SC22/WG20 N894

SUMMARY OF VOTING ON

Letter Ballot Reference No: SC22 N3265
Circulated by: JTC 1/SC22
Circulation Date: 2001-07-13
Closing Date: 2001-09-15

SUBJECT: Summary of Voting on SC 22 N 3265, Letter Ballot on CEN/TC 304 Request for 30 Cultural Specifications to be Registered with ISO/IEC 15897

The following responses have been received on the subject of registration:

"P" Members supporting registration without comment

5 (Czech Republic, Denmark, Ireland, Norway, Russian Federation)

"P" Members supporting registration with comments

-

"P" Members not supporting registration

3 (Netherlands, United Kingdom, United States of America)

"P" Members abstaining

2 (France, Japan)

"P" Members not voting

11 (Austria, Belgium, Brazil, Canada, China, Egypt, Finland, Germany, Romania, Slovenia, Ukraine)

Note: O-members Republic of Korea and Sweden voted to support the registration.

_____ end of summary, beginning of NB comments _____

United Kingdom

The UK does not support the registration. It considers it is an inappropriate request as the CEN Workshop Agreement 14051-2 has not been validated by the CEN Members.

The Netherlands votes NO on
N3265 Letter Ballot on CEN/TC304 Request for 30 Cultural Specifications to be registered with ISO/IEC 15897
for the following procedural, political and technical reasons.

Procedural:

1. N3265 is submitted to be registered under the regulations as set out in ISO/IEC 15897. The latter document prescribes which organisations (called "Sponsoring Authorities") are allowed to submit specifications for registration. The current version of ISO/IEC 15897 only considers (Associated) Member Bodies of CEN or JTC1, and JTC1 itself and any of its SC's or WG's as Sponsoring Authorities. CEN Workshops and CEN TC304 clearly do not qualify, and hence N3265 should not have been allowed in the process.

2. The Netherlands has no opinion on (the contents of) locales from other countries (language areas). Hence we feel it to be inappropriate to cast a vote on for instance the Italian locale (although ISO/IEC 15897 explicitly allows this to happen). Still we feel that we must be able to vote on the Dutch locale and the Euro locale. Therefore we would like to see the documents separated in different documents on which individual votes are possible.

Political:

1. This documents describes locales for a subset of the languages currently considered to be part of 'Europe'. Choosing languages as the starting point however, raises the problem of finding authorities to agree with these specifications for those countries where the mentioned language is spoken as a majority language and for those countries where the mentioned languages are spoken as one of many languages or even as a minority language. An example may be the Dutch language which is a majority language in the Netherlands, but is one of the three official languages used in Belgium.

2. Rules for formatting monetary data are an important part of those locales. However, locales based upon the Euro have been provided for some countries not switching to the Euro on short notice.

3. Specifications for monetary formatting in the euro locale should be based upon a publicly available document from the ECB at Frankfurt. This document should then be referred to in section 2.

4. In a couple of years Europe is being extended with some countries. It is unclear whether some specifications given here may interfere with specifications needed for these countries. An example may be the use of YyOo/N for affirmative/negative answers. It may be the case that some of these countries have a word for 'yes' starting with the letter 'N'. It seems inappropriate to produce at this moment a set of specifications for a part of western Europe only.

Technical:

1. It is understood that the POSIX specification describes the functionality of an operating system and the interface between application programs and that operating system. It is also understood that this document describes a set of locales, a locale being a very limited user interface; a set of input/output rules for users of a computer running the POSIX operating system. Later developments in operating systems such as micro-kernel operating systems and in software architecture (two and three-tier systems) have led to the solid belief that separating the user interface from the operating system is a worthwile goal in constructing maintainable software. It is therefore questioned whether the specification of locales in connection with an operating system is still a state-of-the-art practice.

2. The document is very unclear about what is input and what is output. It is assumed here that clauses on Numeric formatting, Monetary formatting, Date and time conventions describe output specifications and that clauses on affirmative/negative answers describe input specifications. Given that, the document is unclear about the specific output devices involved, and worse, seems to consider only a keyboard as an input device, therewith foregoing input devices like special purpose devices (like teller machine keyboards), touch screens, speech input or even webcams. Considering the use of POSIX (and derivatives) in embedded systems this seems somewhat of a old-fashioned restriction.

