

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

Doc. Type: Draft disposition of comments

Title: Draft disposition of comments on DAM2 text 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: 2015-10-11

Distribution: WG2

Reference: SC2 N4428 WG2 N4667 N4684

Medium: Paper, PDF file

Comments were received from Ireland, Japan, UK, and US. The following document is the draft disposition of those comments.

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 **Underlined Bold Serif text**, *with explanatory text in italicized serif.*

Ireland: Negative

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

Technical comments

T1. Page 14, Row 08A0: Arabic Extended-A.

Ireland requests the change of the name of 08E2 from DISPUTED END OF AYAH to ARABIC DISPUTED END OF AYAH. Compare 06DD ARABIC END OF AYAH.

Propose acceptance

T2. Page 65, Row 17000: Tangut.

With reference to N4667 “Review of Tangut repertoire in DAM ballot” Ireland requests that the revisions to the Tangut encoding be revised according to the recommendations in that document.

WG2 discussion

See also comments T1 to T5 from UK and their dispositions which describe in details the changes to be made. The change concerning U+17CF0 is controversial and will require further discussion.

Editorial comments

E1. Page i, Title page.

Ireland requests that the name “Osage” be added to the title of the Amendment.

Accepted in principle

The title can be changed on the document, but it may be too late to change the project sub-division title which is typically aligned with the document title.

E2. Page 24, Row 1C80: Cyrillic Extended-C.

Ireland requests that the glyphs be improved slightly. Some experts have provided comments and an updated font is available.

Accepted in principle

Pending the actual delivery and review of the updated font.

E3. Page 25, Row 20A0: Currency Symbols.

Ireland requests a change to the glyph of 20BE LARI SIGN so that the right hand curve is longer and thicker. Also, on the following page most of the glyphs are not shown.

Accepted

E4. Page 41, Row 3CED: CJK Unified Ideographs Extension-A.

Ireland requests that the winding errors for 4C88-K3-3661 and 4C89-K3-3663 be corrected.

Accepted in principle

The editor could not reproduce the error. It should be noted that the amendment only reproduces a very small fraction of that block.

E5. Page 144, Row 1F300: Miscellaneous Symbols and Pictographs.

Ireland requests that the glyph for 1F3F9 BOW AND ARROW be centred properly.

Accepted

E6. Page 144, Row 1F300: Miscellaneous Symbols and Pictographs.

Ireland requests a cross-reference for 1F3FA AMPHORA to U+100E8 LINEAR B IDEOGRAM VESSEL B209.

Accepted

Japan: Negative

Technical comments

T1. Page 29, Sub-clause 23.1 List of source references.

Out of CJK J source characters having JIS X0212 source and JA source, the source reference of the characters also encoded in JIS X0213 are changed. In order to represent the meaning of source reference clearly, JNB would like to change the “index” of each source reference. For example, from “J3B” to “J13” as follows.

J3B JIS X 0213:2004 level-3 characters replacing J1 characters
to

J13 JIS X 0213:2004 level-3 characters replacing J1 characters

Here, “1” of “J13” above means it was originally J1 source character and “3” of “J13” means it is encoded in J3 source afterward.

Proposed change by Japan.

Please change as follows.

J3B JIS X 0213:2004 level-3 characters replacing J1 characters
change to

J13: JIS X 0213:2004 level-3 characters replacing J1 characters

J3C JIS X 0213:2004 level-3 characters replacing JA characters
change to

JA3: JIS X 0213:2004 level-3 characters replacing JA characters

J4B JIS X 0213:2004 level-3 characters replacing J1 characters
change to

J14: JIS X 0213:2004 level-4 characters replacing J1 characters

J4C JIS X 0213:2004 level-3 characters replacing JA characters
change to

JA4 JIS X 0213:2004 level-4 characters replacing JA characters

Newly add

J13A: JIS X 0213:2004 level-3 character addendum from JIS X 0213:2000 level-3 replacing J1 characters"

Propose acceptance

The only change that requires clarification is for the newly added J13A. It is the understanding of the editor that this new source only concerns two characters: U+7626 and U+7E6B which would see their source go from a J3A type to a J13A type. The other 8 characters having now J3A sources (U+4FF1, U+525D, U+541E, U+5653, U+59F8, U+5C5B, U+5E77, and U+20B9F) keep that value.

T2. Page 29, Sub-clause 23.1 List of source references.

It is mentioned that

“Replace all J1 and JA sources that are also included in JIS X 0213:2004 by the values defined in JIS X 0213:2004 using the prefixes described above (J3B, J3C, J4B, and J4C).”

