Row 00 of Plane 00 of Group 00. A claim of conformance referring to a selected subset shall list the collections chosen as defined in ISO/IEC 10646.

### **3.2 Abbreviations**

The following abbreviations apply:

*BMP*: Basic Multilingual Plane (Plane 00 of Group 00 of ISO/IEC 10646-1)

*MES-1*: Multilingual European Subset No. 1

*MES-2*: Multilingual European Subset No. 2

*MES-3*: Multilingual European Subset No. 3

NOTE: Clauses 6 and 7 below specify two repertoires based on MES-3. The first, MES-3A, is a non-fixed collection and will allow a user to implement MES-3 dynamically with reference to a particular version of ISO/IEC 10646-1 and its Amendments. The second, MES-3B, is a fixed collection, and is a “snapshot” which can be implemented without reference.

*UCS*: Universal Character Set



## 4. The Multilingual European Subset No. 1 (MES-1)

### 4.1 Specification for the Multilingual European Subset No. 1

Collection Name: MES-1  
Type of Collection (Fixed or Non-fixed): FIXED  
Number of characters 335

Plane 00

#### Rows Positions (Cells)

00	20–7E A0–FF
01	00–13 16–2B 2E–4D 50–7E
02	C7 D8–DB DD
20	15 18–19 1C–1D AC
21	22 26 5B–5E 90–93
26	6A

NOTE: The subset defined here can be used in combination with other collections or characters in ISO/IEC 10646 for specific applications. For example, MES-1 might be combined with UCS Collection 39 (Mathematical Operators) for educational needs.

### 4.2 Characters of the MES-1 contained in UCS collections from annex A of ISO/IEC 10646-1

The characters of MES-1 are contained in the following UCS collections:

1	BASIC LATIN	0020–007E
2	LATIN-1 SUPPLEMENT	00A0–00FF
3	LATIN EXTENDED-A	0100–017F
6	SPACING MODIFIER LETTERS	02B0–02FF
32	GENERAL PUNCTUATION	2000–206F
34	CURRENCY SYMBOLS	20A0–20CF
36	LETTERLIKE SYMBOLS	2100–214F
37	NUMBER FORMS	2150–218F
38	ARROWS	2190–21FF
47	MISCELLANEOUS SYMBOLS	2600–26FF

## 5. The Multilingual European Subset No. 2 (MES-2)

### 5.1 Specification for the Multilingual European Subset No. 2

Collection Name: MES-2  
Type of Collection (Fixed or Non-fixed): FIXED  
Number of characters 1052

Plane 00

#### Rows Positions (Cells)

00	20–7E A0–FF
01	00–7F 8F 92 B7 DE–EF FA–FF
02	18–1B 1E–1F 59 7C 92 BB–BD C6–C7 C9 D8–DD EE
03	74–75 7A 7E 84–8A 8C 8E–A1 A3–CE D7 DA–E1
04	00–5F 90–C4 C7–C8 CB–CC D0–EB EE–F5 F8–F9
1E	02–03 0A–0B 1E–1F 40–41 56–57 60–61 6A–6B 80–85 9B F2–F3
1F	00–15 18–1D 20–45 48–4D 50–57 59 5B 5D 5F–7D 80–B4 B6–C4 C6–D3 D6–DB DD–EF F2–F4 F6–FE
20	13–15 17–1E 20–22 26 30 32–33 39–3A 3C 3E 44 4A 7F 82 A3–A4 A7 AC AF
21	05 16 22 26 5B–5E 90–95 A8
22	00 02–03 06 08–09 0F 11–12 19–1A 1E–1F 27–2B 48 59 60–61 64–65 82–83 95 97
23	02 10 20–21 29–2A
25	00 02 0C 10 14 18 1C 24 2C 34 3C 50–6C 80 84 88 8C 90–93 A0 AC B2 BA BC C4 CA–CB D8–D9
26	3A–3C 40 42 60 63 65–66 6A–6B
FB	01–02
FF	FD

### 5.2 Characters of the MES-2 contained in UCS collections from annex A of ISO/IEC 10646-1

The characters of MES-2 are contained in the following UCS collections:

1	BASIC LATIN	0020–007E
2	LATIN-1 SUPPLEMENT	00A0–00FF
3	LATIN EXTENDED-A	0100–017F
4	LATIN EXTENDED-B	0180–024F
5	IPA EXTENSIONS	0250–02AF
6	SPACING MODIFIER LETTERS	02B0–02FF
8	BASIC GREEK	0370–03CF
9	GREEK SYMBOLS AND COPTIC	03D0–03FF
10	CYRILLIC	0400–04FF
30	LATIN EXTENDED ADDITIONAL	1E00–1EFF
31	GREEK EXTENDED	1F00–1FFF
32	GENERAL PUNCTUATION	2000–206F
33	SUPERSCRIPTS AND SUBSCRIPTS	2070–209F
34	CURRENCY SYMBOLS	20A0–20CF
36	LETTERLIKE SYMBOLS	2100–214F
37	NUMBER FORMS	2150–218F
38	ARROWS	2190–21FF
39	MATHEMATICAL OPERATORS	2200–22FF
40	MISCELLANEOUS TECHNICAL	2300–23FF
44	BOX DRAWING	2500–257F
45	BLOCK ELEMENTS	2580–259F
46	GEOMETRIC SHAPES	25A0–25FF
47	MISCELLANEOUS SYMBOLS	2600–26FF
63	LTR PRESENTATION FORMS	FB00–FB4F
70	SPECIALS	FFF0–FFFD

NOTE: Implementors may wish to exclude the Hebrews characters in the range FB18–FB4F in collection 63. A collection identifier for LTR Presentation Forms has requested from ISO/IEC JTC1/SC2/WG2 which should replace collection 63 in the specification of the MES-2.

