

ISO/IEC JTC1/SC2/WG2
Coded Character Set
Secretariat: Japan (JISC)

Doc. Type: Draft disposition of comments

Title: Draft disposition of comments on PDAM2 to ISO/IEC 10646 4th edition

Source: Michel Suignard (project editor)

Project: JTC1 02.10646.00.02.00.04

Status: For review by WG2

Date: 2014-09-24

Distribution: WG2

Reference: SC2 N4340 4354 WG2 N4588R N4592 N4597 N4610 N4613 N4614 N4627 N4632

Medium: Paper, PDF file

Comments were received from the following members: China, Ireland, Japan, UK, and USA. The following document is the disposition of those comments. The disposition is organized per country.

Note – With some minor exceptions, the full content of the ballot comments have been included in this document to facilitate the reading. The dispositions are inserted in between these comments and are marked in **Underlined Bold Serif text**, with explanatory text in italicized serif.


China: Positive

Technical comment

T1. About 5 “Birga”s in Mongolian code chart

China does not agree to encode 5 “BIRGA”s (U+181A through U+181E). There are up to 14 “BIRGA”s (all are variants of U+1800 “MONGOLIAN BIRGA”) found in literatures, which means there is not enough space for them in Mongolian code chart. A proposal on usage of VARIATION SELECTORS for Mongolian, including that for “BIRGA”s will be submitted to SC2/WG2 soon.

[Text of WG2 N4632 added here]

Inputting the Birga (U+1800) + ZWJ (U+200D) is not an ideal way of implementing the , so we should modify it. This needs to be discussed and developed.

According to our further investigation into the Birga of the Mongolian Script, there are at least 14 forms. They appear in the Todo Mongolian and traditional Mongolian literature. These 14 variants are base forms. There are more stylistic forms beyond the 14 variants. One thought of encoding is to place all of the 14 variants on line 181* as individual code-points. But it is not possible to encode them all there.

We think that all of the Birga variants are not the same character. To determine whether a Birga deserves an individual encoding we need to look at its function. According to our investigation, each of the 14 forms carries unique sets of functionality. Functions may include indicating the head of the article or page. In manuscript revision, the birga also indicates whether the article is the original or has actually been revised. It may also indicate an opening paragraph. The functions will vary between Todo and Mongolian with a greater variety of function on the side of the Todo usage.

Names like “Birga one”, “Birga two”, ... “Birga nine” could be an option for naming the variants. There are no intuitive methods to naming the variants. In the early of Mongolian Script usage, we find no descriptive discussion regarding nomenclature. They are not assigned names throughout history, so it is not appropriate and may not be accepted by the public to give individual names to each birga.

We are drafting a report suggesting the use of variation selectors (FE00-FE0F) in the Mongolian Encoding as below.

[Proposal showing the use the Variation Selectors (a mix of Mongolian Free Variations Selectors: FVS1 to FVS3 and regular Variations Selectors VS4 to VS13) associated with the base character 1800 MONGOLIAN BIRGA.]

Proposed change by China

Remove 5 Birgas from Amendment.

WG2 discussion

See also comment E.5 from U.K. concerning Birga names.

Several points worth mentioning:

1. Current implementations use non-conformant sequences with FVS1 to FVS3 and ZWJ to represent 4 of the 5 proposed birgas. China recognizes the non-ideal situation for ZWJ, is proposing to use VS4 instead, and then VS5 to VS13 for the additional birgas.
2. One argument for using VS instead of discrete characters is the lack of space in the Mongolian code chart. However there is still ample space in that block to encode these additional 8 birgas. An alternative is to create a new block ‘Mongolian Supplement’ to encode these 13 code points (5 in amendment 2 +8 proposed by China).
3. Prof. Quejingzhabu (author of the letter included in WG2 N4632) is saying: “We think that all of the Birga variants are not the same character. To determine whether a Birga deserves an individual encoding we need to look at its function. According to our investigation, each of the 14 forms carries unique sets of functionality.” At the same time ISO/IEC 10646 mentions in sub-clause 16.6.1 that “the purpose [of VS] is to indicate a specific variant form of graphic symbol for that base character”. Based on this, because the birgas are different characters performing separate functions (unlike typical variation sequences), the usage of VSs (and in this case a mix of VS from various blocks) seems inappropriate.
4. Based on the points above, the encoding of these 13 birgas in a new block seems a better solution.
5. Concerning the names, the preference of the experts for naming symbols has been to use descriptive names. A good example is the set of Siddham section marks part of Amendment 115CA..115D7. Experts should explore the two alternative, either digits 1 to 13 to express the set such as BIRGA SYMBOL ALT-1 to BIRGA SYMBOL ALT-13, or descriptive names like for Siddham section marks. See the U.K. comment E.5 about using consistently the adjective ‘ornamented’ in the name.

