ISO/IEC JTC 1/SC22
Programming languages, their environments and system software interfaces
Secretariat: U.S.A. (ANSI)

ISO/IEC JTC 1/SC22
N2917

TITLE:

Summary of Voting on Second FCD Ballot for FCD 14652
Information technology - Programming languages, their environments
and system software interfaces
Specification Method for Cultural Conventions

DATE ASSIGNED:
1999-04-28

SOURCE:
Secretariat, ISO/IEC JTC 1/SC22

BACKWARD POINTER:
N/A

DOCUMENT TYPE:
Summary of Voting

PROJECT NUMBER:
JTC 1.22.30.02.03

STATUS:
WG20 is requested to prepare a Disposition of Comments Report and make a
recommendation on the further processing of the FCD.

ACTION IDENTIFIER:
FYI

DUE DATE: N/A

DISTRIBUTION: Text

CROSS REFERENCE: N2869

DISTRIBUTION FORM: Def

Address reply to:
ISO/IEC JTC 1/SC22 Secretariat
William C. Rinehuls
8457 Rushing Creek Court
Springfield, VA 22153 USA
Telephone: +1 (703) 912-9680
Fax: +1 (703) 912-2973
email: rinehuls@radix.net

SUMMARY OF VOTING ON

Letter Ballot Reference No: SC22 N2869
Circulated by: JTC 1/SC22
Circulation Date: 1998-12-24
Closing Date: 1999-04-26

SUBJECT: Second FCD Ballot for FCD 14652 - Information technology -
        Programming languages, their environments and system software
        interfaces - Specification Method for Cultural Conventions

----
The following responses have been received on the subject of approval:

"P" Members supporting approval
    without comment   7

"P" Members supporting approval
    with comment   1

"P" Members not supporting approval   4

"P" Members abstaining   3

"P" Members not voting   7

"O" Members supporting approval
    without comment   1

"O" Members not supporting approval   1

Secretariat Action:

WG20 is requested to prepare a Disposition of Comments Report and make a
recommendation on the further processing of the FCD.

The comment accompanying the abstention vote from France was: "Due to
lack of resources."

The comments accompanying the affirmative vote from Denmark are attached
along with the comments accompanying the negative votes from Germany,
Japan, Sweden, the United Kingdom and the United States of America.

ISO/IEC JTC1/SC22 LETTER BALLOT SUMMARY

PROJECT NO: JTC 1.22.30.02.03

SUBJECT: Second FCD Ballot for FCD 14652 - Information technology -
          Programming languages, their environments and system software
          interfaces - Specification Method for Cultural Conventions

Reference Document No: N2869   Ballot Document No: N2869
Circulation Date: 1998-12-24   Closing Date: 1999-04-26

Circulated To: SC22 P, O, L   Circulated By: Secretariat

SUMMARY OF VOTING AND COMMENTS RECEIVED

	Approve	Disapprove	Abstain	Comments	Not Voting
'P' Members					
Australia	( )	( )	(X)	( )	( )
Austria	( )	( )	( )	( )	(X)
Belgium	( )	( )	( )	( )	(X)
Brazil	( )	( )	(X)	( )	( )
Canada	(X)	( )	( )	( )	( )
China	( )	( )	( )	( )	(X)
Czech Republic	(X)	( )	( )	( )	( )
Denmark	(X)	( )	( )	(X)	( )
Egypt	( )	( )	( )	( )	(X)
Finland	(X)	( )	( )	( )	( )
France	( )	( )	(X)	(X)	( )
Germany	( )	(X)	( )	(X)	( )
Ireland	( )	( )	( )	( )	(X)
Japan	( )	(X)	( )	(X)	( )
Netherlands	(X)	( )	( )	( )	( )
Norway	(X)	( )	( )	( )	( )
Romania	(X)	( )	( )	( )	( )
Russian Federation	(X)	( )	( )	( )	( )
Slovenia	( )	( )	( )	( )	(X)
UK	( )	(X)	( )	(X)	( )
Ukraine	( )	( )	( )	( )	(X)
USA	( )	(X)	( )	(X)	( )
'O' Members Voting					
Korea Republic	(X)	( )	( )	( )	( )
Sweden	( )	(X)	( )	(X)	( )

Comments from Denmark

From: Pia Junker Hviid <ph@ds.dk>

Subject: Danish vote on SC22 N 2869 – FCD 14652

We can inform you that the Danish vote on SC22 N2869 - FCD 2 ISO/IEC 14652 - Specification Method for Cultural Conventions, is "Yes" with the following comments.

1. Three new keywords for LC_CTYPE should be introduced in 4.2.

1.a: keyword "charclass" defines the extra set of keywords used in the LC_CTYPE category, examples "gaiji" to specify some custom Japanese characters, "alphabet" to specify what is the native alphabet of the language in question.
Syntax:
    charclass "gaiji";"alphabet";"class-n"
This is industry practice in for example GNU C.

1.b: keyword "width" should be added to specify the width of characters.
Syntax:
    width   (<charrange>;integer-width);...
This to support functionality in ISO C.

1.c keyword "alnum" should be introduced to specify what is alphabetic and numeric characters
Syntax: alnum <char1>;<char2>....

2. LC_DATE
In 4.6 for LC_DATE new keywords should be introduced era_d_t_fmt – analogeous to d_t_fmt for era era_t_fmt - analogeous to t_fmt for era This is to have a full set of formatting for era as for normal specifcation.