## 6. The Multilingual European Subset No. 3A (MES-3A)

### 6.1 Specification for the Multilingual European Subset No. 3A

Collection Name: MES-3A  
Type of Collection (Fixed or Non-fixed): NON-FIXED

The following UCS collections from annex A of ISO/IEC 10646-1 comprise MES-3A:

No.	Collection name	hex range
1	BASIC LATIN	0020–007E
2	LATIN-1 SUPPLEMENT	00A0–00FF
3	LATIN EXTENDED-A	0100–017F
4	LATIN EXTENDED-B	0180–024F
5	IPA EXTENSIONS	0250–02AF
6	SPACING MODIFIER LETTERS	02B0–02FF
7	COMBINING DIACRITICAL MARKS	0300–036F
8	BASIC GREEK	0370–03CF
9	GREEK SYMBOLS AND COPTIC	03D0–03FF
10	CYRILLIC	0400–04FF
11	ARMENIAN	0530–058F
27	BASIC GEORGIAN	10D0–10FF
30	LATIN EXTENDED ADDITIONAL	1E00–1EFF
31	GREEK EXTENDED	1F00–1FFF
32	GENERAL PUNCTUATION	2000–206F
33	SUPERSCRIPTS AND SUBSCRIPTS	2070–209F
34	CURRENCY SYMBOLS	20A0–20CF
35	COMBINING DIACRITICAL MARKS FOR SYMBOLS	20D0–20FF
36	LETTERLIKE SYMBOLS	2100–214F
37	NUMBER FORMS	2150–218F
38	ARROWS	2190–21FF
39	MATHEMATICAL OPERATORS	2200–22FF
40	MISCELLANEOUS TECHNICAL	2300–23FF
42	OPTICAL CHARACTER RECOGNITION	2440–245F
44	BOX DRAWING	2500–257F
45	BLOCK ELEMENTS	2580–259F
46	GEOMETRIC SHAPES	25A0–25FF
47	MISCELLANEOUS SYMBOLS	2600–26FF
63	ALPHABETIC PRESENTATION FORMS	FB00–FB4F
65	COMBINING HALF MARKS	FE20–FE2F
70	SPECIALS	FFFO–FFFD

NOTE: Implementors may wish to exclude the Hebrews characters in the range FB18–FB4F in collection 63. A collection identifier for LTR Presentation Forms has requested from ISO/IEC JTC1/SC2/WG2 which should replace collection 63 in the specification of the MES-3A.

## 7. The Multilingual European Subset No. 3B (MES-3B)

### 7.1 Specification for the Multilingual European Subset No. 3B

Collection Name: MES-3B  
Type of Collection (Fixed or Non-fixed): FIXED  
Number of characters: 2819

Plane 00

#### Rows Positions (Cells)

00	20–7E A0–FF
01	00–FF
02	00–1F 22–33 50–AD B0–EE
03	00–4E 60–62 74–75 7A 7E 84–8A 8C 8E–A1 A3–CE D0–D7 DA–F3
04	00–86 88–89 8C–C4 C7–C8 CB–CC D0–F5 F8–F9
05	31–56 59–5F 61–87 89–8A
10	D0–F6 FB
1E	00–9B A0–F9
1F	00–15 18–1D 20–45 48–4D 50–57 59 5B 5D 5F–7D 80–B4 B6–C4 C6–D3 D6–DB DD–EF F2–F4 F6–FE
20	00–46 48–4D 6A–70 74–8E A0–AF D0–E3
21	00–3A 53–83 90–F3
22	00–F1
23	00–7B 7D–9A
24	40–4A
25	00–95 A0–F7
26	00–13 19–71
FB	00–06 13–17
FE	20–23
FF	F9–FD

### 7.2 Characters of the MES-3B contained in UCS collections from annex A of ISO/IEC 10646-1

The characters of MES-3B are contained in the following UCS collections:

No.	Collection name	hex range
1	BASIC LATIN	0020–007E
2	LATIN-1 SUPPLEMENT	00A0–00FF
3	LATIN EXTENDED-A	0100–017F
4	LATIN EXTENDED-B	0180–024F
5	IPA EXTENSIONS	0250–02AF
6	SPACING MODIFIER LETTERS	02B0–02FF
7	COMBINING DIACRITICAL MARKS	0300–036F
8	BASIC GREEK	0370–03CF
9	GREEK SYMBOLS AND COPTIC	03D0–03FF
10	CYRILLIC	0400–04FF
11	ARMENIAN	0530–058F
27	BASIC GEORGIAN	10D0–10FF
30	LATIN EXTENDED ADDITIONAL	1E00–1EFF
31	GREEK EXTENDED	1F00–1FFF
32	GENERAL PUNCTUATION	2000–206F
33	SUPERSCRIPTS AND SUBSCRIPTS	2070–209F
34	CURRENCY SYMBOLS	20A0–20CF
35	COMBINING DIACRITICAL MARKS FOR SYMBOLS	20D0–20FF
36	LETTERLIKE SYMBOLS	2100–214F
37	NUMBER FORMS	2150–218F
38	ARROWS	2190–21FF
39	MATHEMATICAL OPERATORS	2200–22FF
40	MISCELLANEOUS TECHNICAL	2300–23FF
42	OPTICAL CHARACTER RECOGNITION	2440–245F
44	BOX DRAWING	2500–257F
45	BLOCK ELEMENTS	2580–259F

## 7.2 Characters of the MES-3B contained in UCS collections (*continued*)

46	GEOMETRIC SHAPES	25A0–25FF
47	MISCELLANEOUS SYMBOLS	2600–26FF
63	ALPHABETIC PRESENTATION FORMS	FB00–FB4F
65	COMBINING HALF MARKS	FE20–FE2F
70	SPECIALS	FFF0–FFFD

NOTE: Implementors may wish to exclude the Hebrews characters in the range FB18–FB4F in collection 63. A collection identifier for LTR Presentation Forms has requested from ISO/IEC JTC1/SC2/WG2 which should replace collection 63 in the specification of the MES-3B.

