CEN WORKSHOP AGREEMENT Draft for CWA TC304/P10:1998

1998-05-30

ICS: 35.040

Descriptors: Data processing, information interchange, text processing, text communication, graphic characters, character sets, representation of characters, coded character sets, architecture

English version

Information Technology -

Multilingual European Subsets of ISO/IEC 10646-1

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

Informationstechnologie -
Mehrsprächige europäische Untermengen von ISO/IEC 10646-1

This CEN Workshop Agreement has been drawn up by the Technical Committee CEN/TC 304.

NOTE: reviewers of this DRAFT document are asked to look carefully at the repertoires. Some rationale and explanation for each of the subsets is given in annexes at the back, but this may be sketchy or not the full kind of rationale we may wish.

Some of the other text was taken from ENV 1973:1995 and either should be deleted from this draft or be revised.

The editor invites comment.

This CEN Workshop Agreement was established by CEN in one official version (English). 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 version.

CEN members are the national bodies of Austria, Belgium, 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 1973:1998 E

This page has been intentionally left blank.

Contents

Foreword

This CEN Workshop Agreement has been prepared by the WS/MES under the auspices of Technical Committee CEN/TC304 “Technologies for Information and Communication: European Localization Requirements” of which the secretariat is held by STRÍ.

It was approved in the month of _____ 1998.

This CEN Workshop Agreement differs from previously approved standards and interim standards in the field of European character sets by having a much larger number of characters even if it is only a small subset of the corresponding international standard.

0. Introduction

0.1 Requirements

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

The countries in question have a large number of official languages and officially-accepted native minority languages. These employ a large number of letters. In addition, a large number of other graphic characters are required for day-to-day computer use in Europe.

0.2 Towards large character sets

The Multilingual Subset No. 1 specified by this CEN Workshop Agreement provides for the minimal needs of some national governmental administrations in Europe. The Multilingual Subset No. 2 specified by this CEN Workshop Agreement provides a greater step towards the implementation of large character sets in Europe. The Multilingual Subset No. 3, provides a second step by covering all characters belonging to European scripts. The implementation of all the world's characters (that is, of the entire Universal Character Set) is the ultimate goal.

It is not assumed that CEN nor Europeans in general have the expertise to make provision for languages originating outside Europe. In other parts of the world, work regarding the definition and implementation of locally-important scripts (*e.g.* Hebrew, Arabic, Devanagari, Korean) is ongoing. To date, however, coherent specification of the living scripts of Europe (*i.e.* Latin, Cyrillic, Greek, Armenian, Georgian) has not been made.

This CEN Workshop Agreement is one of the results of earlier CEN/TC304 work on the characters and the scripts of Europe. It is based in part on the results of another work item that specifies the characters used by the indigenous languages of Europe.

0.3 ISO/IEC 10646-1 Universal Character Set (UCS), Basic Multilingual Plane (BMP)

Prior to the UCS, there was no character set standard that included all the characters and scripts needed by Europe. The possibilities to combine several standardized character sets – 7-bit or 8-bit coded – in the same IT-systems with existing code extension techniques have proven to be impractical and insufficient.

Part 1 of ISO/IEC 10646-1 (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.

0.4 Implementing the Universal Character Set

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

Implementation of ISO/IEC 10646-1 characters requires resources for supporting *e.g.*:

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

The resources needed are material or human, *e.g.*:

- 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

0.5 Migration path

Some of the implementations problems discussed above could be solved by subsets of the UCS. The work on European subsets 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 will be costly in the first stages of UCS use, and that 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 all characters for use in European languages plus other frequently used characters. Such subsets may also be useful with regard to further standardization work (for example, on sorting specifications), so that the work is reasonably limited and still useful in a European environment. Only a limited number of subsets should be specified.

1. Scope

This CEN Workshop Agreement specifies three *limited subsets* of ISO/IEC 10646-1 for use in Europe. Multilingual European Subset No. 1 (MES-1) and Multilingual European Subset No. 2 (MES-2) are *limited subsets* that include the characters used by a number of European languages. Multilingual European Subset No. 3 (MES-3) is a limited subset that includes all of the characters belonging to the modern European scripts found in ISO/IEC 10646-1. It is useful to note that the MES-3 differs from the MES-1 and MES-2 mainly by basing its selection of characters by script and function, rather than by use in particular languages.

NOTE: Rationale for the composition of each of the subsets is found in annexes C, D, and E.