Same as above, we request to change the source reference information.

Proposed change by Japan.

Please change the source reference information in the code charts.

- From J3B to J13
- From J3C to JA3
- From J4B to J14
- From J4C to JA4
- From J3BA to J13A

Propose acceptance in principle

The last change is incorrect, there is no J3BA sources; J13A would be a new source. Two of the current J3A sources (identified in disposition of the comment T.1 would become J13A.

T3. Page 29, Sub-clause 23.1 List of source references.

As described in Annex S.1.6, JIS X 0212-1990 is one of the source standards that the source code separation rule is applied.

Since the source reference information for some of “JIS X0212-1990” source characters is changed as proposed on this amendment, it would be difficult to recognize JIS X0212 characters when the source code separation rule is applied. JNB requests to add some text to clarify it.

Proposed change by Japan.

Please see the proposed text in the comment on Annex P below.

Propose acceptance in principle

There is a note added by this amendment for the sub-clause S.1.6 ‘Source separation rule’ that indicates that the collection 372 JAPANESE IDEOGRAPHICS SUPPLEMENT identifies the JIS X 0212-1990 characters. Therefore, it should not be difficult to recognize these characters. This does not prevent some additional clarification as suggested for Annex P.

T4. Page 29, Sub-clause 23.1 List of source references.

It is mentioned that

“in the line for kIRG_JSource, insert the additional third field format values:

J3B-hhhh, J3C-hhhh, J4B-hhhh, and J4C-hhhh.

As commented above on sub-clause 23.1, we request to change the source reference information.

Proposed change by Japan.

Please replace

J3B-hhhh, J3C-hhhh, J4B-hhhh, and J4C-hhhh.

with

J13-hhhh, JA3-hhhh, J14-hhhh, JA4-hhhh, and J13A-hhhh.

Propose acceptance

T5. Page 29, Sub-clause 23.1 List of source references.

It is mentioned that

“Provide a new reference file with updated references.

(The following text is identical to ISO/IEC 10646:2014 but is linked to the new file)”

Then, new “CJKSrc.txt” is provided.

As commented above, please change the source reference information in “CJKSrc.txt”.

Proposed change by Japan.

Please correct the source reference information in “CJKSrc.txt” file.

- From J3B to J13
- From J3C to JA3
- From J4B to J14
- From J4C to JA4
- From J3BA to J13A

Proposed acceptance in principle

The last change is incorrect, there is no J3BA sources. J13A is a new source.

T6. Page 2393, Annex A.4.3 371 JIS 2004 IDEOGRAPHICS EXTENSION

Because new indexes for J source reference are added, the statement should be updated.

Proposed change by Japan.

Please change

“The code points of this collection are identified by 1 249 J3 Kanji J sources, 10 J3A Kanji J sources, and 2 436 J4 Kanji J sources in the Source Reference file for CJK Unified Ideographs (CJKSrc.txt). See 23.1 and 23.2 for details.

to

“The code points of this collection are identified by J3 Kanji J sources (J3, J3A, J13, J13A and JA3) and J4 Kanji J sources (J4, J14 and JA4) in the Source Reference file for CJK Unified Ideographs (CJKSrc.txt). See 23.1 and 23.2 for details.”

Proposed acceptance

T7. Page 2393, Annex A.4.3 372 JAPANESE IDEOGRAPHICS SUPPLEMENT

The number of characters that are common between the collections 371 and 372 is wrong.

Proposed change by Japan.

Please change

“NOTE – 2 742 characters are common between the collections 371 and 372.”

to

“NOTE – 2 743 characters are common between the collections 371 and 372.”

Proposed acceptance

T8. Page 2423, Annex I.2 Syntax of an ideographic description sequence

“A coded stroke”, the character on the CJK STROKES block (U+31C0 - U+31EF), is proposed as one of DCs (Description Component).

JNB understands some characters such as 𠂇𠂈 can be represented as IDS sequences if the CJK stroke (U+31E3 CJK STROKE Q) is allowed.

However, we think the following impact negates the above merit.

Since CJK STROKES are too elementary as a glyphic component to describe the structure of CJK Ideographs, there could be several variations of IDS for one ideograph character if using CJK STROKES. It would introduce unexpected ambiguity into IDS. As one of practical usages for IDS, IRG has the procedure to verify if the proposed ideograph is not duplicated with encoded ideographs by comparing their IDSs. By introducing CJK STROKES as DC, it is difficult to verify because the difference on stroke information does not necessarily mean whether they should be unified or not.