## Annex A. Coded character set repertoires covered by MES-2 (Informative)

The Multilingual European Subset No. 2 covers at least the repertoires of the following coded character sets.

### A.1 Coded character sets in the ISO 2375 IR

The Multilingual European Subset No. 2 includes the repertoires of registered coded character sets listed in the following table:

IR number	IR name	Comment
ISO-IR 6	Basic G0 Set	ASCII
ISO-IR 100	Latin alphabet No. 1, supplementary set	Latin-1
ISO-IR 101	Latin alphabet No. 2, supplementary set	Latin-2 (most of Central Europe)
ISO-IR 109	Latin alphabet No. 3, supplementary set	
ISO-IR 110	Latin alphabet No. 4, supplementary set	Nordic (not in use)
ISO-IR 126	Greek supplementary set	
ISO-IR 144	Cyrillic supplementary set	
ISO-IR 148	Latin alphabet No. 5, supplementary set	
ISO-IR 154	Supplementary set for Latin alphabets No. 1 or No. 5, and No. 2	
ISO-IR 156	Supplementary set of ISO/IEC 6937	
ISO-IR 157	Latin alphabet No. 6, supplementary set	Nordic (not in use)
ISO-IR 158	Sámi supplementary set	Sámi and Old Icelandic
ISO-IR 179	Baltic Rim supplementary set	Latin-7 (Baltic Rim)
ISO-IR 199	Latin alphabet No. 8 (Celtic)	Latin-8 (Celtic)
ISO-IR 203	Latin alphabet No. 9	Latin-9

### A.2 ISO standards

By including these repertoires, the subset covers all characters coded in the following ISO standards:

- ISO/IEC 8859 parts 1, 2, 3, 4, 5, 7, 9, and 10, 13, 14, and 15 (Parts 6 and 8 encode Latin/Arabic and Latin/Hebrew respectively and are not covered by the Multilingual European Subset No. 2).
- ISO/IEC 6937 as described in Table 4 of the standard.
- ISO/IEC 10367 as described in its normative annex C, minus the Arabic and Hebrew letters.

### A.3 Vendors' coded character sets

In addition the subset covers a large number of vendors' coded character sets, commonly called "code pages" or "code tables". Some of these are listed below.

Example of some code pages which match the ISO standards:

- 813 (same as ISO/IEC 8859-7, Greek)
- 819 (same as ISO/IEC 8859-1, Latin-1)
- 912 (same as ISO/IEC 8859-2, Latin-2)
- 915 (same as ISO/IEC 8859-5, Cyrillic)

Examples of code pages used in the personal computer environment:

- 437, 850, 1004 (internal to OS/2), and 1252 (used in America and Western Europe)
- 737, 851, 869, and 1253 (used in Greece)
- 775 and 1257 (used in Latvia and Lithuania)
- 852 and 1250 (used in Central Europe, Latin-2)
- 855, 866, and 1251 (used in countries using the Cyrillic script)
- 857 and 1254 (used in Turkey)
- 860, 861, 863, and 865 (legacy Latin-based code pages)
- Roman-8 (used in Hewlett-Packard Laser Jet printers and compatibles)

Further examples are the following Apple Macintosh character sets:

- Extended Roman (Roman)
- Icelandic Roman (IS)
- Celtic Roman (PQ)
- Gaelic Roman (GS)
- Croatian Roman (HR)
- Romanian Roman (RO)
- Turkish Roman (TU)
- East European (CE)
- Cyrillic (RU)
- Greek (GR)

Still further examples are the following EBCDIC code pages:

- 037 (used mostly in North America; similar to Latin-1)
- 273 (used in Germany)
- 277 (used in Denmark and Norway)
- 278 (used in Finland and Sweden)
- 280 (used in Italy)
- 284 (used in Spain)
- 285 (used in the United Kingdom)
- 297 (used in France)
- 500 (used mostly in Western Europe, similar to Latin-1)
- 870 (used in Central Europe, Latin-2)
- 871 (used in Iceland)
- 875 (used in Greece)
- 1025 (Slavic Cyrillic)
- 1026 (used in Turkey)

NOTE: The code tables mentioned here are available commercially in various implementations from a number of vendors. Because they are widely available, their repertoires were considered when selecting characters for inclusion in these subsets. This information is given for the convenience of users of this CEN Workshop Agreement and does not constitute an endorsement by CEN/ISSS of any particular product.