This CEN Workshop Agreement covers the repertoires of all options of EN 1923, ENV 41503:1990, ENV 41505:1991, and ENV 41508:1990. It also covers the repertoires of commonly used registered coded character sets and commonly used vendor coded character sets.

NOTE: Examples of these character sets are found in Annex C.

2. Normative references

EN 1923:1998 Title.

ISO/IEC 8824-1:1995 Information technology – Abstract Syntax Notation One (ASN.1) – Specification of Basic Notation (third edition).

ISO/IEC 10646-1:1993 Information technology – Universal Multiple-Octet Coded Character Set (UCS) – Part 1: Architecture and Basic Multilingual Plane; and its amendments 1–12.

3. Definitions and abbreviations

3.1 Definitions

For the purposes of this CEN Workshop Agreement the basic definitions of ISO/IEC 10646-1 clause 4 apply. The following are reproduced here for ease of reference:

abstract syntax: The specification of Application Layer data or application-protocol-control-information by using notation rules which are independent of the encoding technique used to represent them (ISO/IEC 8822:1994).

character abstract syntax: Any abstract syntax whose values are specified as the set of character strings of zero, one or more characters from some specified collection of characters (ISO/IEC 8824-1:1995).

character transfer syntax: Any transfer syntax for a character abstract syntax for which all bit patterns are an integral multiple of eight bits (ISO/IEC 8824-1:1995).

coded character: A character together with its coded representation (ISO/IEC 10646-1).

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 (ISO/IEC 10646-1).

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 (ISO/IEC 10646-1).

concrete syntax: Those aspects of the rules used in the formal specification of data which embody a specific representation of that data (ISO 7498-1:1994).

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 (ISO/IEC 10646-1 §13.1).

object identifier: A value (distinguishable from all other such values) which is associated with an information object (ISO/IEC 8824-1:1995).

script: A set of graphic characters used for the written form of one or more languages. (ISO/IEC 10646-1 §4.34).

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 (ISO/IEC 10646-1 §13.2).

transfer syntax: That concrete syntax used in the transfer of data between open systems (ISO 7498-1:1994).

3.2 Abbreviations

The following abbreviations apply:

3.2.1 *BMP:* Basic Multilingual Plane

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

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

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

3.2.5 *UCS:* Universal Character Set

4 Compliance, conformance, and procurement

4.0 Compliance

Clauses 5, 6, and 7 of this CEN Workshop Agreement is in compliance with the provisions of ISO/IEC 10646-1 clause 13 for the specification of a limited subset.

4.1 Conformance to 10646

Implementations claiming conformance to this CEN Workshop Agreement shall also conform to the conformance provisions of ISO/IEC 10646-1. A claim of conformance shall be supported by a completed Implementation Conformance Statement (ICS) proforma (see Annex A).

4.2 Support of the Multilingual European Subsets

4.2.1 The supported subsets shall be one or more than one of the Multilingual European Subsets, as identified in clauses 5, 6, and 7.

> NOTE: The MES-3 is a true superset of both the MES-2 and the MES-1. The MES-2 is a true superset of the MES-1. Support of the MES-3 implies support of the MES-2 and the MES-1. Support of the MES-2 implies support of the MES-1.

4.2.2 A sending implementation shall be capable of coding the MES-1 and/or the MES-2 at level 1 of ISO/IEC 10646-1 using one of the codings as specified in ISO/IEC 10646-1, and/or shall be capable of coding the MES-3 at level 3 of ISO/IEC 10646-1 using one of the codings as specified in ISO/IEC 10646-1.

4.2.3 A receiving implementation shall be capable of supporting the coding of the MES-1 and/or the MES-2 at level 1 of ISO/IEC 10646-1 using one of the codings as specified in ISO/IEC 10646-1, and/or shall be capable of supporting the coding of the MES-3 at level 3 of ISO/IEC 10646-1 using one of the codings as specified in ISO/IEC 10646-1.

4.3 Receiving devices

4.3.1 A receiving device shall be able to present to an internal application the coded representation of any received character within the supported subset, in such a way that the internal application can identify the character.

4.3.2 A receiving device with retransmission capability shall be able to retransmit the coded representation of any character of the supported subset without change to its coded representation unless modified by the user.

4.3.3 Coded representation of characters not contained in the supported subset may or may not be retransmitted.

4.4 Originating devices

An originating device shall allow its user to supply any characters from the supported subset and be capable of transmitting their coded representation.