Therefore, JNB thinks the characters on CJK STROKES should not be allowed to use as DC.

Proposed change by Japan.

Delete the item “a coded stroke, which consists of any coded character from the CJK STROKES block.”

Proposed non acceptance

There are clearly arguments pro and con for allowing the CJK Strokes in DCs. However, the fact that some CJK Ideographs cannot be represented in their absence seems a compelling argument. To reiterate the argument made in the disposition for a similar comment made in page 6 of N4664 ‘Disposition of comments on PDAM2.3 to ISO/IEC 10646 4th edition):

CJK strokes should not be removed from the list of IDC elements, as characters such as 𠂇𠂈 can only be represented as IDS sequences if the 〇 stroke (U+31E3 CJK STROKE Q) is allowed (some people use U+3007 IDEOGRAPHIC NUMBER ZERO 〇, but that is not correct).

The argument that allowing CJK strokes will introduce ambiguous sequences as some strokes are visually the same as some encoded characters is not convincing as already all Kangxi radicals and most CJK radicals supplement characters are visual clones of unified ideographs, and so a process dealing with IDS sequences already needs to deal with visual ambiguity.

Concerning their elementary shape, it is true that they should only be used in cases as above where they bring shapes not available elsewhere and should not be used to decompose characters in lengthy sequences where more complex components (such as radicals or other ideographs) should be used instead. In other words, CJK STROKES should only be used sparingly but they nevertheless have some good use cases.

T9. Page 2453, Annex P Additional information on CJK Unified Ideographs

There is the description about the J1 source characters and JA source characters which source reference information is changed as follows.

“The set consists of 2828 characters, of which 2723 were part of JIS X 0212-1990 (referred as J1 sources in this International Standard) and 85 were part of the ‘Unified Japanese IT Vendors Contemporary Ideographs, 1993’ (referred as JA in this International Standard).”

JNB thinks the list of all 2828 characters are helpful to the user.

Proposed change by Japan.

Please attach the file containing the list of 2828 characters. JNB is separately sending the list to the project editor.

Proposed acceptance in principle

The project editor has not received yet the file. However, it seems better to create a collection in Annex A describing these characters and refer to that collection in Annex P. In other words, enumerated list of code points should be all described in Annex A for consistency.

T10. Page 2453, Annex P Additional information on CJK Unified Ideographs

JNB requests to revise the following text in Annex P to reflect the comments above that are related to Annex P.

“The Japanese Industrial Standard JIS X 0213:2004 contains characters in its levels 3 and 4 which were part of previous JIS standards. The set consists of 2828 characters, of which 2723 were part of JIS X 0212-1990 (referred as J1 sources in this International Standard) and 85 were part of the ‘Unified Japanese IT Vendors Contemporary Ideographs, 1993’ (referred as JA in this International Standard). Among these 2828 characters, 205 had their glyph representation slightly modified. These 2828 characters use the ‘J3B, J3C, J4B, and J4C’ notations in their prefix four source identification as described in subclause 23.1.”

Proposed change by Japan.

Please change the description of Annex P as follows. (Comment to project editor: Some part of the figure below are truncated. Please contact JNB for whole picture of figure.)

The Japanese Industrial Standard JIS X 0213:2004 includes a set of characters which were part of previous JIS standards. The set consists of 2828 characters, of which 2743 were part of JIS X 0212-1990 (referred as J1 sources in the previous version of this International Standard) and 85 were part of the ‘Unified Japanese IT Vendors Contemporary Ideographs, 1993’ (referred as JA in the previous version of this International Standard). These 2828 characters use the ‘J13, J13A, J14, JA3, JA4’ notations in their prefix five source identification as described in sub-clause 23.1. The glyph representation of them also reflects the one shown in JIS X 0213:2004.

The list of 2828 characters is provided in machine-readable format that is accessible as a link from 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, each line contains the following information organized in fields delimited by the TAB character:

- 1st field: UCS code point in the format 'hhhh', 'h' being a hexadecimal unit,
- 2nd field: Source reference in previous editions of this standard.
- 3rd field: Source reference in this standard.

Click on this highlighted text to access the reference file.

NOTE – The content is also available as a separate viewable file in the same file directory as this document. The file is named: “xxxx.txt”. (To project editor: please assign the appropriate name.)

Out of the pairs of ideographs which are exceptions to the unification rules described in S.1, the source reference of either ideograph of the pairs shown below are JIS X 0213:2004 level-3 characters replacing J1 characters.