## Annex B. List of languages covered by MES-1 (Informative)

The Multilingual European Subset No 1 is believed to cover at least the languages listed here:

Afrikaans	Finnish	Luxemburgish	Romanian ( <i>with restrictions; has Ș ș Ț ț though Ș ș Ț ț are preferred</i> )
Albanian	French	Maltese	Scottish Gaelic
Basque	Frisian	Manx Gaelic	Slovak
Breton	Galician	Moldavian ( <i>new orthography, with restrictions; has Ș ș Ț ț though Ș ș Ț ț are preferred</i> )	Slovenian
Catalan	German	Northern Sámi	Lower Sorbian
Croatian	Greenlandic	Norwegian	Upper Sorbian
Czech	Hungarian	Occitan	Spanish
Danish	Icelandic	Polish	Swedish
Dutch	Irish Gaelic ( <i>new orthography</i> )	Portuguese	Turkish
English	Italian	Rhaeto-Romanic	Welsh ( <i>with restrictions; only Ŵŵ Ț ț Ț ț Ț ț and Ț ț</i> )
Esperanto	Latvian		
Estonian	Lithuanian		
Faroese			

## Annex C. List of languages covered by MES-2 (Informative)

In addition to the languages listed in annex B, the Multilingual European Subset No. 2 is believed to cover at least the languages listed in C.1-C.3.

### C.1 Latin script

Arumanian  
Asturian  
Azerbaijani (*new orthography*)  
Cornish  
Friulian  
Inari Sámi  
Irish Gaelic (*old and new orthographies*)  
Istro-Romanian  
Karelian  
Kashubian  
Ladin  
Latin  
Lule Sámi  
Megleno-Romanian  
Northern Sámi  
Romani  
Romanian  
Skolt Sámi  
Southern Sámi  
Vepsian  
Votic  
Welsh

### C.2 Greek script

Greek

### C.3 Cyrillic script

Abaza  
Abkhaz  
Adyge  
Altai  
Avar  
Azerbaijani (*old orthography*)  
Balkar  
Bashkir  
Belarussian  
Bulgarian  
Buryat  
Chechen  
Chukchi  
Chuvash  
Crimean Tatar  
Dargwa  
Dungan  
Even  
Evenki  
Gagauz  
Hill Mari  
Ingush  
Kabardian  
Kalmuk  
Kalmyk  
Karaim  
Karakalpak  
Kazakh  
Khakas  
Khanty  
Komi  
Komi-Permyak  
Koryak  
Kumyk  
Kyrgyz

Lak  
Lezgian  
Mansi  
Meadow Mari  
Moksha  
Moldavian (*old orthography*)  
Nanai  
Nenets  
Nogai  
Ossetian  
Romani  
Russian  
Rutul  
Serbian  
Siberian Yupik  
Slavic Macedonian  
Tabasaran  
Tajik  
Tatar  
Tati  
Türkmen  
Tuva  
Udmurt  
Uighur  
Ukrainian  
Uzbek  
Yakut



## Annex D. Rationales for inclusion of characters in this CWA (Informative)

### D.1 Rationale for inclusion of characters in MES-1

This subset is intended to be identical to the LL8 repertoire defined in EN 1923:1998, with the addition of 20AC EURO SIGN. Note that the LL8 repertoire is in turn based on ISO/IEC 6937:1994 with the addition of 00D0 LATIN CAPITAL LETTER ETH.

### D.2 Rationale for inclusion of characters in MES-2

The CEN/ISSS WS/MES selected the following repertoires for inclusion in the MES-2 for reasons of language coverage and backward compatibility with widely-used character sets:

EN 1923:1998 LL8	ISO/IEC 8859-13
ISO/IEC 8859-1	ISO/IEC 8859-14
ISO/IEC 8859-2	ISO/IEC 8859-15
ISO/IEC 8859-3	ISO-IR 158
ISO/IEC 8859-4	cp437
ISO/IEC 8859-5	cp850
ISO/IEC 8859-7	cp1004 (internal OS/2 page)
ISO/IEC 8859-9	cp1252
ISO/IEC 8859-10	Macintosh Roman

#### D.2.1 These character sets contain the following 690 characters:

Plane 00

**Rows**    **Positions (Cells)**

00	20–7E A0–FF
01	00–7F 92 B7 DE–EF FA–FF
02	92 BC–BD C6–C7 C9 D6 D8–DD
03	84–8A 8C 8E–A1 A3–CE
04	01–0C 0E–4F 51–5C 5E–5F 90–91
1E	02–03 0A–0B 1E–1F 40–41 56–57 60–61 6A–6B 80–85 F2–F3
20	13–15 17–1E 20–22 26 30 32–33 39–3A 3C 3E 44 7F A3–A4 A7 AC
21	05 13 16 22 26 2E 5B–5E 90–95 A8
22	02 06 0F 11–12 15 19–1A 1E–1F 29 2B 48 59 60–61 64–65
23	02 10 20–21
25	00 02 0C 10 14 18 1C 24 2C 34 3C 50–6C 80 84 88 8C 90–93 A0–A1 AA–AC B2 BA BC C4 CA–CB CF D8–D9 E6
26	3A–3C 40 42 60 63 65–66 6A–6B
FB	01–02
FF	FD

**D.2.2** The following characters, additional to the repertoires listed above, were also selected for inclusion in the MES-2. Each is followed by a brief parenthetical statement indicating its rationale.