4.5 Receiving devices with rendering capabilities

A receiving device with rendering capability shall be able to display to the user in a recognizable form any character of the supported subset.

4.6 Procurement

To be supplied.

5. Multilingual European Subset No. 1 (MES-1)

5.1 Specification for the Multilingual European Subset No. 1

Collection Name: MES-1
Type of Collection (Fixed or Open): FIXED

Plane 00

Rows	**Positions (Cells)**
00	20-7E, A0-FF
01	00-13 16-2B 2E-4D 50-7F B7 EE EF
02	18-1B 1E 1F 7C 92 C7 D8-DB DD
1E	02 03 0A 0B 1E 1F 40 41 56 57 60 61 6A 6B 80-85 9B F2 F3
20	15 18 19 1C 1D AC
21	22 26 5B-5E 90-93
26	6A

5.2 Collections containing the MES-1

The following UCS collections from Annex A of ISO/IEC 10646 containing the MES-1:

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
30	LATIN EXTENDED ADDITIONAL	1E00 - 1EFF
32	GENERAL PUNCTUATION	2000 - 206F
34	CURRENCY SYMBOLS	20A0 - 20CF
36	LETTERLIKE SYMBOLS	2100 - 214F
47	MISCELLANEOUS SYMBOLS	2600 - 26FF

5.3 Identification of this subset.

The following object identifier can be used for the Multilingual European Subset No. 1:

```
     {ISO(1) standard(0) 10646 part-(1) implementation -level(1) 1 2
3 4 5 6 30 32 34 36 47}
```

The MES-1 can also be designated and invoked by making use of the escape sequences defined in clause 17.3 of ISO/IEC 10646-1.

6. Multilingual European Subset No. 2 (MES-2)

6.1 Specification for the Multilingual European Subset No. 2

Collection Name: MES-2
Type of Collection (Fixed or Open): FIXED

Plane 00

Rows	Positions (Cells)
00	20-7E, A0-FF
01	00-7F 92 B7 C4 C6 C7 C9 CA CC DE DF E4-E9 EE EF F1 F3-F5 FA-FF
02	18-1B 1E 1F 7C 92 C6 C7 C9 D8-DD
03	74 75 7A 7E 84-8A 8C 8E-A1 A3-CE
04	00-5F 90 91
1E	02 03 0A 0B 10 11 1E 1F 30 31 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	10 13-15 17-1E 20-23 26 28-30 32 33 39 3A 3C 3E 43 44 7F A3 A4 A7 AC
21	05 16 22 26 2E 5B-5E 90-95 A8
22	02 06 0F 11 12 19 1A 1E 1F 29 2B 48 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 AC B2 BA BC C4 CA CB D8 D9
26	3A-3C 40 42 60 63 65 66 6A 6B
FB	01 02

6.2 Collections containing the MES-2

The following UCS collections from Annex A of ISO/IEC 10646 containing the MES-2:

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
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	Alphabetical Presentation Forms	FB00-FB4F

6.3 Identification of this subset.

The following object identifier can be used for the Multilingual European Subset No. 2:

```
     {ISO(1) standard(0) 10646 part-(1) implementation -level(1) 1 2
3 4 5 6 8 10 30 31 32 33 34 36 37 38 39 40 44 45 46 47 63}
```

The MES-2 can also be designated and invoked by making use of the escape sequences defined in clause 17.3 of ISO/IEC 10646-1.

7. Multilingual European Subset No. 3 (MES-3)

7.1 Specification for the Multilingual European Subset No. 3

Collection Name: MES-3
Type of Collection (Fixed or Open): FIXED

Plane 00

Rows	Positions (Cells)
00	20-7E, A0-FF
01	00-FF
02	00-1F 50-A8 B0-DE E0-E9
03	00-45 60 61 74 75 7A 7E 84-8A 8C 8E-A1 A3-CE D0-D6 DA DC DE E0 E2-F3
04	00-86 90-C4 C7 C8 CB CC D0-EB EE-F5 F8 F9
05	31-56 59-5F 61-87 89 8A
10	A0-C5 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	00-47 6A-70 74-83 A0-AD D0-E1
21	00-38 53-82 90-EA
22	00-F1
23	00 02-7A
24	00-24 40-4A
25	00-95 A0-EF
26	00-13 1A-6F
27	01-04 06-09 0C-27 29-4B 4D 4F-52 56 58-5E 61-67 76-94 98-AF B1-BE
FB	00-06 13-17
FE	20-23
FF	FC FD

