# ISO/IEC JTC1/SC2/WG2 N **5245**Date: 2024-01-03

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

Doc. Type: Disposition of comments

Title: Disposition of comments on CDAM2.3 to ISO/IEC 10646 6th edition

Source: Michel Suignard (project editor)
Project: JTC1 02.10646.00.02.00.06
Status: For SC2 consideration

Date: 2024-01-03
Distribution: WG2
Reference: SC2 N4891
Medium: Paper, PDF file

Comments were received from the following members: Canada, Czech Republic, Japan, Korea (ROK), Poland, United Kingdom, and USA. The following document is the disposition of those comments.

The disposition introduces 10 new characters to the repertoire:

- 1 Khitan Small Script character as:
  - 18CFF KHITAN SMALL SCRIPT-18CFF,
- 7 pictographs as:
  - o 1FA89 HARP
  - o 1FA8F SHOVEL
  - o 1FABE LEAFLESS TREE
  - o 1FAC6 FINGERPRINT
  - 1FADC ROOT VEGETABLE
  - 1FADF SPLATTER
  - 1FAE9 FACE WITH BAGS UNDER EYES
- 2 CJK Stroke characters as:
  - o 31E4 CJK STROKE HZXG
  - 31E5 CJK STROKE SZP

It also modifies the glyph of 4 Tangut characters (U+17121, U+17C51, U+17D0B, U+180DF), one Khitan Small Script character (U+18BD2) and one K source horizontal extension (U+25A38, K source: K6-110D).

It also adds 8 standardized variation sequences and moves two J sources JMJ references newly introduced in this amendment.

These new characters are described in the following table which includes all the proposed new ranges with pointers to the relevant documentations:

Name	Ranges	Document(s)
Khitan Small Script	18CFF	https://www.unicode.org/L2/L2023/23065-n5205-khitan-small-blank.pdf
Symbols and Pictographs Extended-A	1FA89 1FA8F 1FABE 1FAC6 1FADC 1FADF 1FAE9	https://www.unicode.org/L2/L2023/23236r-esc-report-q4-2023.pdf
CJK Strokes	31E4-31E5	https://www.unicode.org/L2/L2023/23220-new-cjk-strokes.pdf

(Details of the comments and disposition in the following pages)

Date:2024-01-03	Document ISO/IEC 10646:2020	Project:
Date.2024-01-03	(Ed.6)/CDAM 2.3	1.02.10646.00.02.00.06

MB/ NC¹	Line number	Clause/ Subclause	Paragraph/ Figure/Table	Type of comment <sup>2</sup>	Comments	Proposed change	Observations of the secretaria
US- 001		page 183ff.	Tangut	te	Glyphs for the following four Tangut characters should be corrected: U+17121, U+17C51, U+17D0B, U+180DF, based on glyphs in the recently-published dictionary of Tangut words, Xīxiàwén Cídiǎnby, and shown in WG2 N5237. The document, by Andrew West, has been reviewed and the glyphs confirmed by Tangut experts Dr. Sun Bojun and Prof. Nie Hongyin.	Correct the glyphs for the four Tangut characters U+17121, U+17C51, U+17D0B, U+180DF, as documented in WG2 N5237.	Accepted
US- 002			Khitan Small Script	te	The glyph for the Khitan Small Script character U+18BD2 KHITAN SMALL SCRIPT CHARACTER-18BD2 should be corrected to the form shown on the original epigraphic source, as documented in L2/23-199. This change has been agreed to by Dr. Sun Bojun.	Correct the glyph for Khitan Small Script character U+18BD2, as documented in L2/23-199.	Accepted
US- 003			Khitan Small Script	te	The character KHITAN SMALL SCRIPT CHARACTER-18CFF, documented in WG2 N5205, is urgently needed. We request it be added to AMD 2, which will also make ISO/IEC 10646 be in sync with Unicode version 16.0.	Add the character.	Accepted
US- 004			Symbols and Pictographs Extended-A	te	The following 7 characters have been included in the proposed repertoire for Unicode version 16.0. In order for AMD 2 to stay in sync with Unicode version 16.0, we recommend the following 7 characters be added (with names and code points as documented in L2/23-236):  1FA89 HARP (L2/23-256)  1FA8F SHOVEL (L2/23-259)  1FABE LEAFLESS TREE (L2/23-255)  1FAC6 FINGERPRINT (L2/23-258)  1FADC ROOT VEGETABLE (L2/23-258)  1FADF SPLATTER (L2/23-261)  1FAE9 FACE WITH BAGS UNDER EYES (L2/23-260)	Add the 7 characters.	Accepted
US- 005		p. 24	CJK Strokes	te	Two CJK strokes have been identified as being very prevalent in use, but are incorrectly unified	Add the two characters.	Accepted