填 填	TJ
5861 586B	

[U+5861] J1-3834 → [J3B-2F58](#) [U+586B] J0-4536

擊 擊	TJ
6483 64CA	

[U+6483] J0-3762 [U+64CA] J1-4123 → [J3B-7522](#)

瘦 瘦	J
75E9 7626	

[U+75E9] J0-4169 [U+7626] J1-4D77 → [J3BA-7E7D](#)

菑 菑	TJ
83D1 8458	

[U+83D1] J1-583B → [J3B-7B21](#) [U+8458] J1-5858

蔣 蔣	GJ
848B 8523	

[U+848B] J0-3E55 [U+8523] J1-5936 → [J3B-7B36](#)

醬 醬	J
91A4 91AC	

[U+91A4] J0-3E5F [U+91AC] J1-6273 → [J3B-7C79](#)

顛 顛	J
985A 985B	

[U+985A] J1-6833 → [J3B-7E23](#) [U+985B] J0-453F

Proposed acceptance in principle

As hinted in the disposition of comment T9, it is preferable to move the content definition of the collection of 2828 characters to Annex A, using a new collection number (for example collection 374 under a new A.4.5 subclause). Concerning the graphic description above, it duplicates what is already described in Annex S.3 Source separation examples. Therefore, it makes more sense to add a note to S.3, covering these 7 pairs by indicating that some of the J sources of these pairs have J13 (not J3B) source type indicating their original source as JIS X 212-1990). The note could be phrased as follows:

NOTE – Seven pairs described below have J sources which were originally part of the JIS X 212-1990 level-3, therefore covered by the Source separation rule, which are also part of the JIS X 213:2004 with source identified as J13. The pairs (J13 character code points emphasized) are:
U+5861-U+586B, **U+6483-U+64CA**, **U+75E9-U+7626**, **U+83D1-U+8458**, **U+848B-U+8523**,
U+91A4-U+91AC, and **U+985A-U+985B**.

United Kingdom: Negative

Technical comments

T1. Tangut U+17CF0

As described in WG2 N4667, there is an error in the glyph for 17CF0.

Proposed change by U.K.

Correct the glyph for 17CF0 (no reordering is required).

WG2 discussion

See comment T2 from Ireland.

The document N4667 requests a change for the middle component while document N4684 from Chinese experts requests a change for the left component.

T2. Tangut U+17F24

As described in WG2 N4667, 17F24 is incorrectly ordered.

Proposed change by U.K.

Reorder 17F24 to 17F20, and move 17F20..17F23 to 17F21..17F24.

Propose acceptance

See comment T2 from Ireland.

T3. Tangut U+18316

As described in WG2 N4667, 18316 is incorrectly ordered.

Proposed change by U.K.

Reorder 18316 to 18314, and move 18314..18315 to 18315..18316.

Propose acceptance

See comment T2 from Ireland.

T4. Tangut U+186A1 and U+186A2

As described in WG2 N4667, 186A1 and 186A2 are insignificant glyph variants of the same character, and should have been unified according to the unification principles in WG2 N4522.

Proposed change by U.K.

Remove 186A2, and move 186A3..187ED to 186A2..187EC.

Use the 186A2 glyph for 186A1 for consistency with 186A0.

Change the source reference for 186A1 to L2008-0523-4565.

Propose acceptance

See comment T2 from Ireland.

This is also supported by document N4684 (Chinese experts).

T5. Tangut Components

As described in WG2 N4667, two Tangut components proposed in N4326 were removed in N4636. The removal of these two components was a mistake as they are required to represent Tangut stroke types.

Proposed change by U.K.

Insert the two Tangut components shown in N4667 at 18817 and 18818, and move 18817..18AF0 to 18819..18AF2..

Propose acceptance

See comment T2 from Ireland.

United States: Positive with comments

Technical comment

T1. Tamil Supplement

The US requests the Tamil Supplement block be removed from the ballot until additional review and study can be conducted.

Proposed change by U.S.

Remove the Tamil Supplement block. If te.1 is accommodated, the US changes its vote to Yes.

Propose acceptance

The US comment is already positive, so no vote change is expected. There is no urgency to the encoding of these characters and some of the request from postponement originally came from the Tamil communities. While the block will be removed from the Amendment 2 it will placed back into the next committee level ballot (i.e. 5th edition of 10646, either a new CD or amendment 1).

Editorial comment

E1. Page 11 Annex I

In the section under “Page 2423, Annex I.2” on page 11 (=PDF page 13), the following sentence has an incorrect subject/verb agreement: "The subgroups and their contents is as follows:"

Proposed change by U.S.

Change “is” to “are”.

Accepted