7.2 Collections containing the MES-3

The following UCS collections from Annex A of ISO/IEC 10646 containing the MES-3:

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
28	Georgian Extended	10A0-10CF
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
41	Control Pictures	2400-243F
42	Optical Character Recognition	2440-245F
43	Enclosed Alphanumerics	2460-24FF
44	Box Drawing	2500-257F
45	Block Elements	2580-259F
46	Geometric Shapes	25A0-25FF

47	Miscellaneous Symbols	2600-26FF
48	Dingbats	2700-27BF
63	Alphabetical Presentation Forms	FB00-FB4F
65	Combining Half Marks	FE20-FE2F
70	Specials	FFF0-FFFD

7.3 Identification of this subset.

The following object identifier can be used for the Multilingual European Subset No. 3:

```
{ISO(1) standard(0) 10646 part-(1) implementation -level(3) 1 2
3 4 5 6 7 8 9 10 11 27 28 30 31 32 33 34 35 36 37 38 39 40 41 42 43
44 45 46 47 48 63 65 70}
```

The MES-3 can also be designated and invoked by making use of the escape sequences defined in clause 17.3 of ISO/IEC 10646-1.

8. Abstract and transfer syntax

8.1 Character abstract syntax

The following allocation of object identifier values for the Multilingual European Subset No. 1 of ISO/IEC 10646-1 has been made by EWOS/TG.

```
{iso(1) identified-organization(3) EWOS(16) eg(2) tlg(0)
character-abstract-syntax(1) iso10646-option1 (0)}
```

8.2 Character abstract syntax

The following allocation of object identifier values for the Multilingual European Subset No. 2 of ISO/IEC 10646-1 has been made by EWOS/TG.

```
{iso(1) identified-organization(3) EWOS(16) eg(2) tlg(0)
character-abstract-syntax(1) iso10646-option1 (0)}
```

8.3 Character abstract syntax

The following allocation of object identifier values for the Multilingual European Subset No. 3 of ISO/IEC 10646-1 has been made by EWOS/TG.

```
{iso(1) identified-organization(3) EWOS(16) eg(2) tlg(0)
character-abstract-syntax(1) iso10646-option1 (0)}
```

8.4 Character abstract syntax

The following object descriptors for the Multilingual European Subsets of ISO/IEC 10646-1 have been made.

```
ISO-10646-Multilingual-European-Subset-1
ISO-10646-Multilingual-European-Subset-2
ISO-10646-Multilingual-European-Subset-3
```

8.5 Character transfer syntax

Annex M of ISO/IEC 10646-1 assigns ASN.1 object identifier values to the character transfer syntaxes that represent the coding methods defined within that International Standard. Any of these character transfer syntaxes may be used with the character abstract syntax of these European Subsets of ISO/IEC 10646-1.

Annex A. Implementation Conformance Statement (ICS) proforma (Normative)

A.1 Classification

The classification in the status column is as follows:

- m: Support for the feature is mandatory
- o: Support for the feature is optional.
- c: Support for the feature is conditional. The predicate expressions are contained at the end of the relevant tables.
- x: Support for the feature is excluded.
- i: Support for the feature is out of scope.
- -: Support for the feature is not applicable in this context.

The status and support columns are divided into:

- R: to designate Receiving capabilities
- S: to designate Sending capabilities
- O: Originating device
- RR: Receiving device with Rendering capabilities

When filling in the answers to the ICS in the support column, the following classification shall be used:

- y: supported
- n: not supported

A.2 Implementation Details

Ref	Description	Answer
1	Date	
2	Name of organization	
3	Address	
4	Telephone number	
5	Fax number	
6	E-mail	
7	Identification of implementation	
8	Description of implementation	

A.3 Requirements

Ref	Description	Status				Support			
		S	R	O	RR	S	R	O	RR
1	Are the conformance requirements of clause 4.2.1 supported?	m	m	m	m				
2	Are the conformance requirements of clauses 4.2.2 and 4.2.3 supported?	m	m	-	m				
3	Are the conformance requirements of clause 4.3.1 supported?	-	m	-	m				
4	Are the conformance requirements of clause 4.3.2 supported?	m	m	-	m				
5	Are the conformance requirements of clause 4.3.3 supported?	i	i	-	i				
6	Are the conformance requirements of clause 4.4 supported?	-	-	m	-				
7	Are the conformance requirements of clause 4.5 supported?	-	-	-	m				

Annex B. Coded character sets whose repertoires are covered by the MES-2 (Informative)

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

B.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	also known as ASCII set
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 155	Basic box drawing set	
ISO-IR 156	Supplementary set of ISO/IEC 6937	
ISO-IR 157	Latin alphabet No. 6, supplementary set	Nordic (not in use)
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

B.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, 14, and 15 (Parts 6 and 8 encode respectively Latin/Arabic and Latin/Hebrew 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 the normative Annex C, minus the Arabic and Hebrew letters.