3. Monetary formatting:

(a) The convention prescribed for negative amounts is very uncommon in the Netherlands.

(b) The rules for writing negative amounts are in conflict with the rules given for writing negative numbers. Are negative amounts not negative numbers?

(c) The convention for writing Euro amounts gives poorly organized results when denoting tables. Using the E for the Euro-sign:

```
|  E78,90 |
| E123,77 |
|    E,23 |
```

instead of the much nicer

```
| E 78,90 |
| E123,77 |
| E   ,23 |
```

(d) The latter text line also shows that the rules given are incomplete. How to write 23 eurocents? E000,23 or E0,23 or E,23 or E23?

(e) It is also unclear why only Monetary formatting is so very detailed using e.g. no-break-space.

4. 5.2.11 nl_EU
   The rules given here are unacceptable for the Netherlands as they do not properly handle the Dutch digraphs ij and IJ.

   A short note on how to handle these digraphs is available on request.

5. Affirmative/negative aswers

(a) As indicated above it is highly arbitrary that a locale for the Netherlands should allow the English yY and the French oO (for oui).

(b) The use of a plus and minus sign is to be avoided as these could mean the addition or substraction of an amount in financial applications.

(c) Clause 6 states that "Since these answers are used at the user interface and the question is being asked in a natural language, the answer should be given in the same language". The possibilities for answering are however limited to simple key presses '+' or 'y'. Key presses are not language! Under the real meaning of language, much more elaborate answers should be allowed, e.g. (using the English language) 'yes' or 'surely' or 'certainly' or 'shaken, not stirred' (to the robot bar tender running on a Posix compliant embedded system)

US National Body votes to Disapprove the registration on CEN/TC 304 Request for 30 Cultural Specifications to be Registered with ISO/IEC 15897

Comments:
Please find the comments to this ballot below:

Comments on SC22 N3265
European generic locales - Part 2: Narrative cultural specifications, POSIX locales, and repertoiremap)

The following comments refer to the repertoire map first, the _EU locale, and then the country-specific locales.

****************IN THE REPERTOIRE MAP:**

* The repertoire map says that it is MES-2, but there are multiple characters in it that are not in the official definition of MES-2 (CWA 13873...MES-2). They are:

U02D6 Modifier Letter Plus Sign
U2113 Script Small L
U212E Estimated Symbol
U2215 Division Slash
U2501 Box Drawing Heavy Horizontal
U2571 Box Drawing Light Diagonal Upper Right to Lower Left
U2572 Box Drawing Light Diagonal Upper Left to Lower Right
U25A1 White Square
U25AA Black Small Square
U25AB White Small Square
U25CF Black Circle
U25E6 White Bullet
U25E2 Black Lower Right Triangle
U25E3 Black Lower Left Triangle

These should be removed from the repertoire map.

* The repertoire map should not use the Danish mnemonics. It should use only the Uxxxx identifiers. This would be consistent with ISO/IEC 14651 and with ISO/IEC 10646.

* Near the end of the repertoire map, some characters are repeated, but with different mnemonics. They are:

Character	1st mne.	2nd mne.	
NUMBER SIGN	<Nb>	<H->	
DOLLAR SIGN	<DO>	<!S>	
COMMERCIAL AT	<At>	<@>	(also includes <Oa> as 3rd mne.
CENT SIGN	<Ct>	<!C>	
POUND SIGN	<Pd>	<L->	
CURRENCY SIGN	<Cu>	<Xo>	
YEN SIGN	<Ye>	<Y->	
BROKEN BAR	<BB>	<!B>	
SECTION SIGN	<SE>	<So>	
NOT SIGN	<NO>	<7!>	

PILCROW SIGN <PI> <9I>

These should not be repeated. Remove them.

* At the very end of the repertoire map, there is a group of box drawing characters. Earlier in the map, a larger group of such characters is defined. At the end, it includes the same subset of characters in range the U2500..U253C as were defined earlier, but here adds U2501. It also adds U2571, U2572, U25E2, and U25E3, and then repeats U266A. As noted previously, some of these characters are not part of the official definition of MES-2 and so should be removed, but it also is confusing that part of the box drawing section is repeated. These characters should only be defined once. Remove the extra definitions.