#### D.2.2.1 The following 20 characters:

012C	LATIN CAPITAL LETTER I WITH BREVE (Latin)
012D	LATIN SMALL LETTER I WITH BREVE (Latin)
017F	LATIN SMALL LETTER LONG S (older documents in English, French, German, Irish Gaelic, etc.)
01FA	LATIN CAPITAL LETTER A WITH RING ABOVE AND ACUTE (Danish official orthography and lexicography)
01FB	LATIN SMALL LETTER A WITH RING ABOVE AND ACUTE (Danish official orthography and lexicography)
01FC	LATIN CAPITAL LETTER AE WITH ACUTE (Danish official orthography and lexicography)
01FD	LATIN SMALL LETTER AE WITH ACUTE (Danish official orthography and lexicography)
01FE	LATIN CAPITAL LETTER O WITH STROKE AND ACUTE (Danish official orthography and lexicography)
01FF	LATIN SMALL LETTER O WITH STROKE AND ACUTE (Danish official orthography and lexicography)
0387	GREEK ANO TELEIA (Greek orthography and grammar)
0490	CYRILLIC CAPITAL LETTER GHE WITH UPTURN (Ukrainian)
0491	CYRILLIC SMALL LETTER GHE WITH UPTURN (Ukrainian)
201B	SINGLE HIGH-REVERSED-9 QUOTATION MARK (ENV 1973)

2032	PRIME (ENV 1973)
2033	DOUBLE PRIME (ENV 1973)
203E	OVERLINE (ENV 1973)
20A3	FRENCH FRANC SIGN (ENV 1973)
20A4	LIRA SIGN (ENV 1973)
2105	CARE OF (ENV 1973)
2212	MINUS SIGN (ENV 1973)

**D.2.2.2** The following 233 characters used in Greek *polytonikó* orthography:

1F 00-15 18-1D 20-45 48-4D 50-57 59 5B 5D 5F-7D 80-B4 B6-C4 C6-D3 D6-DB DD-EF F2-F4 F6-FE

**D.2.2.3** The following 42 additional characters:

NOTE: Parentheses are given around the hex representation of characters under ISO/IEC JTC1/SC2 ballot.

018F	LATIN CAPITAL LETTER SCHWA (Azerbaijani ( <i>new orthography</i> ))
0259	LATIN SMALL LETTER SCHWA
0218	LATIN CAPITAL LETTER S WITH COMMA BELOW (Romanian)
0219	LATIN SMALL LETTER S WITH COMMA BELOW
021A	LATIN CAPITAL LETTER T WITH COMMA BELOW (Romanian)
021B	LATIN SMALL LETTER T WITH COMMA BELOW
021E	LATIN CAPITAL LETTER H WITH CARON (Finnish Romani)
021F	LATIN SMALL LETTER H WITH CARON
027C	LATIN SMALL LETTER R WITH LONG LEG (older documents in Irish Gaelic; compatibility character)
02EE	MODIFIER LETTER DOUBLE APOSTROPHE (Nenets)
0374	GREEK NUMERAL SIGN (Greek orthography and grammar)
0375	GREEK LOWER NUMERAL SIGN (Greek orthography and grammar)
037A	GREEK YPOGEGRAMMENI (Greek orthography and grammar)
037E	GREEK QUESTION MARK (Greek orthography and grammar)
03D7	GREEK KAI SYMBOL (Greek sign for 'and')
03DA	GREEK LETTER STIGMA (used as a number in legal contexts)
03DB	GREEK SMALL LETTER STIGMA
03DC	GREEK LETTER DIGAMMA (archaic Greek letter; Greek linguistics)
03DD	GREEK SMALL LETTER DIGAMMA
03DE	GREEK LETTER KOPPA (used as a number in legal contexts)
03DF	GREEK SMALL LETTER KOPPA
03E0	GREEK LETTER SAMPI (used as a number in legal contexts)
03E1	GREEK SMALL LETTER SAMPI
1E9B	LATIN SMALL LETTER LONG S WITH DOT ABOVE (older documents in Irish Gaelic; compatibility character)
204A	TIRONIAN SIGN ET (Irish Gaelic sign for 'and')
2083	SUBSCRIPT TWO (common chemical formulae)
20AF	DRACHMA SIGN (Greek currency sign)
2200	FOR ALL (mathematics: logic)
2203	THERE EXISTS (mathematics: logic)
2208	ELEMENT OF (mathematics: set theory)
2209	NOT AN ELEMENT OF (mathematics: set theory)
2227	LOGICAL AND (mathematics: logic)
2228	LOGICAL OR (mathematics: logic)
222A	UNION (mathematics: set theory)
2259	ESTIMATES (mathematics: commonly used corresponds-to symbol)
2282	SUBSET OF (mathematics: set theory)
2283	SUPERSET OF (mathematics: set theory)
2295	CIRCLED PLUS (mathematics: generic algebraic operator)
2297	CIRCLED TIMES (mathematics: generic algebraic operator)
2329	LEFT-POINTING ANGLE BRACKET (mathematics: encloses sequences of numbers)
232A	RIGHT-POINTING ANGLE BRACKET (mathematics: encloses sequences of numbers)
FFFD	REPLACEMENT CHARACTER (signals characters outside the supported set)

**D.2.2.4** The following 87 additional UCS Cyrillic characters from CEN/CENELEC IT/PT004 additional to those Cyrillic characters found in ISO/IEC 8859-5 and in WGL4:

