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

Doc. Type: **Disposition of comments**

Title: **Draft Disposition of comments on CDAM1.2 to ISO/IEC 10646 6th edition**

Source: **Michel Suignard (project editor)**
Project: JTC1 02.10646.00.02.00.06
Status: For SC2 consideration
Date: 2022-06-08
Distribution: WG2
Reference: SC2 N4783
Medium: Paper, PDF file

Comments were received from the following members: China, Indonesia, Japan, 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 has been included in this document to facilitate the reading. The dispositions are inserted in between these comments and are marked in **<u>Underlined Bold Serif text</u>,** *with explanatory text in italicized serif*.

The following items reflect the changes to Amendment 1 between CDAM1.2 and the planned DAM1 based on the ballot disposition:

Addition

UCS	Name
2B739	CJK UNIFIED IDEOGRAPH-2B739 (V, M, and U sources)

Glyph changes through this ballot

UCS	IRG source
5F50	new kIRG_Vsource VN-05F50, previous V-source V1-5454 moved to U+2B739
5F50	kIRG_Usource UTC-01005 moved to U+2B739
32351	kIRG_GSource GXM-00265

Annotations, cross reference, source reference added/modified

UCS	Name or IRG source
0675	ARABIC LETTER HIGH HAMZA ALEF (preamble modified for 0675..0678)
1BBD	SUNDANESE LETTER BHA (formal alias missing)

China: Positive with comments

China votes YES to SC2N4783 [CDAM1.2] with comments.

Technical comment:

T1. CJK Unified Ideographs Extension H
The glyph of U+32351 (GXM-00265) should be changed as below:

鱙 vs | 32351 鱙
鱼 195'.7
GXM-00265

Stroke Count should changed to 6 consequently.

This is a simplified form of U+9C59 鱙 thus the "dot" at upper right of the current glyph in code chart should be removed, according to the normalization convention.

Proposed change by China:
Glyph change.

Propose accepted
See also comments TE.2 from USA.

Editorial comment:

E1. On "Digraphs letters for Kazakh"
The preferred spelling forms for U+0675 through U+0678 have been included, that are acceptable and suitable. We still argue for removing the sentence "Use of these characters is discouraged." under the subtitle "Digraphic letters for Kazakh". These four characters are the proper characters in the Chinese national standards, and have been used widely in China for so many years, which are treated as the single characters in Kazakh alphabet. The word "discouraged" will make the users confused. On the other hand, other pre-combined characters have not been marked as "discouraged", such as U+00E0 à, so there is no reason to add this for Kazakh-used Arabic letters. The preferred spelling forms are not equal to the UCD mappings. When the users use the new preferred spelling forms in words, they can not search words used with the four characters.

Proposed change by China:
The note "Use of these characters is discouraged. ……" for U+ U+0675 through U+0678 should be removed.

Proposed accepted in principle
There was a similar comment made for CDAM1, it results in clarification concerning these four letters.

New text (second sentence of preamble is new, and text concerning 'preferred spelling' is also new):

Digraphic letters for Kazakh

Use of these characters is discouraged. They were encoded for Kazakh digraphs, but their decompositions do not reflect the preferred order of representation. Accordingly, the representation of these Kazakh digraphs should instead use the preferred two-character spellings with the correct order of elements.

0675 ٵ ARABIC LETTER HIGH HAMZA ALEF
• preferred spelling is 0674 ٴ 0627 ا
≈ 0627 ا 0674 ٴ

0676 ٶ ARABIC LETTER HIGH HAMZA WAW
• preferred spelling is 0674 ٴ 0648 و
≈ 0648 و 0674 ٴ

0677 ٷ ARABIC LETTER U WITH HAMZA ABOVE
• preferred spelling is 0674 ٴ 06C7 ۇ
≈ 06C7 ۇ 0674 ٴ

0678 ٸ ARABIC LETTER HIGH HAMZA YEH
• preferred spelling is 0674 ٴ 0649 ى
≈ 064A ي 0674 ٴ

There is no dispute about the fact that these digraph letters are in use for Kazakh. The note refers to the encoding of these letters using the precomposed characters. Their compatibility decomposition is however anomalous in using the incorrect order, that cannot be changed and their usage may cause implementation problems when using some normalization forms such as NFKD.. Thus, the representation of these letters should use character sequences starting with U+0674, not ending with that code point.