T2. About Tangut code chart

1) The radical of  should be 

Proposed acceptance


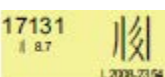
See also comment T3 from Ireland and T10 from UK.

Also requested in document WG2 N4588R (Tangut glyph corrections), page 10.

2) There is no Tangut glyph like , it should be taken place by 

WG2 discussion

This is not supported by document WG2 N4588R. Furthermore they have separate source references and it seems prudent to keep both forms.

3) The glyph of  is wrong, it should be taken place by 

WG2 discussion

This is not supported by document WG2 N4588R. Furthermore they have separate source references and it seems prudent to keep both forms.

4) The radical of  should be 

Proposed acceptance

Also requested in document WG2 N4588R (Tangut glyph corrections), page 9.

Ireland: Negative

Ireland **disapproves** the draft with the editorial comments given below. Acceptance of these comments and appropriate changes to the text will change our vote to approval

Technical comments

T1. Page 9, Row 0B8: Tamil

Ireland requests that the six characters at 0BDF, 0BFB, 0BFC, 0BFD, 0BFE, and 0BFF be moved to 11FF0..11FF5 in the Tami Supplement block, since they are little-used archaic characters. Ireland also notes that there have been a number of UTC documents suggesting name and annotation improvements for the Tamil characters, and recommends that WG2 make whatever changes might be agreed after further discussion.

Propose acceptance

T2. Page 9, Row 1030: Old Italic

The Irish NB continues to be of the opinion that the addition of the character 1032F OLD ITALIC LETTER TTE implies an implicit unification of Old Italic and North Italic scripts whose technical merit has not yet been agreed. We agree that a TTE should be encoded for writing North Italic, but do not believe that the ramifications of unification with Old Italic have been agreed by all of the stakeholders. This issue is similar to that of Phoenician and Hebrew. Ireland requests that the encoding of 1032F OLD ITALIC LETTER TTE be delayed for further study.

WG2 discussion

This has been going on for too long.

T3. Page 48, Row 1700: Tangut

With reference to ISO/IEC JTC1/SC2/WG2 N4588 “Tangut glyph corrections”, Ireland requests that the glyph changes and reordering of characters recommended in that document be implemented in the next version of the amendment.

WG2 discussion

See also comments T2 from China and T10 from UK.

Japan: Negative

Technical comments

T1. Page 37-39 Clause 32 – Zanabazar Square block

The new Zanabazar Square block may be incomplete. See WG2 N4613 for detail. Japan wants WG2 to consider the Zanabazar Square block based on the comments in N4613.

Proposed change by Japan

No concrete change proposal is given at this time.

WG2 discussion

See also comment T7, E8, and E9 from UK.

The document WG2 N4613 does not contain any substantial comments, just indicating the need for ‘further investigation’. The original proposal (WG2 N4541 quoted incorrectly as N4511 in one place in N4613) is a thorough document and a solid base for the encoding. What may be prudent is to keep the repertoire at the committee level and not push it to enquiry stage (DAM2) at this moment.

T2. Page 110-116 Clause 32 – Nushu block

The Nushu block may be still incomplete. See WG2 N4610 for detail. Japan wants WG2 to consider the Nushu based on the comments in N4610.

Proposed change by Japan

No concrete change proposal is given at this time.

WG2 discussion

The document WG2 N4610 is a thorough review of the Nushu proposal. One of its recommendation is: “He Yanxin glyph list” might be one of the most compact sources, so the cross section between PDAM2 and it would be useful to reduce the future works for the stabilization.”