- 0492 CYRILLIC CAPITAL LETTER GHE WITH STROKE (Azerbaijani (*old orthography*), Bashkir, Karakalpak, Kazakh, Khakas, Romani, Tajik, Uighur, Uzbek)
- 0493 CYRILLIC SMALL LETTER GHE WITH STROKE
- 0494 CYRILLIC CAPITAL LETTER GHE WITH MIDDLE HOOK (Abkhaz, Yakut)
- 0495 CYRILLIC SMALL LETTER GHE WITH MIDDLE HOOK
- 0496 CYRILLIC CAPITAL LETTER ZHE WITH DESCENDER (Dungan, Kalmyk, Tatar, Türkmen, Uighur)
- 0497 CYRILLIC SMALL LETTER ZHE WITH DESCENDER
- 0498 CYRILLIC CAPITAL LETTER ZE WITH DESCENDER (Bashkir)
- 0499 CYRILLIC SMALL LETTER ZE WITH DESCENDER
- 049A CYRILLIC CAPITAL LETTER KA WITH DESCENDER (Abkhaz, Karakalpak, Kazakh, Khanty, Koryak, Tajik, Uighur, Uzbek)
- 049B CYRILLIC SMALL LETTER KA WITH DESCENDER
- 049C CYRILLIC CAPITAL LETTER KA WITH VERTICAL STROKE (Azerbaijani (*old orthography*))
- 049D CYRILLIC SMALL LETTER KA WITH VERTICAL STROKE
- 049E CYRILLIC CAPITAL LETTER KA WITH STROKE (Abkhaz)
- 049F CYRILLIC SMALL LETTER KA WITH STROKE
- 04A0 CYRILLIC CAPITAL LETTER BASHKIR KA (Bashkir)
- 04A1 CYRILLIC SMALL LETTER BASHKIR KA
- 04A2 CYRILLIC CAPITAL LETTER EN WITH DESCENDER (Bashkir, Dungan, Kalmyk, Kazakh, Khakas, Khanty, Kyrgyz, Mansi, Tatar, Türkmen, Tuvin, Uighur)
- 04A3 CYRILLIC SMALL LETTER EN WITH DESCENDER
- 04A4 CYRILLIC CAPITAL LIGATURE EN GHE (Altai, Hill Mari, Meadow Mari, Yakut)
- 04A5 CYRILLIC SMALL LIGATURE EN GHE
- 04A6 CYRILLIC CAPITAL LETTER PE WITH MIDDLE HOOK (Abkhaz)
- 04A7 CYRILLIC SMALL LETTER PE WITH MIDDLE HOOK
- 04A8 CYRILLIC CAPITAL LETTER ABKHASIAN HA (Abkhaz)
- 04A9 CYRILLIC SMALL LETTER ABKHASIAN HA
- 04AA CYRILLIC CAPITAL LETTER ES WITH DESCENDER (Bashkir, Chuvash)
- 04AB CYRILLIC SMALL LETTER ES WITH DESCENDER
- 04AC CYRILLIC CAPITAL LETTER TE WITH DESCENDER (Abkhaz)
- 04AD CYRILLIC SMALL LETTER TE WITH DESCENDER
- 04AE CYRILLIC CAPITAL LETTER STRAIGHT U (Azerbaijani (*old orthography*), Bashkir, Buryat, Dungan, Kalmyk, Karakalpak, Kazakh, Kyrgyz, Tatar, Türkmen, Tuvin, Uighur, Yakut)
- 04AF CYRILLIC SMALL LETTER STRAIGHT U
- 04B0 CYRILLIC CAPITAL LETTER STRAIGHT U WITH STROKE (Kazakh)
- 04B1 CYRILLIC SMALL LETTER STRAIGHT U WITH STROKE
- 04B2 CYRILLIC CAPITAL LETTER HA WITH DESCENDER (Abkhaz, Karakalpak, Tajik, Uzbek)
- 04B3 CYRILLIC SMALL LETTER HA WITH DESCENDER
- 04B4 CYRILLIC CAPITAL LIGATURE TE TSE (Abkhaz)
- 04B5 CYRILLIC SMALL LIGATURE TE TSE
- 04B6 CYRILLIC CAPITAL LETTER CHE WITH DESCENDER (Abkhaz, Khakas, Tajik)
- 04B7 CYRILLIC SMALL LETTER CHE WITH DESCENDER
- 04B8 CYRILLIC CAPITAL LETTER CHE WITH VERTICAL STROKE (Azerbaijani (*old orthography*))
- 04B9 CYRILLIC SMALL LETTER CHE WITH VERTICAL STROKE
- 04BA CYRILLIC CAPITAL LETTER SHHA (Azerbaijani (*old orthography*), Bashkir, Buryat, Kalmyk, Kazakh, Tatar, Uighur, Yakut)
- 04BB CYRILLIC SMALL LETTER SHHA
- 04BC CYRILLIC CAPITAL LETTER ABKHASIAN CHE (Abkhaz)
- 04BD CYRILLIC SMALL LETTER ABKHASIAN CHE
- 04BE CYRILLIC CAPITAL LETTER ABKHASIAN CHE WITH DESCENDER (Abkhaz)
- 04BF CYRILLIC SMALL LETTER ABKHASIAN CHE WITH DESCENDER
- 04C0 CYRILLIC LETTER PALOCHKA (Abaza, Adyge, Avar, Chechen, Dargwa, Ingush, Kabardian, Lak, Lezgian, Rutul, Tabasaran, Tati)
- 04C1 CYRILLIC CAPITAL LETTER ZHE WITH BREVE (Moldavian (*old orthography*))
- 04C2 CYRILLIC SMALL LETTER ZHE WITH BREVE
- 04D0 CYRILLIC CAPITAL LETTER A WITH BREVE (Chuvash)
- 04D1 CYRILLIC SMALL LETTER A WITH BREVE
- 04D2 CYRILLIC CAPITAL LETTER A WITH DIAERESIS (Abkhaz, Azerbaijani (*old orthography*), Gagauz, Hill Mari, Kalmyk, Khanty, Mountain Mari)
- 04D3 CYRILLIC SMALL LETTER A WITH DIAERESIS
- 04D4 CYRILLIC CAPITAL LIGATURE A IE (Ossetian)
- 04D5 CYRILLIC SMALL LIGATURE A IE
- 04D6 CYRILLIC CAPITAL LETTER IE WITH BREVE (Chuvash)
- 04D7 CYRILLIC SMALL LETTER IE WITH BREVE
- 04D8 CYRILLIC CAPITAL LETTER SCHWA (Abkhaz, Azerbaijani (*old orthography*), Bashkir, Dungan, Kalmyk, Kazakh, Tatar, Türkmen, Uighur)
- 04D9 CYRILLIC SMALL LETTER SCHWA