3. Alignment
In 4.6.2 alignment with the new C standard 9899:1999 should be sought with respect to %O and %E formats in LC_DATE. FDIS 9899 is expected to be available primo May 1999.

4. Coll weights
In 4.3.1 coll weights need not be in ascending order, as replace-after should be usable to rearrange the weights without the need to rearrange the order the lines of the specification is given in. Line 1869 and 2 more lines should be replaced with: "The weights for each of the collation elements determines the character collation sequence - such that each collation statement does not need to be in collation order, and weights could be rearranged via for example the "replace-after" keyword. No character has any specific predetermined placement in the collation sequence."

Comments from Germany

```
From: WACHTENDORF <Wachtendorf@NI.DIN.DE>

Subject: German vote on 2nd FCD 14652 - Comments


The German member body disapproves of ISO/IEC FCD_14652.2

Introduction
```

General

```
Germany has opposed to this draft standard from the very beginning. It
considers WG20 to be the place where general information on
internationalization is to be made available to other working groups
of SC22 and beyond. It would prefer to see the potentially
valuable information inherent in ISO/IEC_FCD_14652.2 to be made
available in narrative form in a technical report, rather than mixing
the discussion about the <em>contents</em> of internationalization
with that of its POSIX specific presentation form.

Furthermore, it is open to debate if some of the categories which
are present in FCD_14652.2 should not better be dealt with on an
application level. Examples for this are entries such as
<tt>LC_PAPER</tt>. For other entries such as <tt>LC_NAME</tt> the
formalization of its presentation does rather a disservice to the
user.
```

Comments from Japan

From: Tomomi HARUHANA <haruhana@itscj.ipsj.or.jp>

Comments on FCD 14652.2

The National Body of Japan disapproves FCD 14652.2 for the reasons below.

    NOTATIONS

        1) The expression "#xxxx" stands for a line number used in
        the printed and distributed version of SC22/WG20 N634, though
        the line numbers should be removed in the final text.

        2) The following abbreviations are used:

                POSIX.1 -- ISO/IEC 9945-1:1990
                POSIX.2 -- ISO/IEC 9945-2:1993

J-01) Introduction, #61-66:

>From the sentence

       This International Standard defines a general mechanism to
     ...
     formatting, telephone number handling, measurement handling,
     and a way to specify how much is covered and the status of it.

"measurement handling" should be removed because LC_MEASURE has been
abandoned.

J-02) Introduction, #81-95, Internationalization:

The item

       Internationalization   An internationalized application needs
                           to be designed and implemented as
                           cultural neutral, so that, at run time,
                           it draws on the cultural conventions of
                           the user thus giving the application
                           the ability to support cultural
                           conventions of many different cultures.
                           This standard specifies those cultural
                           conventions ...

should be changed to

      Productivity

> This standard specifies those cultural
> conventions and how to specify data for
> them. With those data an application developer
> is relieved from getting the different
> information to support all the cultural
> environments for the expected customers
> of the product. The application
> developer is thus ensured of culturally
> correct behavior as specified by the
> customer, and possibly more markets may
> be reached as customers may have the
> possibility to provide the data
> themselves for markets that were not
> targeted.

because

- the first sentence of the old item is ambiguous and overlaps
  with the previous item,

- "Internationalization" is not an appropriate subtitle here,

J-03) Introduction, #97-108, Uniform behaviour:

The item

> Uniform behaviour   When an application has been
>   internationalized, it is dependent on
>   the operating system support for
>   internationalization what level of
>   service is available to the user. ...

discusses too much on implementation variants and the benefit is not
clear. It should be changed to

> Uniform behaviour
>
>   When a number of applications share one cultural specification,
> which may be supplied from the user or a built-in nature,
> their behaviour for cultural adaptation become uniform.

considering the true intent of the Canadian comments on FCD.1 that cultural
specification needs not always be given by users.

J-04) Introduction, #109-112:

The sentence

> It is expected that the primary areas of use is within
> the POSIX operating system, ...

should be removed because there is no extension programme in POSIX
for this matter.

J-05) Introduction, #109-112:

In the sentence

        A number of cultural conventions, such as spelling,
        hyphenation rules and terminology, and classification of
        characters such as Japanese gaiji characters, are not
        specifiable with this standard, ...

the text "classification of characters such as Japanese gaiji
characters" should be removed because an user or a system can
specify for what classes the extended characters belongs.

        NOTE: "gaiji" is not an English word
        and it should not be included in a standard document
        without sufficient explanation.

J-06) Introduction, #121-122:

The sentence

        This International Standard defines a format compatible with
        the one used in the International String Ordering standard,
        ISO/IEC 14651. This International Standard is backwards

should be removed because it now becomes incompatible (see later comments).

J-07) Introduction, #131:
  2 Normative References, #174:
  4.2.1 Basic keywords, #887:

The word "10646" should be changed to "10646-1".

J-08) 1 Scope, #143-144:

The sentence

        The descriptions is intended to also be of use in other systems
        than POSIX

should be removed because it suggests the description is of use in POSIX.

J-09) 2 Normative References, #180:

This standard, ISO/IEC 15897:1998, contains no provisions which constitute
provisions of ISO/IEC 14652. ISO/IEC 15897:1998 gives only some helpful
hints in Clause 4.0 #483-484 and is used in a rationale in Clause 6, #3730.
It should be put into BIBLIOGRAPHY.

>   NOTE) This standard may be revived if one of Japan's comment is
>   accepted later.

J-10) 3.1.1 byte, #189:

The text "application defined" should be changed to "implementation defined"
because applications may specify the minimal number of bits but it does not
define the number.