Based on this, a good approach would be to adopt that recommendation as a new base for the Nushu repertoire and use the list shown in page 4 and 5 with positions marked ‘V’ and ‘D’ removed and the glyphs fixed for the positions marked ‘F’.

Editorial comments

E1. Page 29 Clause 23 – Source references for CJK Ideographs

These two sources should be included in the list of Hanzi G sources, not Hanzi H sources.

Proposed change by Japan

Place these two sources in the list of Hanzi G sources.

Accepted

E2. Page 29 Clause 23 – Source references for CJK Ideographs

It mentions GGFZ-dddddd and GFC-dddddd are inserted in the list for kIRG_Source, but “kIRG_Source” is incorrect.

Proposed change by Japan

Insert GGFZ-dddddd and GFC-dddddd in the list for kIRG_GSource.

Accepted

E3. Sub-clause 24.2 – Source reference file for Tangut Ideographs

As 1st field, it mentions “UCS code point in the format (U+hhhh) or (U+hhhhh).” However, Tangut ideographs are allocated in the SMP.

Proposed change by Japan

Remove “(U+hhhh)”.

Accepted

E4. Sub-clause 24.3 – Source reference presentation for CJK Ideographs

The title of 24.3 is incorrect.

Proposed change by Japan

Change to “24.3 Source reference presentation for Tangut Ideographs”.

Accepted

E5. Page 37 Clause 24 – Character names and annotations

The sub-clause “24.7 Character names for Hangul syllables” already exists. So, this should be renumbered after the insertion of new 25.7 and 25.8.

Proposed change by Japan

Renumber “24.7 Character names for Hangul syllables” to “25.9 Character names for Hangul syllables”.

Accepted in principle

The editor instructions read: Insert two new sub-clauses after 25.6 Character names for CJK Ideographs. An insertion implies the renumbering of any possible following sub-clause. To be more explicit, the instruction will read as follows:

Insert two new sub-clauses after 25.6 Character names for CJK Ideographs and renumber the following sub-clause accordingly.

E6. Sub-clause 25.8 – Character names for Nushu Characters

There is ideograph character after the statement of 25.8.

Proposed change by Japan

Remove it.

Accepted

UK: Negative

Technical/Editorial comment (T or E prefix):

E1. Sub-clause 24.3 – Source reference presentation for CJK Unified Ideographs

Subclause title incorrectly refers to CJK ideographs.

Proposed change by UK:

Change "Source reference presentation for CJK Unified Ideographs" to "Source reference presentation for Tangut Ideographs".

Accepted

See also comment E4 from Japan.

E2. Sub-clause 25.8 – Character names for Nushu Characters

Spurious CJK character at the end of the paragraph.

Proposed change by UK:

Remove "𪛗" at end of paragraph.

Accepted

See also comment E6 from Japan.

E3. Sub-clause 31.2 (32.2) – Code chart

The subclause references the wrong clauses for CJK and Tangut code chart formats.

Proposed change by UK:

Change "See Clause 24 and 25" to "See Clause 23 and 24".

Accepted

E4. Clause 31 (32) – Code charts and list of character names

The list of code points added to existing blocks has some mistakes.

Proposed change by UK:

Change "08B6-08B6" to "08B6-08B7".

Remove "A8FD FE2E-FE2F" (Amd. 1 additions)..

Accepted

T5. Clause 31 (32) – Mongolian

As 181A is named "MONGOLIAN ORNAMENTED BIRGA", the double and triple birga signs should also be named as "ORNAMENTED BIRGA" for consistency. Note, it is conceivable that unornamented double and triple birga symbols could exist, and they would then be named "MONGOLIAN DOUBLE BIRGA" and "MONGOLIAN TRIPLE BIRGA".

Proposed change by UK:

Change the name of 181C from "MONGOLIAN DOUBLE BIRGA" to "MONGOLIAN DOUBLE ORNAMENTED BIRGA".

Change the name of 181D from "MONGOLIAN TRIPLE BIRGA" to "MONGOLIAN TRIPLE ORNAMENTED BIRGA".

WG2 discussion

See also comment T1 from China. Should we encode the additional 13 birgas in a new block as an outcome of the disposition of the Chinese comment, using the convention suggested by UK makes sense.

