

AUSTRIA

Approval with comments

On: A.16.

heading: »letter with diacritical mark« (not: marks.)

[Reaction] Rejected as MES-2 contains a very small number of letters with more than one diacritical mark, for example the letter á (U01FB). The definition of the base repertoire is out of the scope of this standard

On: A.8.2,

Table 3 – Equivalent forms: The clusters dz, DZ; dz, DZ; ij, J; lj, Lj; nj, NJ; in the list seem to be just a random sample of diagraphs which in a given spelling system rank as »letters« for the purpose of alphabetic ordering. There are more such cases which are equally qualified as candidates for inclusion in the list (e.g., Czech ch, Hung. cs, sz). Therefore either delete the above mentioned, or try to find out all others.

[Reaction] Rejected as such, as this Standard does not specify characters and just orders those specified in the MES Standard. However, additional text is clearly needed, as the comment is made by other NIMBs, too, and the present draft is obviously too imprecise on the rationale for the selection of characters

On: A.1.4, Note 9:

Reference to Dal's Russian dictionary (1880) [correct the spelling:does not seem relevant for the subject, i.e. a comprehensive list of European cyrillic letters.

[Reaction] Agreed. All footnotes beginning with <DRAFT> (in the previous version with <NOTE>...</NOTE>) are only there to facilitate reviewing and will be removed for the final version (cf. footnote 3)

On: A.1.5:

Proposed new wording: »[...] graphical marks placed over, under, or next to a basic letter which modify pronunciation or some other feature of the basic letter. [...]

Comments:

- a) »Structure« is an abstract relationship between items. However, diacritical marks are palpable (visible) dots, strokes etc. Thus: »marks« (perhaps: »signs«).
- b) Instances of a diacritical mark place »through« (across) a letter are being classified as »modified letters« under A.1.8, A.8.3/Table 4. Therefore cancel »through«.

[Reaction] Accepted

- c) »[...] not significantly modify the shape [...]« – wording is not sufficiently explicit and meaningful. Moreover, the word »meaning« does not adequately describe any of the functions of a letter as member of an alphabet. Also, it is not clear what is to be understood by »a valid letter«. – Delete starting from »does not« until including »structures«.

[Reaction] Agreed in principle. The passage will be reformulated

BELGIUM

Approval without comments

CZECH REPUBLIC

Approval with comments

On: 3

In clause 3 Normative references: the numbers for ISO/IEC standards should be written »FDIS« and »FCD« (and in the notes »to be published«).

[Reaction] Accepted. Normative references will be presented according to CEN rules.

On: 4

In clause 4 Definitions: it should be more legible to write here the whole definitions texts.

[Reaction] Agreed. Other NMB's make this point, too. The relevant definitions are included.

On: 5

In clause 5 Conformance: it should be more legible to write here the whole conformance clause.

[Reaction] Agreed in principle. The relevant part of the conformance clause of ISO/IEC FDIS 14651 has been included.

DENMARK

Disapproval with comments

On: General

The specification is based on standards that are not fully approved at this time, such as ISO/IEC 14651 (at FCD stage) and the European subset (MES-1).

[Reaction] Noted. MES-1 will hopefully soon have stabilized. It must, however, be ascertained if a CWA can formally be referenced in an ENV. If not, other ways of reference must be found (including the option of using the repertoire of ENV 1973:1995 as the original Terms of Reference suggests)

The specification does not take into account the specification on the area that was previously sent to TC304 for comment for enquiry. This document was sent to TC304 in October 1997. The consensus made on that document should be utilized, preferably by using that document as the base document.

[Reaction] Rejected. N747 played an important rôle in the creation of the EOR-draft. Otherwise, the Danish comment is at variance with the decisions of TC304.

On: Annex A

This is predominant in annex A, where there are problematic definitions like »letter«, »basic letter«, »small letter«, »equivalent letter form«.

[Reaction] Accepted in principle. Some of the definitions will be reformulated in line with concrete proposals of other NMBs

On: General

The letters „ , and , should be treated as basic letters for this standard to be adopted in Denmark.

[Reaction] In the email, no letters are shown here. This is just one of the quirks of Internet email: the upper-ASCII letters were just stripped out somewhere. The project team assumes that the LETTER AE is meant.