J-11) 3.1.15 affirmative response, #246-248:
   3.1.16 negative response, #250-252:

The definitions are tautology. They should be removed.

J-12) 3.2.1   Notation for defining syntax, #269:

The text "the POSIX-2 standard" should be changed to "ISO/IEC 9945-2:1993"
because the abbreviation is not declared in this standard. The same kind of
change should be done in Annex B.1 FDCC-set Rationale, #6215.

J-13) 3.2.2   Continuation of lines, #292-296:

The contents of this subclause 3.2.2 should be moved to Clause 4 because the
line continuation is used not in this specification but in FDCC-sets defined
in Clause 4.

J-14) 3.2.2   Continuation of lines, #294:

(This comment should be neglected if the previous comment is accepted)

The expression "a specification" is ambiguous. It should be clarified.

J-15) 3.2.3 Portable character set, #300-302:

In this subclause, there is no explanation for what "the portable character set" is and how and where it is used in this specification.

The text should be changed to

> A set of symbolic names for characters in Table 1, which is called the portable character set, is used in character description text of this specification.

J-16) 3.2.3 Portable character set, Table 1, #309-316:

The symbolic names from <NUL> to <form-feed> are not defined in ISO/IEC 10646-1.Change the table as follows

Symbolic name   Glyph   Description

```
        <NUL>                              NULL (NUL)
        <alert>                            BELL (BEL)
        <backspace>                        BACKSPACE (BS)
        <tab>                              CHARACTER TABULATION (HT)
        <carriage-return>                  CARRIAGE RETURN (CR)
        <newline>                          LINE FEED (LF)
        <vertical-tab>                     LINE TABULATION (VT)
        <form-feed>                        FORM FEED (FF)
        <space>                            <U0020>   SPACE
        <exclamation-mark>    !            <U0021>   EXCLAMATION MARK
                ...
```

and add some explanation e.g.

> The first eight entries in Table 1 are defined in ISO/IEC 6429 and others are defined in ISO/IEC 10646-1.

J-17) 3.2.3 Portable character set, #421-#430:

The text

> This standard places only the following requirements on the encoded values of the characters in the portable character set:
>
> (1) ....
>
> (2) ...

should be removed because there is no need for restricting the encoding. The notion of FDCC-set should be applicable to the systems using the character set not satisfying these requirements -- e.g. EBCDIC code set.

J-18) 4 FDCC-set, #464-465:

The statement here

      This standard also defines an FDCC-set named "i18n" with
      values for each of the above categories.

should be changed to

      This standard also defines an FDCC-set named "i18n" with
      values for some of the above categories in order to simplify
      FDCC-set descriptions for a number of cultures. The contents
      of "i18n" categories should not be considered as the most commonly
      accepted values or as the recommendation.

because the aim of the FDCC-set is not to develop a global standard and
some categories will not be in agreement even with this explanation.

J-19) 4.0 FDCC-set definition, #435:

The subclause numbering should start from '1'.

J-20) 4.0 FDCC-set definition, #493-496:

The text

      The category body shall consist of one or more lines of text.
      Each line shall contain an identifier, optionally followed by
      one or more operands. Identifiers shall be either keywords,
      identifying a particular FDCC, or collating elements, or
      section symbols, or transliteration statements.

should be changed to

      The category body shall consist of one or more lines of text.
      Each line shall be one of the following:

         - a line containing an identifier, optionally followed by
           one or more operands. Identifiers shall be either keywords,
           identifying a particular FDCC, or collating elements, or
           section symbols,
         - one of transliteration statements defined in 4.2.

because transliteration statements are not identifiers.

      NOTE) This text should be changed again if one of Japan's
      comment is accepted later.

11

J-21) 4.0.1   Character representation, #516-518:

The requirement

> The left angle bracket (<) is a reserved symbol, denoting
> the start of a symbolic name; when used to represent itself
> it shall be preceded by the escape character

is different from that in Clause 6

> If a right angle bracket or an escape character is used
> within a symbolic name, it shall be preceded by the escape
> character

which allows names like

```
<<>     <U003C>     LESS-THAN SIGN
<<(>    <U005B>     LEFT SQUARE BRACKET
```

and so on. There is no need to have different syntax in FDCC-set and
repertoiremap. They should be aligned.

J-22) 4.0.2.1   comment_char, #606-608:

The sentence

> Blank lines and lines containing the <comment char> in the first
> position, and the remainder of a line with a <comment char>
> occurring where an end of line may occur, shall be ignored

should be changed to

> Blank lines and lines containing the <comment char> in the first
> position shall be ignored

Rationale:

Comments not beginning from the top of the line interferes with the syntax
notations such as

```
"%s %s;%s;...;%s\n",<collating-identifier>,<weight>,<weight>,...
```
or
```
"copy %s\n", <FDCC-set-name>
```

which specify the exact sequence of characters. Someone may say such a
syntax notation applies to the result of comment removal. But it will not
work because "where an end of line may occur" depends on syntax notations.

Generally speaking, a comment introducer will be allowed where it is easily
detectable and not confused with its literal usage, e.g. by its physical
position in the case of POSIX.2. In the case of the language C, the
characters "/*" introduce a comment except within a character constant, a
string literal or a comment all of which can be easily detected by its
carefully designed syntax.

Comments not beginning from the top of the line might be allowed if all the character constants and character strings in FDCC-sets were enclosed in some separator pairs. But it is not the case here.

This problem was pointed out in J-13 comment on FCD.1 and the disposition rationale

        Rejected. This is requested by experts of other NBs during the
        development of the standard.