T6. Clause 31 (32) – CJK Unified Ideographs

The characters at 9FCD through 9FE9 have not been reviewed by WG2, and were not added to the repertoire of Amendment 2 by WG2 resolution, but were added at the discretion of the project editor. We approve of the inclusion of 9FCD through 9FD7 in this amendment, but have concerns about the appropriateness of encoding the eighteen characters at 9FD8 through 9FE9.

Evidence of the usage of 9FD8 through 9FE9 has been provided in WG2 N4583, but conclusive evidence that these are actually individual, discrete characters has not been presented. Each of these characters comprises two CJK ideographs placed side-by-side in a single character space to indicate the pronunciation of a particular syllable of a biblical name, with the right hand ideograph indicating the syllable's initial sound, and the left hand ideograph indicating the syllable's final sound. This method of phonetic notation (known as fanqie in Chinese) is very common for representing Sanskrit, Tangut, Tibetan, Mongolian, Manchu and other languages in pre-modern Chinese linguistic texts, where commonly the two fanqie characters are squeezed together into the space normally used to contain a single ideograph.

In the case of 9FD8 through 9FE9, it is quite possible that these are also pairs of fanqie characters squeezed into a single character space rather than individual characters, in which case displaying the pairs of fanqie characters in a single character space is a presentation issue, which can easily be solved by word processors or in html/css. In the absence of any documentation about these characters by their creator, it is difficult to be sure whether they were intended to be two characters placed together as a phonetic notation, or whether they really were intended to be discrete characters. However, the discussion on pp. 3-4 of N4583 implies that in modern horizontal left-to-right layout the positions of the two ideographs within the "character" may be swapped (so that the fanqie pair is read left-to-right), which is an indication that each "character" is indeed a pair of characters rather than a single, discrete character.

We further note that in the online version of the Russian New Testament translation that uses the characters proposed in N4583 ([http://jesus.tw/Orthodox/New Testament](http://jesus.tw/Orthodox/New_Testament)), the text is laid out horizontally left-to-right. Here the characters in question are represented by superscripting the right hand ideograph of the fanqie pair to the right of the full-sized left hand ideograph, which does not accord with the established methods of indicating an unencoded Han ideograph on the internet. One could either interpret this to mean that the editors of the online edition had no way to represent the 9FD8 through 9FE9 and so used this layout as a hack; or one could interpret this to mean that the editors of the online edition recognised that these were two separate characters used together as an indication of pronunciation, and it did not really matter whether they were squeezed together in a single character space or whether they were superscripted, just as long as the presentation format indicates that they are to be read together.

We are concerned that if the eighteen characters at 9FD8 through 9FE9 are encoded, it will set a precedent for encoding an open-ended set of hundreds of fanqie character pairs used for phonetic notation of other languages. We believe that the wider issue of encoding such phonetic character pairs should be addressed before accepting 9FD8 through 9FE9 for encoding.

Proposed change by UK:

Remove 9FD8 through 9FE9 from Amendment 2.

WG2 discussion

See also document WG2 N4627 which addresses the points raised by UK.

Concerning the first point (lack of review by WG2 and addition at the discretion of the project editor), as described in details in N4627, these characters were discussed at WG2 #62, IRG #41 and IRG #42 where they were approved for further processing. So these characters were adequately qualified for a committee level ballot such as pdam.

Furthermore, the document N4627 takes the position that these characters are not 'fanqie' characters, i.e. they are not used for phonetic notations but are rather transliteration syllables.

T7. Clause 31 (32) – Zanabazar Square

The name for 11A29 ZANABAZAR SQUARE LETTER SMALL A is not appropriate, as this is not a small letter. The character should be named "ZANABAZAR SQUARE LETTER -A" for consistency with the corresponding characters in the Tibetan and Marchen scripts (U+0F60 TIBETAN LETTER –A, U+11C88 MARCHEN LETTER -A). The name of the corresponding character in the Phags-pa script (U+A856 PHAGS-PA LETTER SMALL A) is infelicitous, and should not be taken as a model.

Proposed change by UK:

Change the name for 11A29 from "ZANABAZAR SQUARE LETTER SMALL A" to "ZANABAZAR SQUARE LETTER -A".

Propose acceptance

See also comment T1 from Japan.