Rejected as such. Any NMB will have to declare a delta to suit national customs if it desires to base any formal specification on the EOR (which are specifically targeted at multilingual European ordering). Æ and æ are not treated as one letter throughout Europe. It is, however, likely that the treatment of Æ and æ in the final draft will be more in line with Danish expectations.

On: A.3.3.3

A.3.3.3 There should not be any concept of canonical equivalence, as this is an ill thought out concept, and not necessary for this standard that does not use combining characters.

[Reaction] Formal references to the concept of canonical equivalence can be removed if desired, though the draft will ensure that no artificial incompatibilities to Unicode are introduced.

On: General

There are also a number of differences in the sorting specifications from ISO 14651, for example AE OE and ETH are sorted differently in EOR and 14651.

[Reaction] Further harmonization of these tables is likely. However, note that even Danish specifications of this are different too from the emerging international and European standards. In the Danish specification for sorting, as specified in ISO/IEC 15897 (adapted from ENV 12005) note that Danish sorting requirements differ from the international equivalents. Although AE is treated as a letter, THORN is treated as filing as T and H – see ISO/IEC 15897 (ISO/IEC JTC1/SC22/WG20 N 610), page 16, clause 11, and page 29, clause E6. If there are valid reasons for the European Ordering Rules to differ from ISO/IEC FCD 14651 and the Unicode Ordering Algorithm tables used to autogenerate Annex A of ISO/IEC FCD 14651, these will be so specified.

On: General

All in all the draft is immature

[Reaction] Rejected

On: General

and needs further work.

[Reaction] National member bodies ensure by their comments that the draft will be optimized.

FINLAND

Approval with comments

On: General

Special attention needs to be paid throughout the process to ensure harmonization and interdependencies as per the stated objectives.

[Reaction] Accepted

ICELAND

Approval as presented

GERMANY

Approval as presented

GREECE

Disapproval with comments

[Reaction] No comments received to date

IRELAND

No (with comments)

On: General

Ireland votes No on the suitability of this document to go out for TC-enquiry. It is poorly typeset and at the very least if fonts are not available then glyphs must be drawn in by hand.

[Reaction] Noted. The editor disagrees on the topic of typographic quality, but that is a matter of personal taste and has little relevance here.

Glyphs will be added if it can be done without disproportionate amount of work and cost.

Note: MES draft 4 has no glyphs shown! Nor do ISO/IEC FCD 14651 or the Unicode Ordering Algorithm tables.

On: General

We object to the format aAbBcCdD, and request that the format AaBbCcDd be used, as it is natural and easier to read and the format presently used is unnatural and confusing to the ordinary reader.

[Reaction] Noted.

On: General

Further, the ordering is intended to be based on MES-1, which has not been finalized. So how could it possibly be ready for TC-enquiry.

[Reaction] It has always been the intention to synchronize the repertoire for the EOR with that of MES-1 as soon as that is stable and if it is formally possible.

The logic of the EOR as explained in Annex A gives a unique place to each of the new characters in the 4th draft of MES-1. Tables will be adjusted for the third draft

NETHERLANDS

Disapproval with comments

On: General

The draft is considered useless in this form, and not suitable for development into a standard.

1. The ordering rules are being given in a syntax that will deter readers not accustomed to POSIX notation. Potential users, like publishers of lexicons, encyclopedias, gazetteers or name lists will not be prepared to decipher information presented in this form, and thus will ignore the standard. (It is open to doubt anyway that they will adapt their present rules to those specified in this draft.) There is no evidence that they have been consulted.

[Reaction] Rejected. The project team is very well aware that many users of ordering are not versed in formal notation and that the main part of the standard may well be unreadable for them. The Netherlands comments do not show any awareness of the introduction where this problem has explicitly been discussed. The annexes are nothing but a plain-English variant of the tables and this is clearly so stated. Nobody is forced to even look at the formal specification which was, however, demanded by industry representatives.

Members of the project team are lexicographers and/or are in close contact with many users in the field of the publishing of »lexicons, encyclopedias, gazetteers and name lists.« References can be provided if so desired.

On: Annex C and note 21 on page 31

2. The draft confuses the concept of character with that of graphic symbol (for example in note 21 on page 31, and the whole of Annex C).