The Unicode Core Specification Chapter 9, page 394 has the following text:

High Hamza. The characters U+0675 arabic letter high hamza alef, U+0676 arabic letter high hamza waw, U+0677 arabic letter u with hamza above, and U+0678 arabic letter high hamza yeh are not recommended for use. Their compatibility decompositions are anomalous: the decomposed sequences are pairs of letters with right-to-left bidirectional character properties, with U+0674 arabic letter high hamza as the second letter. When processed using the Unicode Bidirectional Algorithm, the hamza will appear to the left of the other letter. Thus, the ordering of characters in the decomposition sequences are the reverse of what is expected. Accordingly, appropriately-ordered pairs of letters beginning with U+0674 arabic letter high hamza should be used. For example, the sequence should be used rather than U+0675 arabic letter high hamza alef.

However, because the issue is not present in all instances (for example it would not present in the case of text using normalization NFC or NFD), it seems that the request from China can be accommodated. The preamble will then read as follows (first sentence removed and text slightly reworked to take into account that removal):

These code points were encoded for Kazakh digraphs, but their decompositions do not reflect the preferred order of representation. Accordingly, the representation of these Kazakh digraphs should instead use the preferred two-characters spellings with the correct order of elements.

Indonesia: Positive with comments

Technical comment:

TE.1. Clause 34. Code charts and lists of character names - Balinese

Charts should show the following three Balinese characters recently recommended by script adhoc (L2/22-068) for future inclusion in the Unicode:

> 1B4E BALINESE INVERTED CARIK SIKI

> 1B4F BALINESE INVERTED CARIK PAREREN

> 1B7F BALINESE PANTI BAWAK

<u>Proposed change by Indonesia:</u>

Please add missing characters

<u>Accepted in principle</u>

Note that these characters are proposed for future inclusion in Unicode, and are not in the current version being under beta review which is Unicode 15.0. It is an implied goal to synchronize Unicode 15.0 with this amendment and therefore it is preferable to consider this inclusion for a future 10646 amendment

TE.2. Clause 34. Code charts and lists of character names – Arabic

One of the Pegon (Javanese-Arabic) diacritic character, *pĕpĕt*, has not been encoded. Some source mentioned that <u>*pĕpĕt* is represented by *maddah*</u>. This replacement seems a temporary alternative because of the lack of true *pepet* shape in the Unicode Standard and ISO.

ê	maddah مدة	ُ	آ ê	اؤ kê	كَرْ krê كَلْ klê	/ə/	ê (e pepet)	ۤ / ۤ ê / (pepet)

<u>Proposed change by Indonesia:</u>

Please encode the *pĕpĕt* diacritic character for Javanese and Sundanese Arabic script.

Name suggestion: ARABIC FATHA WITH ELBOW

Shape: ﹈

Use:

لَكَجَةْ

(lajĕng)

سَاكَتْ مَلَبَتْ

(sagĕt mĕlĕbĕt)

<u>Propose not accepted</u>

Such as request should be made through a proposal document. Among others, it should explain how it is different from a maddah (currently encoded at U+0653 ARABIC MADDAH ABOVE). And if accepted it should probably go in a future amendment.

Editorial comments:

ED.1. Clause 34. Code charts and lists of character names - Sundanese

Three Sundanese characters has been discussed on L2/21-221 and recommended by the script adhoc on L2/21-174. But the ISO draft seems not followed all of the recommendation yet.

Sign

1BBA	ꮺ	SUNDANESE AVAGRAHA
		= gemination mark

Historic letters

1BBB	ꮻ	SUNDANESE LETTER REU
		• vocalic r
1BBC	ꮼ	SUNDANESE LETTER LEU
		• vocalic l
1BBD	ꮽ	SUNDANESE LETTER BHA
1BBE	ꮾ	SUNDANESE LETTER FINAL K
1BBF	ꮿ	SUNDANESE LETTER FINAL M
		• used in a 21st century document
		• for actual final m, use the sequence 1B99 ꮺ
		1BAA ○

Proposed change by Indonesia:
Please make correction based on consensus L2/21-167.