04DA	CYRILLIC CAPITAL LETTER SCHWA WITH DIAERESIS (Khanty)
04DB	CYRILLIC SMALL LETTER SCHWA WITH DIAERESIS
04DC	CYRILLIC CAPITAL LETTER ZHE WITH DIAERESIS (Udmurt)
04DD	CYRILLIC SMALL LETTER ZHE WITH DIAERESIS
04DE	CYRILLIC CAPITAL LETTER ZE WITH DIAERESIS (Udmurt)
04DF	CYRILLIC SMALL LETTER ZE WITH DIAERESIS
04E0	CYRILLIC CAPITAL LETTER ABKHASIAN DZE (Abkhaz)
04E1	CYRILLIC SMALL LETTER ABKHASIAN DZE
04E2	CYRILLIC CAPITAL LETTER I WITH MACRON (Tajik)
04E3	CYRILLIC SMALL LETTER I WITH MACRON
04E4	CYRILLIC CAPITAL LETTER I WITH DIAERESIS (Udmurt)
04E5	CYRILLIC SMALL LETTER I WITH DIAERESIS
04E6	CYRILLIC CAPITAL LETTER O WITH DIAERESIS (Altai, Gagauz, Hill Mari, Karaim, Khakas, Khanty, Komi, Komi-Permyak, Meadow Mari, Udmurt)
04E7	CYRILLIC SMALL LETTER O WITH DIAERESIS
04E8	CYRILLIC CAPITAL LETTER BARRED O (Azerbaijani ( <i>old orthography</i> ), Bashkir, Buryat, Kalmyk, Karakalpak, Kazakh, Khanty, Kyrgyz, Tatar, Türkmen, Tuvin, Uighur, Yakut)
04E9	CYRILLIC SMALL LETTER BARRED O
04EA	CYRILLIC CAPITAL LETTER BARRED O WITH DIAERESIS (Khanty)
04EB	CYRILLIC SMALL LETTER BARRED O WITH DIAERESIS
04EE	CYRILLIC CAPITAL LETTER U WITH MACRON (Tajik)
04EF	CYRILLIC SMALL LETTER U WITH MACRON
04F0	CYRILLIC CAPITAL LETTER U WITH DIAERESIS (Altai, Gagauz, Hill Mari, Kalmyk, Karaim, Khakas, Khanty, Meadow Mari)
04F1	CYRILLIC SMALL LETTER U WITH DIAERESIS
04F2	CYRILLIC CAPITAL LETTER U WITH DOUBLE ACUTE (Chuvash)
04F3	CYRILLIC SMALL LETTER U WITH DOUBLE ACUTE
04F4	CYRILLIC CAPITAL LETTER CHE WITH DIAERESIS (Udmurt)
04F5	CYRILLIC SMALL LETTER CHE WITH DIAERESIS
04F8	CYRILLIC CAPITAL LETTER YERU WITH DIAERESIS (Hill Mari)
04F9	CYRILLIC SMALL LETTER YERU WITH DIAERESIS

#### **D.2.2.5** The following 10 additional Cyrillic characters:

NOTE: Parentheses are given around the hex representation of characters under ISO/IEC JTC1/SC2 ballot.

0400	CYRILLIC CAPITAL LETTER IE WITH GRAVE (Slavic Macedonian)
040D	CYRILLIC SMALL LETTER E WITH GRAVE
0450	CYRILLIC CAPITAL LETTER IE WITH GRAVE (Slavic Macedonian)
045D	CYRILLIC SMALL LETTER E WITH GRAVE
04C3	CYRILLIC CAPITAL LETTER KA WITH HOOK (Chukchi, Khanty, Koryak)
04C4	CYRILLIC SMALL LETTER KA WITH HOOK
04C7	CYRILLIC CAPITAL LETTER EN WITH HOOK (Nenets)
04C8	CYRILLIC SMALL LETTER EN WITH HOOK
04CB	CYRILLIC CAPITAL LETTER KHAKASSIAN CHE (Khakas)
04CC	CYRILLIC SMALL LETTER KHAKASSIAN CHE

### D.3 Rationale for inclusion of characters in MES-3

This subset is intended to cover entire collections of the UCS, as a collection-based superset of MES-1 and MES-2. It includes all characters belonging to the Latin, Greek, and Cyrillic scripts, plus two collections for the Armenian and Georgian scripts, and four collections of other characters widely used in Europe. These six collections are:

7	COMBINING DIACRITICAL MARKS	0300–036F	used with European scripts and IPA
11	ARMENIAN	0530–058F	Armenian script
27	BASIC GEORGIAN	10D0–10FF	Georgian script
35	COMBINING DIACRITICAL MARKS FOR SYMBOLS	20D0–20FF	publishing and science
42	OPTICAL CHARACTER RECOGNITION	2440–245F	banking
65	COMBINING HALF MARKS	FE20–FE2F	bibliographic use