E8. Clause 31 (32) – Zanabazar Square

11A29 ZANABAZAR SQUARE LETTER SMALL A should not be aliased as "Tibetan 'a-chung".

Proposed change by UK:

For 11A29 ZANABAZAR SQUARE LETTER SMALL A, remove the alias "Tibetan 'a-chung".

Propose acceptance

The fact that it is cross-referenced to 0F60 TIBETAN LETTER –A and referenced implicitly to 0F71 TIBETAN VOWELL SIGN AA (which itself has the a-chung name alias) indicates that the alias is peculiar.

E9. Clause 31 (32) – Zanabazar Square

It would be helpful to provide comprehensive annotations on the usage of the consonant characters for representing Mongolian, Sanskrit and Tibetan, based on information provided in N4541 §4.7.3..

Proposed change by UK:

For 11A0B ZANABAZAR SQUARE LETTER KA, replace the note "Mongolian ga, gamma" with "Mongolian ga, ya; Sanskrit ka; Tibetan ka".

For 11A0C ZANABAZAR SQUARE LETTER KHA, replace the note "Mongolian ka, qa, xa" with "Mongolian ka, qa, xa; Sanskrit kha, Tibetan kha".

For 11A0D ZANABAZAR SQUARE LETTER GA, add the note "Sanskrit ga; Tibetan ga".

For 11A0E ZANABAZAR SQUARE LETTER GHA, add the note "Sanskrit gha; Tibetan gha".

For 11A0F ZANABAZAR SQUARE LETTER NGA, add the note "Sanskrit ña; Tibetan ña".

For 11A10 ZANABAZAR SQUARE LETTER CA, add the note "Tibetan ca".

For 11A11 ZANABAZAR SQUARE LETTER CHA, add the note "Tibetan cha".

For 11A12 ZANABAZAR SQUARE LETTER JA, add the note "Tibetan ja".

For 11A13 ZANABAZAR SQUARE LETTER NYA, add the note "Sanskrit ña; Tibetan ña".

For 11A14 ZANABAZAR SQUARE LETTER TTA, add the note "Sanskrit ṭa; Tibetan ṭa".

For 11A15 ZANABAZAR SQUARE LETTER TTHA, add the note "Sanskrit ṭha; Tibetan ṭha".

For 11A16 ZANABAZAR SQUARE LETTER DDA, add the note "Sanskrit ḍa; Tibetan ḍa".

For 11A17 ZANABAZAR SQUARE LETTER DDHA, add the note "Sanskrit ḍha; Tibetan ḍha".

For 11A18 ZANABAZAR SQUARE LETTER NNA, add the note "Sanskrit ṇa; Tibetan ṇa".

For 11A19 ZANABAZAR SQUARE LETTER TA, replace the note "Mongolian da" with "Mongolian da; Sanskrit ta; Tibetan ta".

For 11A1A ZANABAZAR SQUARE LETTER THA, replace the note "Mongolian ta" with "Mongolian ta; Sanskrit tha; Tibetan tha".

For 11A1B ZANABAZAR SQUARE LETTER DA, add the note "Sanskrit da; Tibetan da".

For 11A1C ZANABAZAR SQUARE LETTER DHA, add the note "Sanskrit dha; Tibetan dha".

For 11A1D ZANABAZAR SQUARE LETTER NA, add the note "Mongolian na; Sanskrit na; Tibetan na".

For 11A1E ZANABAZAR SQUARE LETTER PA, replace the note "Mongolian ba" with "Mongolian ba; Sanskrit pa; Tibetan pa".

For 11A1F ZANABAZAR SQUARE LETTER PHA, replace the note "Mongolian pa" with "Mongolian pa; Sanskrit pha; Tibetan pha".

For 11A20 ZANABAZAR SQUARE LETTER BA, replace the note "Mongolian va" with "Mongolian va; Sanskrit ba; Tibetan ba".

For 11A21 ZANABAZAR SQUARE LETTER BHA, add the note "Sanskrit bha; Tibetan bha".

For 11A22 ZANABAZAR SQUARE LETTER MA, add the note "Mongolian ma; Sanskrit ma; Tibetan ma".