Proposed accepted
See also comment TE.3 from USA
This was an oversight and will be corrected in the next phase of this amendment. The entry for U+1BBD will read as follows:

1BBD	ꮽ	SUNDANESE LETTER BHA
		※ SUNDANESE LETTER ARCHAIC I
		• character was originally misidentified, so its name is a misnomer

ED.2. Clause 34. Code charts and lists of character names - Kawi

Many Kawi glyphs in the name list (p. 179) are inexplicably clipped, such as:

Independent vowels

11F04	अ	KAWI LETTER A
11F05	आ	KAWI LETTER AA
11F06	इ	KAWI LETTER I
11F07	ई	KAWI LETTER II
11F08	उ	KAWI LETTER U
11F09	ऊ	KAWI LETTER UU
11F0A	ऋ	KAWI LETTER VOCALIC R
11F0B	ॠ	KAWI LETTER VOCALIC RR
11F0C	ऌ	KAWI LETTER VOCALIC L
11F0D	ॡ	KAWI LETTER VOCALIC LL
11F0E	ए	KAWI LETTER E
11F0F	ऐ	KAWI LETTER AI
11F10	ओ	KAWI LETTER O

Proposed change by Indonesia:
Please change text settings so that Kawi glyphs are shown in full like the glyphs in the character table

Proposed accepted
The font has been improved with more interlinear spacing and now shows as:

Independent vowels

11F04	अ	KAWI LETTER A
11F05	आ	KAWI LETTER AA
11F06	႒	KAWI LETTER I
11F07	⟨႒	KAWI LETTER II
11F08	ࣿ	KAWI LETTER U
11F09	ࣿ	KAWI LETTER UU
11F0A	႘	KAWI LETTER VOCALIC R
11F0B	⟨႘	KAWI LETTER VOCALIC RR
11F0C	႙	KAWI LETTER VOCALIC L
11F0D	႙	KAWI LETTER VOCALIC LL
11F0E	౭	KAWI LETTER E
11F0F	౭	KAWI LETTER AI
11F10	౭	KAWI LETTER O

Japan: Negative

Technical comment:

TE.1. Clause 34. Code charts… – CJK Unified Ideographs Ext H – 31F68

At CDAM 1 ballot, Japan commented that U+31F68 above in CJK-H (U+31F69 in CDAM 1 text) should be unified with U+26C25.

Disposition of comments for this mentioned

"For that case, https://hc.jsecs.org/irg/ws2017/app/?id=03689note says: not unified to U+26C25, irg50".

and the comment has not been accepted.

However, after IRG#50 (2018.05.21-2018.05.25), there are a few comments on above discussion records saying they are unified, such as "The two characters are variants without a doubt. An identical glyph is also in the Moji-joho database mapped directly to 邦邦 (U+26C25):" on 2021-11-10. Those comments remain unsettled.

<u>Proposed change by Japan:</u>
Please delete U+31F68 from CJK-H and reconsider if it could be unified with U+26C25.

<u>Propose not accepted</u>
As noted in the previous disposition of comments for this amendment (N5173), issues concerning CJK Unified Ideographs are typically discussed/settled by the Ideographic Research Group (IRG) which has been meeting virtually twice a year. This makes easy to all interested parties to participate. In addition, IRG members used the tool located at <u>https://hc.jsecs.org/irg/ws2017/app/</u> to document decisions related to the WS2017 set base for CJK Ext H and Japanese experts have access that tool.
It is rather unfortunate that these comments were not done earlier and in the appropriate forum and were not brought to attention in the last IRG meeting in March 2022.
Strictly speaking, while the latest comments in the tool were asking to unify, no action was taken to that effect in the actual Extension H content provided to the project editor (request to unify is not same thing as unified).

It should also be noted that while this amendment is still in CDAM phase, the corresponding Unicode version 15.0 is in beta review which makes code point changes complicated. In addition, unlike other blocks, it is highly unpractical to leave holes in CJK Unified Ideographs blocks.

Another IRG expert reviewing this comment said: "Although U+31F68 is unifiable to U+26C25, the two characters are not identical in shape, so there is no particular harm in encoding U+31F68. Certainly, there are already dozens of other similar examples of unifiable pairs of characters encoded, so at this late stage I would strongly support leaving U+31F68 in Ext. H."