[Reaction] Disagree. Strictly speaking, Trema and Umlaut are two distinct characters (though not so treated in ISO/IEC 10646), originating from different roots and carrying very different functions in German. That they are therefore traditionally distinguished typographically is only a consequence of this.

Informative Annex C reflects widespread practice in lexicography. The typographic distinction serves in this context as a markup.

On: A.1.5 to A.1.8

It introduces distinctions not found in any ISO/IEC coded character set standard (definitions A.1.5 to A.1.8).

[Reaction] Noted. These distinctions are very common in »traditional« ordering standards. However, a rewording may be useful and might solve some of the problems.

On: A.1.6 and A.1.8

In particular, that between modified letter and letter with a diacritical mark is highly artificial, and cannot be made precise in several cases (see Note 10 on page 24).

[Reaction] The distinction is widely used in »traditional« ordering standards and there has proven its viability. Furthermore, it is a very well known concept to the publishers of »lexicons, encyclopedias, gazetteers and name lists.« whose support the Netherlands comments rightly demand. Perhaps text could be improved so that the distinction is more evident.

On: A.4.2.2 and A.4.2.3

Moreover, it is not functional, as is evident from A.4.2.2 and A.4.2.3.

[Reaction] We do not understand the comment

On: A.4.2.2, A.4.2.3 and A.5.3.2

Both should be called »letter variants«, and an order specified by enumeration. This would also make A.5.3.2 superfluous.

[Reaction] Noted. The term »letter variants« needs further discussion within the project team. To avoid unfair discrimination, the terms »first level letter« and »second level letter« have been chosen. Cf., however, remark above.

How should an order specified by enumeration be realized? It would be unfruitful and at variance with practice in »traditional« ordering standards

if all »letters with diacritical marks« and »modified letters« were enumerated instead of the diacritical marks themselves.

On: Terminology

Terminology must conform strictly to that used in ISO/IEC standards (e.g. the term »canonical equivalence« shall not be used, nor any other from the usage in Unicode, if not occurring in ISO/IEC 10646).

[Reaction] Accepted in principle

On: General

3. The draft cannot proceed before the repertoire to be covered has been fixed.

[Reaction] Accepted in principle

On: General

It now includes still presentation forms, like fi and ff which are not letters.

[Reaction] Noted. These are part of the repertoire which EOR required to sort. They have to be accommodated in ordering as much as AE, OE and IJ. The problem is thus inherent in the base repertoire.

On: A.1.4[!] and A.4.3.4

4. The list of basic Cyrillic letters (A.1.5 and A.4.3.4) is incomplete. Missing are GJE, DZH, KJE, DJE, IO, YI.

[Reaction]

GJE (Ѣ, U403 and U453) is currently treated as a basic letter in the EOR-tables and as a letter with diacritical mark in the annexes. While the resulting ordering is identical, this inconsistency will be corrected. The omission of DZHE (Ѣ, U040F and U045F) in the annexes is an error and will be amended
KJE (Ѣ, U040C and U045C) is currently treated as a basic letter in the EOR-tables and as a letter with diacritical mark in the annexes. While the resulting ordering is identical, this inconsistency will be corrected.

The omission of DJE (Ѣ, U0402 and U0452) in the annexes is an error and will be amended

IO (Ѣ, U0401 and U0451) is currently treated as a basic letter in the EOR-tables and as a letter with diacritical mark in the annexes. While the resulting ordering is identical, this inconsistency will be corrected.
YI ((CYR YI SM), U407 and U457) is currently treated as a basic letter in the EOR-tables and as a letter with diacritical mark in the annexes. While the resulting ordering is identical, this inconsistency will be corrected.

The editor wishes to thank the Netherlands for pointing out these serious shortcomings

On: General

5. See general NL comment: Can it be explained what in the 7 default table under Greek the digamma, stigma and koppa are doing?

[Reaction] The default table has entries for digamma, stigma and koppa, though the following tables do not map to these. They are thus not operational and do not influence the result.

If these are valid characters in the MES repertoires, these will need to be sorted. Recent discussion on the MES list indicates a genuine need at least for stigma and koppa (and sampi: these are numbers 6, 90 and 900 respectively in the Greek numeral system, and are still used in citing Greek laws which are still in force.