        The standard says that comment lines can not be continued with
        the escape character at the end of the line.

did not give an answer to the contradiction but said about the unrealistic desire in the first sentence and irrelevant matter in the second sentence.

NOTE: the comments used in

                upper /
                % TABLE 1 BASIC LATIN
                  <U0041>..<U005A>;/
                % TABLE 2 LATIN-1 SUPPLEMENT
                        ...

is not a case of "comment lines can not be continued ..."
But it may be better to clarify this matter by changing the sentence for "line continuation", now in 3.2.2 #294-296 and Japan requests to move it in Clause 4, to

        A line in a specification can be continued by placing an
        escape character as the last visible graphic character on the
        line; this continuation character shall be discarded from the
        input. The line is continued to the next non comment line.

J-23) 4.0.2.2   escape_char, #610-618:

Add at the end of this subclause a sentence --

        The escape character is used for representing characters in 4.0.1
        and for continuing lines.

J-24) 4.0.2.3   repertoiremap, #622-626:

Add a explanation for name of repertoiremaps allowed in this statement:

        The name shall be one of

                - "i18n" which indicate the "i18n" repertoiremap
                defined in this standard,

                - the name of charmap/repertoiremap registered

by the process defined in ISO/IEC 15897,

- any other name which may be recognized in some
local context -- not being recommended as an international
specification.

The same type of action should be done in "4.0.2.4 charmap" and in all
the "copy" keywords in FDCC-set categories.

J-25) 4.0.2.4 charmap, #635-641:

The text here is confusing. It should be changed to

This keyword gives a hint on which charmaps a FDCC-set is
meant to be supported by.
There may be more than one charmap specification in a FDCC-set.
It is an application's responsibility to decide what
mapping between symbolic character names and character codes
is to be used with that application.
The mapping for an application may be a mapping defined in one of
charmaps which is referred in charmap statements or it may be a
mapping not referred in charmap statements.

J-26) 4.1 LC_IDENTIFICATION, #659-660, #678-679:

The keyword

language   Natural language to which the FDCC-set
                         applies, as specified in ISO 639.

and a note

Note: Only one culture can be addressed with the concepts of
a FDCC-set; to address for example a bilingual culture, one
need to have 2 FDCC-sets

put a unnecessary restriction on the notion of "culture". There are a
number of cultures which allow the use of plural languages simultaneously.

J-27) 4.1 LC_IDENTIFICATION, "language", #659-660:

The explanation of this keyword should be changed to

This keyword specifies natural languages used in that culture.
Each operand may be an ISO 639 identifier or a character string
starting with ':' describing an unstandardized language.

in order to correspond to the wider requests.

J-28) 4.1 LC_IDENTIFICATION, "territory", #661-662:

The explanation of this keyword should be changed to

> territory   The geographic extent where the FDCC-set
>   applies (need not be a national extent),
>   the operand may be a two-letter string form
>   of ISO 3166 or a string starting with ':'
>   describing a non-national area.

in order to correspond to the wider requests.

J-29) 4.1 LC_IDENTIFICATION, #653:

The keyword "contact" should be optional.

J-30) 4.1 LC_IDENTIFICATION, #695-672:

The default value is not needed for this category because the contents here
should not be copied in other FDCC-sets.

If it remains, it should be as follows:

```
LC_IDENTIFICATION
% This is the ISO/IEC 14652 "i18n" definition for
% the LC_IDENTIFICATION category.
%
title                "ISO/IEC 14652 i18n FDCC-set"
source               "ISO/IEC Copyright Office"
address              "Case postale 56, CH-1211 Geneve 20, Switzerland"
contact              ""
email                ""
tel                  ""
fax                  ""
language             ""
territory            ":the area covered by the national bodies of
                      ISO/IEC"
revision             "1.0"
date                 "1999-12-20"
```

J-31) 4.2.1 Basic keywords, #780:

The sentence

> The following keywords shall be defined

should be changed to

        The following keywords shall be recognized

which is used in POSIX.2

J-32) 4.2.1 Basic keywords, #797:

The expression "word-like identifiers for natural languages" sounds queer.
The definition should be changed to

        alpha   Define characters to be classified as used to
                    spell out the words for natural languages;
                    such as letters, syllabic or ideographic

J-33) 4.2.1 Basic keywords, #809-813:

In the definitions of "digit" and "outdigit"

        digit   Define the characters to be classified as numeric
                       ...
                    values. The "digit" keyword is used to specify which
                    characters are accepted as digits in input, and

        outdigit   Define the characters to be classified as numeric

what do the words "input" and "output" mean -- "input" means typing in and
"output" means printing or displaying?

J-34) 4.2.1 Basic keywords, "class", #879-881:

        class   Define characters to be classified in the class with
                    the name given in the first operand, which is a
                    string. This string shall only contain characters of
                    the portable character set that either has the

The use of "either" should be checked by native English writers.

J-35) 4.2.1 Basic keywords, class, #886:

The sentence

        The following two names should be recognized

should be inserted before the explanation of "combining" and
"combining_level3".

16

J-36) 4.2.1 Basic keywords, map, #909:

The example contains errors. It should be changed to

```
     "kana",(<U30AB>,<U304B>);(<U30AC>,<U304C>);(<U30AD>,<U304D>)
```

J-37) 4.2.2 Character string transliteration:

This subclause should be removed because the technical contents defined here
are too premature for international use.

Transliteration depends on the source and destination languages. So the
transliterated values for characters vary depending on the language context
and the current specification neglects this.