<sup>1</sup> MB = Member body / NC = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by \*\*)
2 Type of comment: ge = general te = technical ed = editorial

MB/ NC¹	Line number	Clause/ Subclause	Paragraph/ Figure/Table	Type of comment <sup>2</sup>	Comments	Proposed change	Observations of the secretaria
					with other strokes. To rectify the situation, we recommend adding the following two characters: U+31E4 CJK STROKE HZXG and U+31E5 CJK STROKE SZP, based on the evidence provided in L2/23-220. We recommend they be added to this amendment, since in CJK work, character disunifications are handled as <i>de facto</i> urgently needed characters. In addition, this character will keep ISO/IEC 10646 in sync with Unicode version 16.0.		
US- 006			Standardized variation sequences	te	Handling of the four quotation marks U+2018, U+2019, U+201C and U+201D in East Asian text has long been a problem. To address this issue, standardized variation sequences have been proposed (L2/23-212R) and have been included in Unicode version 16.0. In order to stay in sync with Unicode 16.0, we recommend adding the following 8 standardized variation sequences: 2018 FE00; non-fullwidth form; # LEFT SINGLE QUOTATION MARK 2018 FE01; right-justified fullwidth form; # LEFT SINGLE QUOTATION MARK 2019 FE00; non-fullwidth form; # RIGHT SINGLE QUOTATION MARK 2010 FE01; left-justified fullwidth form; # RIGHT SINGLE QUOTATION MARK 201C FE00; non-fullwidth form; # LEFT DOUBLE QUOTATION MARK 201C FE01; right-justified fullwidth form; # LEFT DOUBLE QUOTATION MARK 201D FE00; non-fullwidth form; # RIGHT DOUBLE QUOTATION MARK 201D FE01; left-justified fullwidth form; # RIGHT DOUBLE QUOTATION MARK 201D FE01; left-justified fullwidth form; # RIGHT DOUBLE QUOTATION MARK	Add the 8 standardized variation sequences.	Accepted
JP1- 007		35.5 Code charts and lists of character		te	After proposing SC2 N 4849, horizontal extension, JNB's experts carefully rechecked it and simultaneously asked for public comment on it. As a result, JNB concluded that the proposed code	Change the code point for JMJ-034451 to 21D14.	Accepted

<sup>1</sup> MB = Member body / NC = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by \*\*)
2 Type of comment: ge = general te = technical ed = editorial

MB/ NC¹	Line number	Clause/ Subclause	Paragraph/ Figure/Table	Type of comment <sup>2</sup>	Comments	Proposed change	Observations of the secretariat
JP2-		names, 24.2 Source references file for CJK Ideographs (CJKSrc.txt) 35.5 Code		te	point 21D12 for JMJ-034451 is not appropriate, and it is necessary to be changed.  21D12  45.6  T5-2876  JMJ-034451  After proposing SC2 N 4849, horizontal extension,	21D14	Accepted
008		charts and lists of character names, 24.2 Source references file for CJK Ideographs (CJKSrc.txt)			JNB's experts carefully rechecked it and simultaneously asked for public comment on it.  As a result, JNB concluded that the proposed code point 8FF3 for JMJ-058848 is not appropriate, and it is necessary to be changed.  8FF3  © 162.5 G0-6549	newly included in extension I on CDAM2.3.	
JP3- 009		35.5 Code charts and lists of character names, 24.2 Source references file for CJK Ideographs (CJKSrc.txt)		ge	The source references In Extension I of CDAM2.3, are reflected those of SC2 N 4874 which is a revised version of SC 2 N4851. But the source references in these documents for what are believed to be the same characters are different. To avoid future confusion, it is necessary to clarify the meaning of "source references" for those characters.  Examples: ISO/IEC JTC 1/SC 2 N 4874(and CDAM2.3):  Glyph Code Source Reference  Code Point GIDC23-004  CEBF4 GIDC23-006  GIDC23-006  GIDC23-006  GIDC23-007		Accepted in principle. (Extension I is a single repertoire from a single source (China), and the identifiers for their G-source prefixes are intended to be sequential within the repertoire. When the repertoire was adjusted, based on comments from China, the source references were re-sequenced. In other words, until a repertoire is frozen, the source references that are associated with its ideographs can be changed.)  Note however that that the source (distinct from GIDC)