For 11A23 ZANABAZAR SQUARE LETTER TSA, replace the note "Mongolian ja, Sanskrit ca" with "Mongolian ja; Sanskrit ca; Tibetan tsa".

For 11A24 ZANABAZAR SQUARE LETTER TSHA, replace the note "Mongolian ca, Sanskrit cha" with "Mongolian ča; Sanskrit cha; Tibetan tsha".

For 11A25 ZANABAZAR SQUARE LETTER DZA, replace the note "Sanskrit ja" with "Sanskrit ja; Tibetan dza".

For 11A26 ZANABAZAR SQUARE LETTER DZHA, replace the note "Sanskrit jha" with "Sanskrit jha; Tibetan dzha".

For 11A27 ZANABAZAR SQUARE LETTER ZHA, add the note "Tibetan ža".

For 11A28 ZANABAZAR SQUARE LETTER ZA, add the note "Tibetan z".

For 11A29 ZANABAZAR SQUARE LETTER SMALL A, add the note "Tibetan 'a".

For 11A2A ZANABAZAR SQUARE LETTER YA, add the note "Mongolian ya; Sanskrit ya; Tibetan ya".

For 11A2B ZANABAZAR SQUARE LETTER RA, add the note "Mongolian ra; Sanskrit ra; Tibetan ra".

For 11A2C ZANABAZAR SQUARE LETTER LA, add the note "Mongolian la; Sanskrit la; Tibetan la".

For 11A2D ZANABAZAR SQUARE LETTER VA, add the note "Sanskrit va; Tibetan wa".

For 11A2E ZANABAZAR SQUARE LETTER SHA, add the note "Mongolian ša; Sanskrit śa; Tibetan śa".

For 11A2F ZANABAZAR SQUARE LETTER SSA, add the note "Sanskrit śa; Tibetan ša".

For 11A30 ZANABAZAR SQUARE LETTER SA, add the note "Mongolian sa; Sanskrit sa; Tibetan sa".

For 11A31 ZANABAZAR SQUARE LETTER HA, add the note "Mongolian ha; Sanskrit ha; Tibetan ha".

For 11A32 ZANABAZAR SQUARE LETTER KSSA, add the note "Sanskrit kṣa; Tibetan kṣa"..

Propose acceptance in principle

While not opposed in principle to these additions, there are limits at what can be expressed in the name list. Some of that information would better be conveyed in a Zanabazar informative section in the Unicode Standard which traditionally plays that role for blocks encoded in ISO/IEC 10646.

Furthermore, there may be accentuated characters that are not representable in the font currently used to represent the annotation in the code chart.

Finally it is important to ensure that the naming convention used for the other scripts in these annotations are identical to the actual ISO/IEC 10646 names for these characters. There are clear differences for Mongolian and Tibetan. For example, the suggestion to annotate respectively the zanabazar letters SSA and SA with ‘Sanskrit sa, Tibetan sa’, and Mongolian sa, Sanskrit sa, Tibetan sa’, while Tibetan has both the letters SA and SSA. Using cross references instead of annotation would insure that the annotation are not ambiguous.

T10. Clause 31 (32) – Tangut

The font used for the Tangut code chart is derived from the font used to print Li Fanwen’s 2008 Tangut-Chinese Dictionary (L2008), but in a few cases the glyph in L2008 and its associated font do not correctly represent the glyph form of the corresponding character as found in primary Tangut sources. A review of the Tangut glyph forms in Amd. 2 was carried out by experts from China, Russia and the UK, the results of which are reported in WG2 N4588. The review has identified twenty-seven characters with incorrect glyphs, which should be corrected. In most cases the glyph changes result in a change of stroke count and/or radical, and so the characters also need to be reordered, as detailed in the report.

Proposed change by UK:

Make glyph changes and reorder characters, as detailed in WG2 N4588.

Update TangutSrc.txt with modified stroke counts and radical numbers as appropriate.

WG2 discussion

See also comments T2 from China and T3 from Ireland.

T11. Clause 31 (32) – Nushu

East Asian character repetition marks, such as U+3005 IDEOGRAPHIC ITERATION MARK, U+309D HIRAGANA ITERATION MARK, U+30FD KATAKANA ITERATION MARK, U+A015 YI SYLLABLE ITERATION MARK, and U+16FE0 TANGUT ITERATION MARK all use the term "iteration mark". For consistency, 1B100 should also use the term "iteration mark" rather than "repetition mark".