If the transliteration is to be contained in this standard, the following
method, which is similar to mapping, seems better:

> The syntax is given as
>
> > ```
> > "translit %s to %s by %s",<source_lang>,<dst_lang>,<rules>
> > ```
>
> and its example is
>
> > ```
> > translit "Russian" to "English" by (<Uxxxx>,<Uyyyy>);\
> > (<Upppp>,<Uqqqq><Urrrr>);....
> > ```
>
> where applications may use language labels to select the appropriate
> rue set.

J-38) 4.2.2 Character string transliteration:

```
     (this comment should be neglected if the comment J-xx is accepted)
```

Converting all the characters not included in a source character subset to
the "default_missing" characters is not a general solution. A new syntax is
needed to specify which characters are not converted and which characters
are converted to "default_missing".

J-39) 4.2.3 "i18n" LC_CTYPE category:

This subclause should be removed because it is too early to define the
default of character classification for all characters in UCS.

The disposition to the same comment from Japan on FCD.1 says

> Rejected. This is a stable definition.

But consider the fact that FCD.1 tried to classify some of CJK characters as
"digit" and only Japan protested and got acceptance. There is no response
>from China and Korea -- of course they share the same concern as many
Western experts agreed to Japan's protest. This makes clear the
unstableness of classifications at this point of time and in the current
commenting system.

J-40) 4.2.3   "i18n" LC_CTYPE category, #1094

"U3EE" should be changed to "U03EF".

J-41) 4.2.3   "i18n" LC_CTYPE category, #1125:

"U0148" should be changed to "U0147".

J-42) 4.2.3   "i18n" LC_CTYPE category, "digit", #1274-1278:

These lines should be changed to

```
        digit /
        % TABLE 1 BASIC LATIN
        <U0030>..<U0039>;/
        % TABLE 15 and 16 ARABIC
        <U0660>..<U0669>;<U06F0>..<U06F9>;/
        % TABLE 17 DEVANAGARI
        <U0966>..<U096F>;/
        % TABLE 18 BENGALI
        <U09E6>..<U09EF>;/
        % TABLE 19 GURMUKHI
        <U0A66>..<U0A6F>;/
        % TABLE 20 GUJARATI
        <U0AE6>..<U0AEF>;/
        % TABLE 21 ORIYA
        <U0B66>..<U0B6F>;/
```

in order to make the table easier to be checked.

J-43) 4.2.3   "i18n" LC_CTYPE category, "space", #1282-1283:

These lines should be changed to

```
        space/
        % ISO 6429
        <U0008>;<U000A>..<U000D>;/
        % TABLE 1 BASIC LATIN
```

```
        <U0020>;/
        % TABLE 35 GENERAL PUNCTUATION
        <U2000>..<U2006>;<U2008>..<U200B>;/
        % TABLE 50 CJK SYMBOLS AND PUNCTUATION, HIRAGANA
        <U3000>
```

in order to make the table easier to be checked.

J-44) 4.2.3 "i18n" LC_CTYPE category, "punct", #1287-1306:

These lines should be rearranged with comments on which UCS Table they
belong in order to make the table easier to be checked.

J-45) 4.2.3 "i18n" LC_CTYPE category, "graph", #1308-1376:

The characters belonging to "upper" and "lower", which are defined to be
automatically included in this class, should be removed from here in order
to make the table simpler as is done in POSIX.2 locale and as is shown in
Annex.2 of Japan's comments on FCD.1.

J-46) 4.2.3 "i18n" LC_CTYPE category, "toupper", "tolower", #1384-1712:

This part of the definition is too difficult to be checked by human readers.

It should be modified by

    1)   introducing a notation
         such as
                 (<U0102>..<U010E>, <U0102>..<U010E>)
         and
                 (<U0102>..(2)..<U010E>, <U0102>..(2)..<U010E>)
         to simplify the sequences with incremental two,

    2)   comment lines should be added for readability

If accepted, Japan will prepare the text.

J-47) 4.3 LC_COLLATE:

The whole contents of this subclause should be put back to that of POSIX in
order to keep upward compatibility and a new subclause LC_COLLATE_14651,
which enables to contain a "delta" specification being defined in ISO/IEC
14651 as a cultural convention.

Rationale:

1) POSIX upward compatibility is lost -- e.g. order-start statement in POSIX

becomes illegal in FCD.2.

2) Incompatibility with 14651 -- 14651 -- tailoring is done only by "delta" declaration,

3) Many new functionality not included in POSIX and 14651 -- e.g. toggling keywords -- which will be an obstacle to the 14651.

J-48) 4.4 LC_MONETARY:

The way of specifying the valid time range of currencies and conversion rates is difficult to use. They should be changed as follows:

  1) the time ranges should be specified uniquely for any case with sufficient precision (as is seen in the examples below).

  2) the valid time range should be specified by the optional parameters of "currency_symbol" and "int_curr_symbol" e.g.

```
    currency_symbol "Foo" from "1976-01-01T12:00Z"

    currency_symbol "Bar" from "2001-01-01T00:00+09:00" to \
             "2001-12-31T24:00+09:00"
```

which mean the currency "Foo" began to valid from the noon of the first day of 1976 in the UTC and the currency "Bar" is valid from the fist minutes of 2001 to the last minutes of 2001 in the local time which is nine hours ahead of Coordinated Universal Time.