For example, the basic law concerning damages and criminal charges from car accidents in Greece is the Act »Gamma-Sampi-Nu« (0375-03B3-03E0-03BD-0374) which was passed in 1911 and is still in force, according to Constantine Stathopoulos, writing on 1998-10-19. It appears that these will get added to the repertoires now called MES-1 and MES-2, according to the MES editor. They should sort in terms of their historically expected sort order.

NORWAY

Approves without comments

PORTUGAL

Abstention

SPAIN

No vote received to date

Information from Þorgeir Sigurðsson:

I phoned Spain, and asked why they had not given us reply to N841! They asked for a week, in the meantime they will have a meeting and after which we will get their reply.

I said that was OK,
My regards
Þorgeir

SWEDEN

Approval with comments

On: General

Introduction [general comment] »... for the people of the European Community and the European Free Trade Area ...«

The text should refer to Europe and European organisations in general. We should not limit the definition of Europe to some specific organisations. The political map of Europe is altered almost every year. CEN has members that isn't covered by this definition and – vice versa – some countries that are covered by the definition are not members of CEN.

[Reaction] Accepted in principle

On: 1 Scope [general comments]

»The ordering rules given here ...« The ordering rules could also be applied when the target is a multicultural audience, even if all the data have a common single-cultural origin. It is not the data origin that defines the suitable ordering rules, it is the source data and the target audience. (E.g. a list of Swedish names could be better presented using an European ordering rule for a pan-European audience that doesn't know the Swedish language.)

[Reaction] Accepted. The scope will be formulated more precisely.

On: Introduction

»They are not meant to replace or influence existing ...«

This sentence conflicts the statement in /6 Tailorability/. The ordering rule could be used as a base for all national standards defining order between characters and scripts not used natively. (E.g. a Swedish ordering rule should be able to use European ordering rules for ordering letters only used in e.g. Greek, Maltese and Serbian.)

[Reaction] Accepted. The introduction will be rewritten to take this into consideration. However, it must be clear that no existing national practices are in any way challenged.

On: A.1.9

In Annex A, 1.9 add before modified letter, equivalent letter form as CAPITAL, i Table 3.

[Reaction] Accepted

On: A.1.4 and A.4.3.4

Annex A (informative) Exposition of relevant principles [technical comments]

A.1.4 basic letter

A.4.3.4 Cyrillic script

The list of basic Cyrillic letters doesn't match the ordering described in /7 Default table/.

Most of the Cyrillic letters origin in Old Slavonic with unaltered shape, use and order.

Many of the remaining letters have diacritical marks (SHORT I can be considered as I with breve), are equivalent forms (NJE can be considered as a combination of EN and SOFT SIGN) or modified letters (GHE WITH UPTURN can be considered as a modified GHE).

The Cyrillic letters (basic letters, letter with diacritical marks, equivalent forms and modified letters) in MIES-2 are usually, with the exception of IO in Russian, ordered as independent basic letters. Make the definition for Cyrillic basic letter consistent with that (cf. final sigma in Greek).

[Reaction] Accepted. For a detailed discussion please cf. remarks on the NL comments

On: A.3.3.2 Equivalent letter forms

See SE comment to A.3.3.2

[Reaction] ?

On: A.8.2 Equivalent forms

Most of the equivalent forms listed are not just a joining of letters. The difference between modified letters and equivalent forms seems to be mainly based to the graphical design – not the usage of the characters. Correctly used they can carry additional information.

E.g. LATIN LETTER NJ would not be used in injekcija (composed of injekcija), but would be used in njiva. This is usage similar to LATIN LETTER

SHARP S (e.g. in Mae / Masse), but this character is regarded as a modified letter and this would affect the secondary ordering.

This difference in equivalent should affect the secondary ordering in the same way as modified letters and diacritical marks.

Two strings should not be ordered fully equivalent, if they are not a binary equal.

[Reaction] Accepted in principle.

The problem is that Table 3 is composed of different kinds of »equivalent forms«: pure ligatures like »fi« which should certainly not be distinguished from »f+i« on any level, forms such as Æ (AE) which are often regarded as modified letters on level 2 (and which we will be so treat in the final draft) and letters such as »DZ« which are simply transliterations of their equivalent Cyrillic letters (ж in this case and ъ in the case of »nj«) and which ought to be treated as single letters in this context. The latter is, however, not understandable to European users not aware of this fact, and that are likely to be the vast majority. For those, the treatment of Table 3 is the natural one.