Proposed change by UK:

Change the name of 1B100 from "NUSHU REPETITION MARK" to "NUSHU ITERATION MARK".

Propose acceptance

See also comments T2 from Japan and TE1 from US.

T12. Clause 31 (32) – Nushu

If a block has a mixture of fixed names and algorithmically-derived names, it can unnecessarily complicate processing of character names by software processes. Therefore it would be best to remove the Nushu iteration mark (1B100) out of the Nushu block.

Proposed change by UK:

Move 1B100 to the new Ideographic Symbols and Punctuation block, at 16FE1.

Reorder 1B101 through 1B28C to 1B100 through 1B28B.

Propose acceptance in principle

See also comments T2 from Japan and TE1 from US.

USA: Positive with comments

Technical comment:

TE.1. Nushu

Based on the questions raised in document WG2 N4610 from Suzuki Toshiya, the US requests the Nushu block be removed from the ballot until additional review and study can be conducted, and the comments from Suzuki are resolved.

Proposed change by US:

Remove Nushu from the ballot. If this request and te.4 are accommodated, the US will change its vote to yes.

Propose acceptance in principle

See also comment T2 from Japan and T12 from UK. It is clear that in all cases another level of committee ballot is required, so the repertoire cannot progress directly at the enquiry level (DAM).

TE.2. Arabic Extended-A

The US requests the following be added to the set of Arabic characters already in the amendment:

U+08D4 ARABIC SMALL HIGH WORD AR - RUB, with glyph and properties as documented in WG2 N4592.

Proposed change by US:

Add the character.

WG2 discussion

TE.3. Arabic Extended-A

The US requests the following 3 characters also be added to the amendment, with glyphs and properties as documented in WG2 N4597:

U+08BB ARABIC LETTER AFRICAN FEH, U+08BC ARABIC LETTER AFRICAN QAF, and U+08BD ARABIC LETTER AFRICAN NOON.

Proposed change by US:

Add the 3 characters.

WG2 discussion

TE.4. Miscellaneous Symbols and Pictographs

Move the character U+1F32D BLACK WANING CRESCENT MOON to U+23FE and rename it "POWER SLEEP SYMBOL"

Rationale: It is expected that this character will co-occur with the 3 other power characters in the PDAM (U+23FB-U+23FD). Placement of this character in the Miscellaneous Technical block will prevent confusion with other emoji moon-like characters also in the Miscellaneous Symbols and Pictographs block.

Proposed change by US:

Move the character and rename it as noted. If this request and te.1 are accommodated, the US will change its vote to yes.

Propose acceptance

Editorial comments:

E.1. Page 1

In the following text, delete “the” as it is unnecessary:

“In the Table 5:”

Proposed change by US:

Delete the unnecessary word.

Accepted

E.2. Page 1

Improve wording in the following text

from:

“These source references are provided in a machine-readable format that is accessible as links to this document. The content pointed by these links is also normative.”

to:

“... format that is accessible as a link from this document. The content pointed to by...”.

Proposed change by US:

Improve the wording as noted.

Accepted

E.3. Clause 23.1 (not in DPDAM2)

A similar wording issue exists in clause 23.1 of the 4th edition:

23.1 List of source references

A CJK Ideograph is always referenced by at least one source reference. These source references are provided in a machine-readable format that is accessible as links to this document. The content pointed by these links is also normative.

Add to Amendment 2 a change to this wording as follows:

“... format that is accessible as a link from this document. The content pointed to by...”.

Proposed change by US:

Add improved wording to Clause 23.1.

Accepted

E.4. Page 1

The following text has an error:

Page 29 Clause 23 Source references for CJK Ideographs In the list of Hanzi H sources, ...

Change “Hanzi H” to “Hanzi G”.

Proposed change by US:

Correct the error.

Accepted

E.5. Page 1

The following line has an error:

“in the line for kIRG_Source”

Change this to:

“in the line for kIRG_GSource”..

Proposed change by US:

Correct the error.