  3) the target currencies of conversion_rate should be specified explicitly as follows:

```
    conversion_rate (120 in "Foo") = (100 in "Bar")
```


J-49) 4.4 LC_MONETARY, "valid_from" and "valid_to", #2638-2650:

        (this comment should be neglected if the comment J-xx is accepted)

The representation like "19980630" should be considered not as an integer but as a character string because the semantic of an integer is not dependent on a specific representation -- octal, decimal or hexadecimal.

Is the validity of currency always beginning from or ending at midnight? And are there some ambiguity for the time zone? If there are some future possibility, it is safe to declare them in a form like "1999-08-16T12:00Z" using UTC form of ISO 8601. Anyway ISO 8601 should be referred here.

J-50) 4.4 LC_MONETARY, "conversion_rate", #2651-2659:

(this comment should be neglected if the comment J-xx is accepted)

The text is ambiguous about what is the currency in the question
and what is the first valid currency (local or international).

J-51) 4.4 LC_MONETARY, #2830-2852:

The "i18n" FDCC-set should not be defined for this category because it is
dangerous to set the decimal point as null.

If this removal is not accepted, then some warning about the usage of this
default category should be given.

J-52) 4.5 LC_NUMERIC, #2888-2898:

The "i18n" FDCC-set should not be defined for these categories because it
violates the definition

        This keyword cannot be omitted and cannot be set to the empty string

J-53) 4.6 LC_TIME, #2901-:

The way of introducing non-Gregorian calendar systems in this draft should
not be approved because

1) it changes the meaning of POSIX locales unstable because it becomes
impossible to judge the semantics of the time system because there may be a
time system which has the same number of months and week days.

2) it disables the usage of the non-Gregorian calendar concurrently with
Gregorian calendar. In Japan, the Gregorian representation of the year and
the representation based on Era system are frequently used even in one
documents and it is enabled in the POSIX system by assigning a different
descriptor for years. But the current specification inhibits such a double
representation of date using non-Gregorian and Gregorian calendars.

Japan will continue to disapprove as long as the specifications developed in
POSIX are changed syntactically or semantically.

Japan recommends, if non-Gregorian calendar is to be supported, this
standard should prepare a new set of keywords and the escape sequences.

J-54) 4.6 LC_TIME, week, #2925-2935:

        (this comment should be neglected if the comment J-xx is accepted)

The text is not understandable as English -- for example, there is no word

corresponding to the clause "which is the first weekday".

J-55) 4.6   LC_TIME, week, #2925-2935:

(this comment should be neglected if the comment J-xx is accepted)

This keyword should be optional in order to accept POSIX locale as a FDCC-set.

J-56) 4.6   LC_TIME, before "era" #2869:

The sentence

The following keywords are all optional

should be inserted between "t_fmt_ampm" and "era" in order to accept POSIX locale as a FDCC-set.

J-57) 4.6   LC_TIME, timezone, #3090:

At the end of this definition, the following note should be added:

NOTE: This way of specifying the timezone is compatible with the format for the environment variable TZ described in Section 8.1.1 of POSIX.1.

J-58) 4.6.1   Date Field Descriptors, #3097

Add the following sentences at the end of main text of this subclause:

This category does not define which timezone -- local time or UTC -- is used in the interpretation of file descriptors. It's the responsibility of each applications to select the appropriate time zone or to support an option for user's selection.

J-59) 4.6.1   Date Field Descriptors, %U, #3125-3126:

The sentence

All days in a new year preceding the first Sunday shall be considered to be in week 0

which exists in POSIX should be inserted at the end.

J-60) 4.9 LC_NAME, #3302-3303:

The explanation for "%d"

    %d   Salutation, using the FDCC-sets conventions, with 1
                    for the name_gen, 2 for name_mr, 3 for name_mrs, 4
                    for name_miss, 5 for name_ms.

is not understandable. Where does the integer between 1 and 5 come from?

J-61) 4.10 LC_ADDRESS, #3309-:

This category should be removed because it is too premature to be
standardized as follows:

  1) no room for representing "state" and "prefecture",

  2) too much dependent on European culture -- use of CEPT-MAILCODE etc.,

J-62) 4.10 LC_ADDRESS, country_post, #3336-3337:

        (this comment should be neglected if the comment J-xx is accepted)

The use of CEPT-MAILCODE should not be admitted in an international standard.

J-63) 4.10 LC_ADDRESS, country_isbn, #3347-3348:

        (this comment should be neglected if the comment J-xx is accepted)

A note to clarify why ISBN code is introduced here.

J-64) 5. CHARMAP, #3345:

The declarations <escseq>, <addset> and <include> should be removed.

RATIONALE:

1) The FDCC-set is a human readable document and needs no consideration
for encoding,

2) The charmap, which maps symbolic names to specific code values,
should be regarded as a old tools for keeping upward compatibility for
POSIX locales and should not be augmented.

The linkage of symbolic character names to a code system based on ISO 2022 environment is a local and/or implementation matter outside of the cultural convention.

This comment is the same as in FCD.1.

The disposition to FCD.1 comment said

> Rejected. The encoding of characters are a cultural element. For example in Denmark it is the cultural convention to employ a specific set of characters, and the encoding, possibly using 2022 techniques is also a specific cultural convention.
>
> The charmaps are necessary for making the FDCC-sets function in an IT environment.

But Japan protests to this because the encoding is not considered as a cultural convention which is defined as

> 3.1.5 cultural convention: A data item for information technology that may vary dependent on language, territory, or other cultural habits.