Fortunately, MIES-1 does not any longer contain »DZ«, »LJ«, »NJ« and their combinations with diacritical marks and thus the problem disappears.

On: A.5 Secondary ordering level

A.5.1 No unique sequence after the primary ordering level

»The secondary ordering level is divided into two parts: ...«

The current text suggests that a modified letter occurs in position M in string S, it would cause string T (with a diacritical mark in position N) to precede string S even if N < M. This is not the case in /Default table/.

[Reaction] Accepted. The wording will be changed to remove this ambiguity

ADDITIONAL SWEDISH COMMENTS

The argument that the current EOR-1 separates non-MES-1 letters from their MES-1 counterparts is a strong one for EOR-1 (based on the still valid MES [ENV1973] and soon on MES-1), but not valid any more for the planned extension EOR-2 which will be based on MES-2, i.e. on a script-based character set. However, the Swedish comment could partly be accommodated by adding a suitable informative annex.

On: Section 7

I [Kent Karlsson] briefly saw the EOR draft at the Tel Aviv WG20 meeting (October 18-22). However, I noticed that while saying that it was a 'delta' from 14651, it is a very big 'delta'. It enumerates all characters in (one of) the MES(?), giving them weights. While formally a 'delta', I find it to defy the idea of having 'deltas'. A 'delta' should be as small as possible. While, for instance, rearrangement of some Asian characters may still result in a very large 'delta', that seems unnecessary for Latin/Greek/Cyrillic compared to the international template. A 'European' delta (for Latin/Greek/Cyrillic) should only mention the very few instances where the international template is really at odds with what is desirable in a pan-European context. I would guess that a true 'delta' would need to mention only a handful of characters, giving them new weights. For instance sorting ð (eth) as a variant of d (same level 1 weight) rather than as a separate letter (just after d at level 1), and some "very few" others.

Thus, instead of having a 20-30 (or more) pages of 'delta' description, significantly less than a quarter of a page for the 'delta' description itself should be sufficient for all the differences from the international template that one can arguably care about in a pan-European context.

The current 'large delta' also separates non-MES(?) Latin letters from the MES(?) Latin letters. Thus, if a non-MES(?) Latin letter occurs, it will be sorted far away (before a, if I remember correctly from my glance) from closely related Latin letters. This is a defect that the international template does not have, and a very small 'delta' would not have that defect either.

[Reaction] Noted. The project team is, however, inclined to keep the current large delta in section 7 for the following reasons:

The tables must always be in line with the informative annexes (which for a great many people not versed in »computerese« will be the »real« standard). However, the tables of 14651 are still in flux and if only a small delta is given, the EOR tables and the annexes could at any time cease to be in sync which would invalidate the idea of the annexes. This reason was, for us, the compelling one.

A small delta is helpful for somebody with a copy of 14651 on his/her desk. This is unlikely to be the case for most users of the EOR. Furthermore, other NMBs have demanded that the EOR should be able to stand on its own to an even larger extent than now. A small delta would be difficult to align with these demands.

On: Section 7

In addition, I would still like to push for a strong recommendation (normative, if possible) of doing a preparation of the strings such that at least numerals for natural numbers in base ten are ordered in a way consistent with the natural number values they represent. See the 'preparation' annex of the current (late October) 14651 draft, and also Swedish Statskontorets alphanumeric sorting norm. (I don't know if there is such a recommendation in the current EOR draft.)

[Reaction] Accepted. A strongly formulated informative recommendation will be added in this matter.

UNITED KINGDOM

Approval with comments

On: 1. Foreword (to be added to the standard)

Include the following in the foreword:

This European standard consists of two main parts: a descriptive part aimed at end users (currently contained in sections 1–6, with much longer informative annexes) and a formal specification, designed to be used directly within computer programs (currently in Section 7). This formal specification provides a Default Table, which can be tailored through the use of the provisions of ISO/IEC FCD 14651.

[Reaction] Agreed. A wording similar to the one suggested will be incorporated into the introduction which shall then be renamed to »Foreword«

On: Sections 1–6

Delete reference to Unicode and replace with reference to ISO/IEC 10646–1.