### Specification for the Multilingual European Subset No. 3A

Collection Name: MES-3A  
Type of Collection (Fixed or Non-fixed): NON-FIXED

The following UCS collections from annex A of ISO/IEC 10646-1 comprise MES-3A:

No.	Collection name	hex range	rationale
1	BASIC LATIN	0020–007E	Latin script
2	LATIN-1 SUPPLEMENT	00A0–00FF	Latin script
3	LATIN EXTENDED-A	0100–017F	Latin script
4	LATIN EXTENDED-B	0180–024F	Latin script
5	IPA EXTENSIONS	0250–02AF	Latin script
6	SPACING MODIFIER LETTERS	02B0–02FF	used with European scripts and IPA
7	COMBINING DIACRITICAL MARKS	0300–036F	used with European scripts and IPA
8	BASIC GREEK	0370–03CF	Greek script
9	GREEK SYMBOLS AND COPTIC	03D0–03FF	Greek script
10	CYRILLIC	0400–04FF	Cyrillic script
11	ARMENIAN	0530–058F	Armenian script
27	BASIC GEORGIAN	10D0–10FF	Georgian script
30	LATIN EXTENDED ADDITIONAL	1E00–1EFF	Latin script
31	GREEK EXTENDED	1F00–1FFF	Greek script
32	GENERAL PUNCTUATION	2000–206F	used with all European scripts
33	SUPERSCRIPTS AND SUBSCRIPTS	2070–209F	publishing and science
34	CURRENCY SYMBOLS	20A0–20CF	banking and general use
35	COMBINING DIACRITICAL MARKS FOR SYMBOLS	20D0–20FF	publishing and science
36	LETTERLIKE SYMBOLS	2100–214F	publishing and science
37	NUMBER FORMS	2150–218F	publishing and science
38	ARROWS	2190–21FF	publishing and science
39	MATHEMATICAL OPERATORS	2200–22FF	publishing and science
40	MISCELLANEOUS TECHNICAL	2300–23FF	publishing and science
42	OPTICAL CHARACTER RECOGNITION	2440–245F	banking
44	BOX DRAWING	2500–257F	legacy character sets
45	BLOCK ELEMENTS	2580–259F	publishing and science
46	GEOMETRIC SHAPES	25A0–25FF	publishing and science
47	MISCELLANEOUS SYMBOLS	2600–26FF	publishing
63	ALPHABETIC FORMS	FB00–FB4F	publishing
65	COMBINING HALF MARKS	FE20–FE2F	bibliographic use
70	SPECIALS	FFF0–FFFD	replacement characters

### Annex E. Bibliography (Informative)

CEN/CENELEC IT/PT004. 1992. *Report from the project team on Definition of a Cyrillic primary set of graphic characters.*

Unicode Consortium. 1996. *The Unicode Standard. Version 2.0.* Reading: Addison-Wesley. ISBN 0-201-48345-9.

## Annex F. Comments accompanying Negative ballots (Informative)

### Comments accompanying Keld Simonsen's Negative ballot:

In creating a Multilingual European Subset it is in my opinion necessary not to include characters that are not for European use, such as letters not used in languages of European origin. An European standardization organization should not make specifications that primarily pertain to conditions outside Europe. The MES-3 specified in this CWA contains such characters, such as letters used in Vietnamese, Asian, or African languages. The following fixed collection is intended to exclude non-European characters.

Collection Name: MES-3KS  
Type of Collection (Fixed or Non-fixed): FIXED

Plane 00

#### Rows Positions (Cells)

00	20-7E A0-FF
01	00-81 8B-8C 8F 92 95 9A-9B 9E-9F A2-A3 A6 AA-AB B5-BB BE-CC D5-D6 DE-F7 FA-FF
02	00-1F 24-27 2A-33 50-AD B0-EE
03	00-4E 60-62 74-75 7A 7E 84-8A 8C 8E-A1 A3-CE D0-D7 DA-F3
04	00-86 88-89 8C-8F 90-C4 C7-C8 CB-CC D0-ED EE-F5 F5 F8-F9
05	31-56 59-5F 61-87 89-8A
10	D0-F6 FB
1E	00-9B F2-F3
1F	00-15 18-1D 20-45 48-4D 50-57 59 5B 5D 5F-7D 80-B4 B6-C4 C6-D3 D6-DB DD-EF F2-F4 F6-FE
20	00-46 48-4D 6A-70 74-8E A0-AF D0-E3
21	00-3A 53-83 90-F3
22	00-F1
23	00-7B 7D-9A
24	40-4A
25	00-95 A0-F7
26	00-13 19-6F 70-71
FB	00-06 13-17
FE	20-23
FF	F9-FD

*The Editor notes that this specification has not been standardized. It is the same as that for MES-3B excluding the following characters:*

Plane 00

#### Rows Positions (Cells)

01	82-8A 8D-8E 90-91 93-94 96-99 9C-9D A0-A1 A4-A5 A7-A9 AC-B4 BC-BD CD-D4 D7-DD F8-F9
02	22-23 28-29
1E	A0-F1 F4-F9

### Comments accompanying Johan van Wingen's Negative ballot:

We (*i.e.* Johan van Wingen and NNI's national committee for standardization of character set coding) have followed with increasing concern the development of this Work Item. We consider reaching consensus impossible anymore. Should the three MESs be approved, then these may be adopted by industry. Then we may be forced to include MES-2 in our systems. We have calculated that this would take NLG 11 million to do that. Thus we consider continuing the Work Item to be in conflict with our national interests.