*****************IN THE GENERIC _EU LOCALE:**

* Multiple mnemonics in the locale do not exist in the repertoire map. Latin letters-with-circumflex have names like <A//> in the locale, but in the repertoire map, the naming convention is <A/>>. This error exists in all letter-related classes and within the collation definition. Thus:

In locale	Should be
<A//>	<A/>>
<E//>	<E/>>
<I//>	<I/>>
<U//>	<U/>>
<C//>	<C/>>
<u//>	<u/>>
<c//>	<c/>>

etc., etc.

Not all mnemonics of the form <*//> are wrong. This is the naming convention for letters-with-stroke. Thus, a name like <O//> is correct for the Scandanavian Ø (O-stroke). However, the mnemonic <O//> appears twice in the upper class; first (incorrectly) in attempting to identify Ô (O-circumflex); second (correctly) meaning Ø (O-stroke).

All incorrect mnemonics for letters-with-circumflex in the locale must be fixed. Of course, as noted earlier, the best solution is to use the Uxxxx names to improve consistency with ISO/IEC 14651 and ISO/IEC 10646 rather than these extremely error-prone mnemonics.

* There also are errors with the Greek mnemonics not matching the names in the repertoire map. This includes any name that starts with <A*/ or <W*/ or <Y*/ (e.g., <A*/;!J> or <W*/;J> or <Y*/;?J>). These probably should not have the slash in them; the probably-matching names in the rep. map are <A*;!J> or <W*;J> or <Y*;?J>. A better solution is to use the Uxxxx names rather than the error-prone mnemonics.

* The LC_COLLATE section defines collating symbols <a8>..<z8> for use in defining the last character in a group of Latin letters. However, it also uses <th8>, but does not define it as one of the collating symbols.

* What authorities provided the Greek collation order?

* What authorities provided the Cyrillic collation order?

* ISO/IEC 14651 lists control characters and ASCII/Latin-1 punctuation first in the common template. The generic _EU locale lists them after all the Latin, Greek, and Cyrillic characters. Although they will sort to the same location, since in both documents they are ignored on the first three passes, it would be clearer to duplicate the 14651 order within the source file.

* There is an error in the LC_TIME list for the last month abbreviation:

```
   abmon "<0><1>";"<0><2>";"<0><3>";"<0><4>";/
          "<0><5>";"<0><6>";"<0><7>";"<0><8>";/
          "<0><9>";"<1><0>";"<1><1>";"<D><2>"
```

That should be <1><2>, not <D><2>.

* The Danish mnemonics in the LC_MESSAGES section is particularly obscure:

```
   LC_MESSAGES
   yesexpr "<<(><+><)/>><.><*>"
   noexpr "<<(><-><)/>><.><*>"
   END LC_MESSAGES
```

It would be more helpful for the people trying to read and understand this source file if the Uxxxx identifiers were used and a comment explaining the meaning was added.

*******************For the country-specific locales:**

* All country-specific locales use the base collation definition defined in the _EU locale's LC_COLLATE section. The need for a pan-European collation definition is recognized, and there are no objections to the way it has been defined in the _EU locale. However, it seems quite inappropriate to use the pan-European collation in all of the country-specific locales without tailoring.

The LC_COLLATE section collates letters-with-diacritics with the base characters. Thus, letters like å, ø, ä (a-ring, o-stroke, a-diaresis) and others sort with the a's or o's. What Danish user would think it correct to sort æ, ø, and å (ae, o-stroke, and a-ring) with the a's, o's, and a's, respectively, rather than after z, as is the case with Danish? Or how is it useful for Finnish to sort å, ä, and ö (a-ring, a-diaresis, o-diaresis) with the a's and o's, rather than after z? Also, there are no collating elements for the Spanish ch and ll. What Spanish speakers would agree the default collation is correct?

Perhaps the argument is that this generic locale is for pan-European support, and each country is giving up a bit of its specific requirements for consistency across Europe. But if that is the case, why are there still language-specific names for months and days in the country-specific locales? For example, what Swedish user will *want* to see Swedish words in a date, want to use non-Swedish rules for å, ä, and ö? Either locales are generic or they aren't. These are a combination of both and will probably cause the most confusion for users.

Country-specific locales should be changed to include appropriate tailoring for collation to match language-specific expectations.

end of comments