B.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, and 1252 (used in America and Western Europe)
- 737, 851, 869, and 1253 (used in Greece)
- 775 and 1257 (to be used in Baltic countries)
- 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 (UA)
- Greek (GR)

Still further examples are the following EBCDIC code pages:

- 037 (used mostly in America, close 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, close to Latin-1)
- 870 (used in Central Europe, Latin-2)
- 871 (used in Iceland)
- 875 (used in Greece)
- 1025 (Cyrillic)
- 1026 (used in Turkey)

NOTE: The code tables mentioned here and in Annex D 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 the MES. This information is given for the convenience of users of this CEN Workshop Agreement and does not constitute an endorsement by CEN of any particular products.

Annex C. Rationale for inclusion of characters in the MES-1 (Informative)

The MES-1 provides Latin script characters, digits, and symbols in simple commercial and telematic use in Europe. It contains all of the characters in ISO 6937, as well as the following characters used for official or regional languages in Europe, and the EURO SIGN.

0110 LATIN CAPITAL LETTER D WITH STROKE (Croatian and Sámi)
0111 LATIN SMALL LETTER D WITH STROKE (Croatian and Sámi)
017F LATIN SMALL LETTER LONG S (Irish Gaelic)
01B7 LATIN CAPITAL LETTER EZH (Sámi)
01EE LATIN CAPITAL LETTER EZH WITH CARON (Sámi)
01EF LATIN SMALL LETTER EZH WITH CARON (Sámi)
0218 LATIN CAPITAL LETTER S WITH CEDILLA (Romanian)
0219 LATIN SMALL LETTER S WITH CEDILLA (Romanian)
021A LATIN CAPITAL LETTER T WITH CEDILLA (Romanian)
021B LATIN SMALL LETTER T WITH CEDILLA (Romanian)
021E LATIN CAPITAL LETTER H WITH CARON (Romany)
021F LATIN SMALL LETTER H WITH CARON (Romany)
027C LATIN SMALL LETTER R WITH LONG LEG (Irish Gaelic)
0292 LATIN SMALL LETTER EZH (Sámi)
1E02 LATIN CAPITAL LETTER B WITH DOT ABOVE (Irish Gaelic)
1E03 LATIN SMALL LETTER B WITH DOT ABOVE (Irish Gaelic)
1E0A LATIN CAPITAL LETTER D WITH DOT ABOVE (Irish Gaelic)
1E0B LATIN SMALL LETTER D WITH DOT ABOVE (Irish Gaelic)
1E1E LATIN CAPITAL LETTER F WITH DOT ABOVE (Irish Gaelic)
1E1F LATIN SMALL LETTER F WITH DOT ABOVE (Irish Gaelic)
1E40 LATIN CAPITAL LETTER M WITH DOT ABOVE (Irish Gaelic)
1E41 LATIN SMALL LETTER M WITH DOT ABOVE (Irish Gaelic)
1E56 LATIN CAPITAL LETTER P WITH DOT ABOVE (Irish Gaelic)
1E57 LATIN SMALL LETTER P WITH DOT ABOVE (Irish Gaelic)
1E60 LATIN CAPITAL LETTER S WITH DOT ABOVE (Irish Gaelic)
1E61 LATIN SMALL LETTER S WITH DOT ABOVE (Irish Gaelic)
1E6A LATIN CAPITAL LETTER T WITH DOT ABOVE (Irish Gaelic)
1E6B LATIN SMALL LETTER T WITH DOT ABOVE (Irish Gaelic)
1E9B LATIN SMALL LETTER LONG S WITH DOT ABOVE (Irish Gaelic)
20AC EURO SIGN

Annex D. Rationale for inclusion of characters in the MES-2 (Informative)

The MES-2 provides Latin, Greek, and Cyrillic characters, digits, and symbols in general use in Europe. It contains, for reasons of backward compatibility, all of the characters in ENV 1973:1995. It also contains characters belonging to the Adobe OpenType specification (excepting the characters F000, F001, and F002, which are private use characters). Further, it contains the following characters used for official or regional languages in Europe, and the EURO SIGN.