TE.2. Clause 34. Code charts … – CJK Unified Ideographs Ext H – 31F4C

According to the source information of U+31F4C on CJK-H and U+2CECB on CJK-F below, both of them are the "abbreviated" form (略記) of frequent Buddhism term "菩提" ("Bodhi"). Thus, they should be considered as a cognate.

Because the difference of their shapes is only the position of the dot, they might be unifiable.

Note that the source information of U+2CECB (MJ-056833) below mentions it is pronounced as "Bodhi."

The evidence for UK-10352

The evidence for JMJ-056833

(Kadokawa Daijigen, ISBN 9784040128009, p.2116)

<u>Proposed change by Japan:</u>

Please delete U+31F4C from CJK-H and reconsider if it could be unified with U+2CECB.

Propose not accepted

Note that the IRG tool at <u>https://hc.jsecs.org/irg/ws2017/app/?id=03653</u> has no comment on this except to possibly change its IDS description.

In addition, the UK expert that originally proposed that ideograph made the following comment:

> *As to U+31F4C, this is only potentially unifiable to U+2CECB, and as far as I can tell there is no UCV to support unification. The Japanese comment suggests the two characters *might* be unifiable because they only differ by the position of the dot; however, that is not the only difference in the glyph forms of the two characters, and unification could only be determined by IRG, not in the disposition of comments by the project editor. This is a UK-submitted character, and we would strongly object to removing it from Ext. H on the basis that theoretically it *might* be unifiable with U+2CECB.*

USA: Positive with comments

Technical Comments

TE.1. 34 (page 292) – CJK Unified Ideographs Extension C

Due to the IRG's recommendation to accept an urgently needed character from Macao Special Administrative Region, China, which necessitated the disunification of U+5F50, a new code point is needed. The code point recommended is U+2B739 in Extension C. See IRG recommendation M58.6 in https://www.unicode.org/L2/L2022/22078-irgn2530-irg58-report.pdf , based on Submission from Macao Special Administrative Region, China document (IRG N2538).

Proposed change by USA:

Add U+2B739 to CJK Unified Ideographs Extension C

Proposed accepted

See also documents:

IRGN2543 at https://www.unicode.org/L2/L2022/22042-irgn2543-utc-unc.pdf ,
IRGN2538R at https://www.unicode.org/L2/L2022/22076-irgn2538r-macao.pdf),
L2/22-067 at https://www.unicode.org/L2/L2022/22067-cjk-unihan-group-utc171.pdf , and
L2/22-091 at https://www.unicode.org/L2/L2022/22091-irgn2509-vietnam-resp.pdf

Result for U+5F50 and the new proposed U+2B739 in Extension C:

Old:

New:

TE.2. 34 (page 346) – CJK Unified Ideographs Extension H – U+32351

The glyph for the G-Source ideograph of U+32351 (GXM-00265) has been identified as an error. The top image below is the current glyph (in yellow), and below it is the correct glyph:

Proposed change by USA:

Correct the glyph for the G-Source ideograph of U+32351 to that shown below:

Proposed accepted

See also comment T1 from China

Page 10

TE.3. 34 (page 59) Sundanese

The formal name alias was missing in the code charts for U+1BBD SUNDANESE LETTER BHA. The formal name alias should be SUNDANESE LETTER ARCHAIC I. Rationale: The name "bha" is now considered by scholars to be incorrect; the correct reading is "I" For further details, see
https://www.unicode.org/L2/L2021/21221-three-sundanese-chars.pdf

Proposed change by USA:
Add the formal name alias SUNDANESE LETTER ARCHAIC I to U+1BBD SUNDANESE LETTER BHA.
Proposed accepted
See also comment ED.1 from Indonesia and its disposition of comment.

Editorial Comments

ED.1. 34.2 (page 6) Kaktovik
Kaktovik is misspelled.

Proposed change by USA:
Change "Kaktovic" to "Kaktovik".
Propose accepted

ED.2. Annex A.1 (page 8) Kaktovik
Kaktovik is misspelled twice on page 8.

Proposed change by USA:
Change "Kaktovic" to "Kaktovik".
Propose accepted

ED.3. Annex A.2.2 (page 9) Kaktovik
Kaktovik is misspelled.

Proposed change by USA:
Change "Kaktovic" to "Kaktovik".
Propose accepted