

Universal Multiple-Octet Coded Character Set
International Organization for Standardization

Doc Type: Working Group Document
Title: Registration of the Multilingual European Subsets in ISO/IEC 10646
Source: Ireland and Finland
Status: National Body request
Date: 1998-09-22

A CEN ISSS (Information Society Standardization System) workshop is about to formally reach a CEN Workshop Agreement (CWA) on three multilingual European Subsets of the UCS.

These subsets are:

- 1) The repertoire of EN 1923 LL8, which is a limited subset of Latin characters defined for the purposes of interoperability. This is a fixed collection.
- 2) The MES-1 repertoire, primarily for business and administration use, particularly by pan-European organizations in their multinational/multilingual applications. The repertoire is defined for use as the set to which specific processing rules can be tailored and applied. This is a fixed collection.
- 3) The MES-2 repertoire is intended to cover the needs of the languages of Europe, although limited to contemporary left-to-right scripts of European origin (but not to characters used in Europe). This is an open collection with a snapshot of it also being registered as a fixed collection.

Draft versions of the specifications are attached. These are not expected to change greatly, but may undergo minor changes in the final balloting process of the CWA, which will be completed well before the next SC 2/WG 2 meeting.

4. The Repertoire of EN 1923 LL8

4.1 Specification for the Repertoire of EN 1923 LL8

Collection Name:	EN 1923 LL8
Type of Collection (Fixed or Open):	FIXED
Number of characters	333

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
21	22 26 5B-5E 90-93
26	6A

4.2 Collections containing the repertoire of EN 1923 LL8

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

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
36	LETTERLIKE SYMBOLS	2100 - 214F
47	MISCELLANEOUS SYMBOLS	2600 - 26FF

WG24

5. The 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 00-7F 8F 92 B7 C4 C6 C7 C9 CA CC DE DF E4-E9 EE EF F1 F3-F5 FA-FF
02 18-1B 1E 1F 59 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 70 74-8E A3 A4 A7 AC
21 02 05 15-16 1A 1D 22 24 26 2E 5B-5E 90-95 A4-A8 D0-D5
22 00-09 0B-0C 0E-0F 11-13 19 1A 1D-1E 1F 27-2B 3C 45 48-49 58 5F 60 61 64 65 6A-6B 82-89 95 A2 A8 AC-AD
23 00 02 08-0B 10 15 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-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 FC-FD

5.2 Collections containing the MES-1

The following UCS collections from Annex A of ISO/IEC 10646 contain the characters of 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
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
70	Specials	FFF0-FFFF

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

6.1 Specification for the Multilingual European Subset No. 2A

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

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	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

OPEN

6.2 Collections containing the MES-2A

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

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 IPA
7	Combining Diacritical Marks	0300-036F	used with IPA and European scripts generally
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

Multilingual European Subsets of ISO/IEC 10646-1

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

7. The Multilingual European Subset No. 2B (MES-2B)

7.1 Specification for the Multilingual European Subset No. 2B

Collection Name: MES-2B
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	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-2B

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

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 IPA
7	Combining Diacritical Marks	0300-036F	used with IPA and European scripts generally
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