0218 LATIN CAPITAL LETTER S WITH CEDILLA (Romanian)
0219 LATIN SMALL LETTER S WITH CEDILLA (Romanian)
021A LATIN CAPITAL LETTER T WITH CEDILLA (Romanian)
021B LATIN SMALL LETTER T WITH CEDILLA (Romanian)
021E LATIN CAPITAL LETTER H WITH CARON (Romany)
021F LATIN SMALL LETTER H WITH CARON (Romany)
02DC SMALL TILDE
0400 CYRILLIC CAPITAL LETTER IE WITH GRAVE (Macedonian)
040D CYRILLIC CAPITAL LETTER I WITH GRAVE (Macedonian)
0450 CYRILLIC SMALL LETTER IE WITH GRAVE (Macedonian)
045D CYRILLIC SMALL LETTER I WITH GRAVE (Macedonian)
1E9B LATIN SMALL LETTER LONG S WITH DOT ABOVE (Irish Gaelic)
20AC EURO SIGN
25A1 WHITE SQUARE
25AA BLACK SMALL SQUARE
25AB WHITE SMALL SQUARE
25CF BLACK CIRCLE
25E6 WHITE BULLET

Annex E. Rationale for inclusion of characters in the MES-3 (Informative)

The MES-3 includes, exhaustively, the collections of ISO/IEC 10646-1 characters containing the Latin, Greek, Cyrillic, Armenian, and Georgian scripts, together with those collections of symbols used academically, commercially, and scientifically in Europe. By including combining characters and phonetic characters of the Latin alphabet (including the International Phonetic Alphabet), the MES-3 also provides for the needs of transliteration and transcription of many of the world's languages into the Latin script, as well as the needs of European publishers who routinely make use of the I.P.A.

Substantial advantages can be expected from the specification and implementation of the MES-3 of ISO/IEC 10646-1, defined here as a guide to implementors, which includes a large number of characters and covers a set of historically-related alphabetic scripts of singular cultural importance to Europe.

The MES-3 provides a European specification for the needs of more specialized groups in government, industry, publishing, academia, and the private sector, than are provided for by the less general MES-1 and MES-2.

The MES-3 will help European developers implement all characters belonging to European scripts. This 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. The non-European scripts are being provided for by non-Europeans.

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

NOTE: The MES-3 is a comprehensive way of migrating to the UCS. The MES-3 can be seen as a way of saying to vendors how to provide the UCS for Europe. First, include the European scripts. Then, include the scientific and publishing characters. Basically, the MES-3 is the UCS minus the non-European scripts – support for which (Arabic, Hebrew, Korean, etc.) can be purchased easily enough as the non-European scripts are being handled in the countries of the world most concerned with them. The MES-3 ensures that government, education, publishing, science, etc. have all the tables they are likely to need: the European scripts, and the international punctuation, dingbats, scientific characters, etc.

One can think of the MES-3 as the "modular" UCS specification for Europe. It is likely that "language kits" will be implemented for the UCS. If one needs Korean, one will buy the Korean language kit, which will have characters, sorting, etc. If one needs European scripts, one will buy the European language kit, which will have the MES-3.

Annex F. Bibliography (Informative)

CEN/TC304 N283 Mapping tables from various coded character sets used in the industry (basis for a technical report).

ENV 12005:1995 Information Technology – Procedure for European registration of cultural elements.

ISO/IEC 7498-1:1994 Information technology – Open Systems Interconnection – Basic reference model: The basic model (second edition).

ISO/IEC 8822:1994 Information processing systems – Open Systems Interconnection – Connection oriented presentation service definition.

ISO/IEC 9646-1:1994 Information technology – Open Systems Interconnection – Conformance testing methodology and framework – Part 1: General concepts.

RFC-1345 "Character mnemonics and character sets" (Internet mapping tables).

RFC-1700 "Assigned numbers" (Internet IANA references to code tables).

UNICODE specification (The Unicode Consortium. 1992. *The Unicode Standard.* Version 2.0. Reading: Addison-Wesley. ISBN 0-201-48345-9.

Annex G. Characters in the Pipeline (Informative)

The following characters are currently in the pipeline for inclusion in the UCS:

To be supplied. Skolt Cyrillic and other characters.