In Section 2 (Notation) Unicode is referenced. As Unicode is not normative (it is referred to only in the bibliography) it should not be referenced in this European standard. ISO/IEC 10646 is normative, and this on its own can be referenced, especially as it uses the same notation as Unicode.

[Reaction] In the present draft the Unicode standard is normatively referenced (cf. section 3). It is, however, likely that references to Unicode will not be retained in the final draft.

On: Annexes A and B

Annexes A and B should be incorporated into the body of the standard, prior to section 7, and made normative by being in the body of the text.

[Reaction] Noted. However, in this case essentially identical rules would be twice in the standard in different formulation and with different definitions. The annexes are meant as a tool to facilitate access for users not versed in the formal language of ISO/IEC FCD 14651 (probably the majority) and should thus best be regarded as an reliable and authoritative interpretation of the standard's body.

On: Section 4 Definitions, section 5 Conformance, and section 6 Tailorability.

These sections are too short and should be expanded. The standard should be able to stand on its own.

[Reaction] Noted. This comment has also been made by other NIMBs. Cf. the relevant points in the response to the Czech comments

On: A.1 Definitions

In particular, section A.1 Definitions, should go in section 4 of the main text of the standard. Some of the definitions used should follow definitions used in other standards.

[Reaction] Noted. Cf. comment above

On: Section 6

Section 6 (Tailorability) needs to be retitled, and rewritten in a clearer and more detailed fashion.

The title should not be metaphorical. An alternative is »Versions« or »variants«. It is incorrect to suggest that users can »tailor« this standard when claiming that their product conforms to it, but they could develop and publish an identified version or variant of it, and then claim conformance to that.

[Reaction] Rejected. »Tailoring« is a $\langle \rangle$terminus technicus</math> for this procedure with well entrenched usage. It ought to be retained

On: 3. Section 7 Default table

Word by word ordering should be the default, reflecting user's default expectations of computer operating systems.

It is anticipated that the table will be compatible with the equivalent table of ISO/IEC FCD 14651.

[Reaction] Rejected. Word-by-word ordering is a case of multi-key ordering. EOR specifies rules for one key. However, this might be explained more clearly.

On: 4. Annexes A and B

As noted above, under Sections 1–6, Annexes A and B should be incorporated into the main body of the standard.

[Reaction] Noted. Cf. answer to comment on section 6

On: 5. Annexes C and D

Delete Annexes C and D. These areas are not sufficiently standardised. It may be useful for this standard to refer to other standardisation work in this area however.

Sorting by style, described in Annex C, is relevant to the specialist area of lexicography and terminology, and thus out of the scope of the EOR standard. However, in the body of the text (e.g. in the Scope) reference can be made to ISO 12199, which does cover that, rather than describing this in the European Ordering Rules.

Annex D should be deleted: this topic could be covered by a phrase in the body of the text stating that if sorted data includes some text which had been transliterated from another script, once the sorting has been done the original can still be used display or print the original script: however such rendering is outside the scope of this standard.

[Reaction] Noted.

On: 6.1 Notes

The number of notes should be reduced.

As the MES CWA draft was not yet available, it should be noted (only in this draft) at the beginning that all tables in this draft are based on the Minimal European Subset (ENV 1973:1995). This can be amended when the MES CWA is available and agreed, and this note can then be removed from this standard.

[Reaction] As footnote 3 explains, none of the footnotes in <NOTE> / </NOTE> (the great majority) is intended for inclusion into the standard and is there to facilitate reviewing only.

On: 6.2 Tables

Brief tables (approximative 1/2 a page) should be included in the current section 6, showing the relevant repertoires of the various MES-repertoires.

There should be some brief general information about languages, or groups of languages, where the preferred sorting order generally differs from that set out in the default table. This was agreed at the EOR meeting at Reykjavik, and could be fairly brief.

[Reaction] First part: Noted. This can be accommodated by including the relevant tables from the MES-CWA.

Second part: The preferred ordering differs in virtually every language from that of the EOR which is a compromise between those. However, an explicit reference to ISO FDIS 12199 and the relevant tables there can be added

On: Format

[JPC] (personal comments noted after UK comments submitted): Tables should follow conventions in ISO/IEC FCD 14651 and Unicode Ordering Algorithm SYMDUMP2.TXT (which is used to autogenerate Annex A of ISO/IEC FCD 14651).

[Reaction] Accepted