J-65) Clause 6. Repertoiremap, #3698-:

Do not use specific mnemonics to specify "i18n" repertoiremap. Whatever wording is used, this description may give an user of this standard

an impression of "this mnemonics is normative".
The mnemonics project proposal was rejected at SC22 WG20 long time ago, so, to sneak in the rejected proposal into JTC1 standard should not be done.

As was pointed out in the previous US comments. this list is arbitrarily chosen, and the principles for characters in it are unstated. If the repertoire file is not going to correspond to one of the named and numbered subsets of ISO/IEC 10646 (and Subset 300, the BMP, would be the obvious choice), then the choice of characters in the repertoire file *must* be justified in 14652.

If the intention is, rather, to just define a bunch of short mnemonics, then most of this entire listing is useless and should be omitted. Introducing mnemonics such as <c*> for GREEK SMALL LETTER XI and <z%> for CYRILLIC SMALL LETTER ZHE and <K%> for HEBREW LETTER FINAL KAF is completely confusing. A very small percentage of these mnemonics has seen widespread use in plaintext reference to accented characters. The rest should be completely abandoned in CD 14652 in favor of use of the hexadecimal value as the unique symbolic identifier for a 10646 characters (e.g. <U0436>).

This comment is the same as in FCD.1.

The disposition to FCD.1 comment said

Rejected. The list of mnemonics builds on existing practice, including POSIX and Internet use.

But Japan considers

-- existing practice is not a rationale for adopting as international standard,

-- POSIX.2 itself does define only a limited number of symbolic names as in its portable character set;   some locale may define more symbolic name as its own cultural convention and it should not be considered as an international default,

-- there are many kinds of Internet use and not unique.

J-66) 6 REPERTOIREMAP, #3716:

The symbolic names <U80000000>..<U8FFFFFFF> for characters not in ISO/IEC 10646 should be changed to <P00000000>..<PF8FFFFFF> as is done in FCD 14651.2

J-67) 6 REPERTOIREMAP, "i18nrep", 3821-3846:

(this comment should be neglected if the comment J-xx is accepted)

The lines

```
<a8>    <U0252>      Weight indicating the position of the last a
...
<z8>    <U0293>      Weight indicating the position of the last z
```

should be removed.

J-68) 6 REPERTOIREMAP, "i18nrep":

(this comment should be neglected if the comment J-xx is accepted)

The following duplication

```
<OC>    <U00A9>      COPYRIGHT SIGN
<OC>    <U009D>      OPERATING SYSTEM COMMAND (OSC)

<OR>    <U2228>      LOGICAL OR
<OR>    <U00AE>      REGISTERED SIGN
```

should be resolved.

J-69) 6 REPERTOIREMAP, "i18nrep", #6026-6071:

        (this comment should be neglected if the comment J-xx is accepted)

The private characters

        <"3>   <U80000000> DIACRITICAL MARK UMLAUT <ISO-IR-53_C9/> (not a
real   ...
        <//c>   <U80000024> JOIN THIS LINE WITH NEXT LINE (Mnemonic)

should not be included.

J-70) Annex C BNF Grammar, #6935-6936, 6941:

The use of "(*" and "*)" for special sequences (ISO/IEC 14977 term) and for
comments should be changed. For special sequences, the character '?'
defined in ISO/IEC 14977 should be used.

J-71) Annex C BNF Grammar, #6950:

The syntactic exception, which is an ISO/IEC 14977 term and is represented
by the symbol '-', should not be used because the concept is not common and
it is used without any explanation.

The rule should be changed to

        graphic_char = ? any character except control_characters and space ?

using the special sequence discussed above.

J-72) Annex C BNF Grammar, Global:

All the identifiers should be written in lowercases because it is common to
use lowercases letters for identifiers for non-terminals as is described in
2.1.2 of POSIX.2. The definitions such as

        elem   = char_symbol | COLLSYMBOL | COLLELEMENT
;
        COLLSYMBOL   = simple_symbol ;

are confusing to many readers.

        NOTE: COLLSYMOL is a terminal (token) in POSIX but it is a non-
        terminal in this standard.

J-73) Annex C BNF Grammar, Global:

The rule

```
        CHAR     = (* any character *);
```

should be changed to

```
        CHAR     = ? any character except those that makes an End Of Line ?
```


J-74) Annex C BNF Grammar:

The rules

```
        EOL                  = (* anything that makes an End Of Line (EOL)
                               in the operating system employed *)
                               | comment EOL ;
        comment              = COMMENT_CHAR CHAR* ;
```

will cause troubles as is already pointed out in the previous Japan's comment.

J-75) Annex C BNF Grammar:

The two rules

```
        portable_graph            = letter ...
        portable_char             = portable_graph | ...
```

should be removed because they are not used in other rules.

J-76) Annex C BNF Grammar:

" CHAR " in

```
        char_symbol               = CHAR | CHARSYMBOL
                                    | OCTAL_CHAR | HEX_CHAR | DECIMAL_CHAR ;
```
should be changed to " graphic_char ".

J-77) Annex C BNF Grammar:

The rule

```
        FDCC_set_definition = [ global_statement* ] category* ;
```

should be changed to

```
        FDCC_set_definition = [ global_statement* ] category category* ;
```

as is defined #438-439.

J-78) Annex C BNF Grammar:

```
#7028 "clarclass_keyword" -> "charclass_keyword".

#7037 "abs_ellipsis" -> "ctype_abs_ellipsis"

#7186 "qouted_string" -> "quoted_string"
```

Comments from Sweden

(Specification method for cultural conventions)

Sweden votes NO on this FCD with the following comments.
(Where the heading says "major" all points, except where otherwise noted
initially, are "major". Note: We see no need to comment on the details of
the FCD2 text, since we very strongly favour a complete rework from
scratch of this CD. Very little text from FCD2 would be present in such a
completely reworked text.)