<sup>1</sup> MB = Member body / NC = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by \*\*)
2 Type of comment: ge = general te = technical ed = editorial

MB/ NC¹	Line number	Clause/ Subclause	Paragraph/ Figure/Table	Type of comment <sup>2</sup>	Comments	Proposed change	Observations of the secretari
					ISO/IEC JTC 1/SC 2 N 4851:  2EBF0		and needs to be documented separately in clause 24. Description: 'GIDC23 GB18030-2022 Supplemental Ideographs', and the syntax should be GIDC23-ddd (the value ddd denoting 3 decimal units).
KR- 010				te	RE: 13 G glyphs containing dagger/seven component (上, U+5315; 七, U+4E03)  - KR requested that the component dagger ('Ł', U+5315) in 63 Korea-origin Hanja chars in UCS be changed to seven (七, U+4E03). (See IRG N2549KRResp1.pdf, WG2 N5186-Request to change glyphs of 63 K-source Hanja chars in UCS.pdf).  - A Korea-origin Hanja char is a Hanja char coined in Korea. Some Korea-origin Hanja chars are used for expressing Korean syllable (sound).  - Among the 63 chars shown in WG2 N5186, each of the following 13 Hanja chars has only two glyphs, K and G. U+517A is shown below as an example.  U+517A, 5391, 54DB, 551C, 551F, 55B8, 55ED, 591E, 5DFC, 5EE4, 65D5, 83BB, 848A.  517A  128  Note. There are five KP glyphs in the 13 chars. However, since KP glyphs were not submitted by DPRK NB, KP glyphs are not taken into consideration in this discussion.  - The source of all the chars except U+551C is GK,		Noted. It is not uncommon to have two components that are distinct when used as a whole ideograph to become part of unifiable ideographs when used a components within another ideographs. In other words, such occurrence may keep the two glyphs cognate. And it may be a source originator preference to use a form or another of the component.

<sup>1</sup> MB = Member body / NC = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by \*\*)
2 Type of comment: ge = general te = technical ed = editorial

MB/ NC¹	Line number	Clause/ Subclause	Paragraph/ Figure/Table	Type of comment <sup>2</sup>	Comments	Proposed change	Observations of the secretaria
					which is "GB12052-89 Korean Character Coded Character Set for Information Interchange". The set of these 13 chars is actually a subset of "92 Korean 'Idu' characters" as mentioned in the clause 27.1 of the first edition of UCS:2003.  - Considering that the source of these G chars (i.e., GK) and the 13 chars are 'Idu', it is almost certain that the G glyphs of these 13 chars are Korea-origin Hanja chars.  - Therefore, KR suggests that China NB consult with the Korean experts in China to see if it is better to modify the component dagger ('Ł', U+5315) in the 13 chars to seven (七, U+4E03).  - KR suggests that this issue be discussed by IRG experts at the next IRG meeting.		
KR- 011				te	RE: 31 JMJ glyphs containing dagger/seven component (Ł, U+5315; Ł, U+4E03)  - Japan requested a horizontal extension of 36422 JMJ chars to UCS (see WG2N5221-Request for 36422 Horizontal Extensions in the Jcolumn for JMJ sources.pdf). These chars came from MOJI-JOHO-KIBAN IDEOGRAPHS-2018.  - Among them, each of the 31 JMJ chars (listed below) in UCS currently has a Korea-origin glyph where the component dagger ('Ł', U+5315) was changed to seven (Ł, U+4E03).  CJK main (11 chars): U+517A, 54DB, 551C, 551F, 55B8, 55ED, 5DFC, 5EE4, 65D5, 83BB, 848A  CJK ExtA (20 chars): U+357E, 358B-358E, 3599-359D, 35AF-35B0, 35B2-35B3, 35DF-35E1, 35EF, 360F, 3612  - U+517A is shown below as an example.		Noted. Please refer to the disposition above. Using either component (dagger/seven) may be a matter of source originator preference and does not make the two ideographs non cognate. However, this has no immediate effect on the repertoire and can be updated later. Accordingly, the JMJ extensions can stay where there are for now, based on the parallel situation with the GK sources above.