Accepted

E.6. Page 2

In the following text, the antecedent of the restrictive relative clause “that specifies..” would be “end of line mark”:

24.2 Source reference file for Tangut Ideographs

The content linked to is a plain text file, using ISO/IEC 646-IRV characters with CARRAGE RETURN/ LINE FEED as end of line mark that specifies the sources references data for all Tangut Ideographs.

Change the wording as follows (and correct the spelling of “CARRIAGE”):

“...CARRIAGE... as end of line mark. The file specifies...”

Proposed change by US:

Improve the wording as noted and correct the spelling of the word “CARRIAGE”.

Accepted

E.7. Clause 23.2 (not in DPDAM2)

A similar wording issue exists in clause 23.2 of the 4th edition:

23.2 Source references file for CJK Ideographs

The content linked to is a plain text file, using ISO/IEC 646-IRV characters with CARRIAGE RETURN/ LINE FEED as end of line mark that specifies the sources references data for all CJK Ideographs.

Add to Amendment 2 a change to the wording as follows:

“... as end of line mark. The file specifies...”

Proposed change by US:

Add improved wording for Clause 23.2.

Accepted

E.8. Annex G (not in DPDAM2)

Similar issues to comments 4, 5, 6, and 7 also exist in Annex G of the 4th edition:

Annex G (informative)

Alphabetically sorted list of character names

The alphabetically sorted list of character names is provided in machine-readable format that is accessible as a link to this document. The content linked to is a plain text file, using ISO/IEC 646-IRV characters with CARRIAGE RETURN/ LINE FEED as end of line mark, that specifies, after a 4-lines header,

Add to Amendment 2 a change to the wording as follows:

“... as a link from this document. The content... as end of line mark. The file specifies, after a 4-line header, ...”

Proposed change by US:

Add improved wording for Annex G comments 4, 5, 6, and 7 as noted.

Accepted

E.9. Page 3

In the following, improve the text with an additional clause:

Page 37 Clause 24 Character names and annotations

The clause is now renumbered 25. Insert two new sub-clauses after 25.6 Character names for CJK Ideographs

Add at the end:

“, renumbering the following clauses:”

Proposed change by US:

Modify the wording as noted.

Accepted in principle

See disposition of comment E5 from Japan.

E.10. Old Italic

The spelling for the header above U+1032F should be corrected to “Raetic” instead of “Rhetic” (cf. WG2 N4395).

Proposed change by US:

Change the spelling in the header above U+1032F to “Raetic”.

Accepted

E.11. Arabic Extended-A

WG2 N4585 has corrected glyph and name errors for U+08DB, U+08DC, U+08E1, and U+08E2 that appeared in an earlier version of the ballot.

These included the following:

- U+08DB had the dot above the skeleton, instead of below.
- The names list had the wrong glyph for U+08E1 ARABIC SMALL HIGH SIGN SAFHA (it had the glyph for U+08E2 DISPUTED END OF AYAH).
- The names list also had the wrong glyph for U+08DC ARABIC SMALL WORD AN-NISF (it had the glyph for U+08DB ARABIC SMALL HIGH WORD AS-SAJDA)

Proposed change by US:

Note to NBs.

Noted

The US comment as stated is incomprehensible. There is no earlier version of the ballot and document WG2 N4585 is not a ballot document. It is just a draft repertoire for a future amendment. It however its code chart/name list fixes mistakes in the official PDAM2 ballot document as follows:

It fixes the glyph for U+08E2 ARABIC DISPUTED END OF AYAH (it had the glyph for U+08E1 ARABIC SMALL HIGH SIGN SAFHA for both chart and name list (reversed of what the US says for the affected code points, and glyphs for both chart and name list need fixes)

It also fixes the glyph for U+08DC ARABIC SMALL HIGH WORD AN-NISF for both chart and name list (US says only the name list is affected, not the chart)

However it does not fix the other error about U+08DB which still has the dot above the skeleton instead of below. This is fixed in the current working set available to the project editor.

E.12. Latin Extended-D

The header for U+A7AE needs to be changed from “Letter for Gabonese orthographies” to “Letter for African languages”. The header was incorrect, since the letter is used for the Kulongo language which is used in Côte d’Ivoire, not Gabon.

Proposed change by US:

Make the correction as noted.

Accepted