SE-1. Relation to 14651 (major)

SE-1.1. The current text in 14652 contains text on how to interpret
collation
tables. The interpretation given in 14652 is different from, and
inconsistent with, that given in (present, CD, and future) 14651. In order
to avoid any inconsistency in interpretation of collation tables when
trying to conform to both 14652 and 14651, it is best to remove all text
implying any kind of interpretation of a collation table, leaving only a
(normative) reference to 14651.

SE-1.2. 14651 (internally) and 14652 might not be using the same table
format
for collation tables. In such case only a table transformation mapping
should be described, still leaving all interpretation description of a
collation table to 14651.

SE-2. Mix of definitions and preference selections (major)

SE-2.1. CD 14652 requires that definitions are intermixed, and confused with,
preference selections. Definitions (of paper sizes, date formats, monetary
formats, etc.) should be clearly separated from preference selections,
where one is choosing among defined (and named) paper sizes (maybe
different ones are used for different purposes, and one should be able to
override the default preference by referring to another definition), date
formats, monetary formats, etc.

SE-2.2. It should be possible to have a hierarchy of preference selections.
E.g. there may be one or more system level preference selections, working
group preference selections that may refer to one of the system preference
selections, and individual preference selections that may refer to another
preference selection for selections not made explicitly by the user.

SE-2.3. CD 14652 requires that one amalgamate definitions for unrelated
categories. E.g. one is required to specify monetary format together with
a collation table, etc. Definitions for unrelated categories must not be
required, maybe not even allowed, to be amalgamated.

SE-2.4. The definitions are not named beyond category name in an "FDCC set",
which makes it impossible to put related definitions of the same category

together. It also makes it impossible for a user to select definitions from several locales, without having to build a new "FDCC set", which would be overwhelmingly taxing for the user. E.g. it must be possible to select Italian monetary unit/format, while using Swedish collation rules, just by selecting such a combination, not defining a new "FDCC set"; etc.

SE-2.5. It must further be possible to put related definitions together. E.g. the definitions of the paper sizes (A4, A3, B4, _, US letter, _) must be possible to put together, rather than having to spread them on multiple "FDCC sets". Likewise, it must be possible to put the collation tailoring definitions together; etc. The user can then make the desired selection by name.

SE-3. Character issues (major)

SE-3.1. The character encoding for any text file describing the definitions or selections must be clear in the file itself, unless one fixes the character encoding on UTF-8 or UTF-16. Compare XML where the character encoding is self-declared in the file. "Platform dependence" is not acceptable.

SE-3.2. 14652 has a large "repertoiremap". This must be removed entirely, as the names defined serves no useful purpose, and are indeed strange and controversial. It is better to use the actual characters, or if need be, reference them by number (compare 'numeric character references' in XML/HTML).

SE-3.3. 14652 allows any "FDCC set" to have it's own list of character properties. Most character properties are fixed (like if the character is a lowercase letter, or a digit, or a _), and are not subject to 'cultural adaptability', though they are subject to versioning (to correct errors, or add character properties). This means that most character properties must not be declarable in an arbitrary FDCC set (only at 'top level' in some way).

SE-3.4. Character encoding mapping tables are missing. These are also not subject to cultural adaptability, but are subject to versioning.

SE-4. Other issues (major)

SE-4.1. 14652 often uses C-printf-like format codes, i.e. % followed by (a) letter(s). Such methods are C-specific, and must not taint any definitions relating to the cultural specifications for man-computer UI.

SE-4.2. 14652 uses its own full syntax for the "FDCC sets". The current, very strong, trend for data files, like the FDCC-sets, is to modularise in the following way: use XML (or SGML) for the general file format, and specify only domain specific syntactic restrictions. Since SGML is an ISO standard, there should be no problem in referencing it normatively.

SE-4.3. UTC leap second correction specifications are missing.

SE-4.4. Geographic limits for time zones are missing (think about mobile computers with a GPS unit).

SE-4.5. Measurements units and unit conversion factors are missing (US vs. SI; typography vs. other things).

SE-5. Conclusion

In short, the entire CD 14652 need to be reworked from scratch, leaving the C/POSIX legacy behind, as that can never be made to cater for a well-designed system of (computer program) internationalisation specifications.

Comments from the UK
From: Robert Yarlett <Robert_Yarlett@BSI.ORG.UK>

Subject: FCD 14652

The UK Votes No to ISO/IEC FCD 14652 However we would support this document being produced as a Technical Report

UK votes a "conditional"
     NO on the FCD.

Unless

I ISO/IEC FCD 14652 is changed to an ISO Technical Report, the UK vote should be changed to a YES one.

Comments from the USA
Susan Bose
for the US P-member JTC 1/SC22

The US National Body votes to Disapprove the Second FCD Ballot for ISO/IEC FCD 14652 - Information technology - Programming languages, their environments and systems software interfaces - Specification Method for Cultural Conventions [SC22 N2869].

USA-1 Many of the U.S. objections to the prior draft were not accommodated in the revised document.

USA-2 The U.S. still objects in principle to the entire approach towards specification of cultural elements represented by the FDCC-set's.

USA-3 The U.S. still objects to the detailed specification of character properties in 14652, since they do not belong there, but rather should be in the purview of SC2/WG2, in conjunction with 10646 itself.

_____end of USA comments _____

_____ end of SC22 N2917 _____