<sup>1</sup> MB = Member body / NC = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by \*\*)
2 Type of comment: ge = general te = technical ed = editorial

MB/ NC¹	Line number	Clause/ Subclause	Paragraph/ Figure/Table	Type of comment <sup>2</sup>	Comments	Proposed change	Observations of the secretaria
					- 1) KR would like to see the actual usage of these 31 chars.  - 2) If these 31 JMJ chars are based on or related with Korea-origin glyphs, KR suggests that Japan NB investigate the possibility of changing the component dagger ('七', U+5315) in the 31 chars to seven (七, U+4E03).  - 3) If these 31 JMJ chars are NOT based on or related with Korea-origin glyphs, K glyphs and J glyphs seem non-cognate. Therefore, it might be better NOT to make a horizontal extension; instead, we could consider the possibility of disunifying K glyphs and J glyphs and assigning separate code positions to J glyphs.  - 4) KR will investigate this issue more carefully and, if necessary, will submit further comments.  - 5) KR suggests that this issue be discussed by IRG experts at the next IRG meeting.		
KR- 012					RE: Some JMJ glyphs containing components Z (bird, U+4E59), □ (IEUNG, U+11BC), or □ (KIYEOK, 11A8)  - Among the 36422 JMJ chars proposed for a horizontal extension in the document WG2N5221, each of some JMJ chars (7 chars are shown as examples below) in UCS currently has a Koreaorigin glyph which includes components Z (bird, U+4E59), □ (IEUNG, U+11BC), or □ (KIYEOK, 11A8).  Components Z (bird, U+4E59, reading = EUL), □ (IEUNG, U+11BC), or □ (KIYEOK, 11A8) in these seven Korea-origin glyphs represents the Korean syllable-final sound ■ (RIEUL, U+11AF), □ (IEUNG, U+11BC), or □ (KIYEOK, 11A8), respectively.		Noted. The Japan's Moji Jōhō Kiban database, which is the basis for the JMJ prefix for the kIRG_JSource property, includes a large number of ideographs for personal names, some of which may be for names of people who immigrated to Japan, or otherwise appear in Japan's personal name database. Showing evidence or usage may not be possible based on privacy consideration and is in all cases not a strong requirements for horizontal extensions.

MB = Member body / NC = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by \*\*)
Type of comment: ge = general te = technical ed = editorial

MB/ NC¹	Line number	Clause/ Subclause	Paragraph/ Figure/Table	Type of comment <sup>2</sup>	Comments	Proposed change	Observations of the secretaria
					3   14-3407   2   52   2   2   2   2   2   2   2		
PL- 013				te	Date in metadata/header in HieroglyphSources.txt file contains date # Date: 2023-8-4	Please make date ISO 8601 compliant, i.e. 2023-08-04 not 2023-8-4.	Accepted
PL- 014				te	metadata/header in HieroglyphSources.txt contains Date, but similar Names.txt file in Unicode contains version number:	Please add version of standard (something like ISO 10646:2023, or ISO 10646 ed 2).	Accepted in principle, next version will contain the

<sup>1</sup> MB = Member body / NC = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by \*\*)
2 Type of comment: ge = general te = technical ed = editorial

MB/ NC¹	Line number	Clause/ Subclause	Paragraph/ Figure/Table	Type of comment <sup>2</sup>	Comments	Proposed change	Observations of the secretariat
					@@@ The Unicode Standard 12.1.0 Version number is better for automatic parsing than date.		mention 10646:2024.
PL- 015				te	metadata lines in Unicode files are prefixed with "@@@", and in HieroglyphSources.txt with "#".	Use same prefix as in Unicode ("@@@")	Not accepted. In ISO 10646 these are comment lines, not formal metadata.
CA1- 016	bottom of pages 2, 4, 6, 8			ed	Copyright notice has incorrect year	Replace 2022 with 2023	Accepted
CA2- 017	page 4	A.1, last line		te	394 CJK UNIFIED IDEOGRAPHS-2024 Collections incomplete	Add 2EBF 0-2EE5D for CJK Unified Ideographs Extension I, per the charts shown on pages 294 to 301; also, should the range end with 2EE5F?	Accepted in principle, the new collection #394 already includes collection #2009 which is CJK Unified Ideographs Extension I 2EBF0-2EE5F
CZ- 018				te	Egyptian Hieroglyphs Extended-A sections contains several blocks with no characters in them: A03, C06, D16, D33, D38, E11, E12, F13, F29, G03, G07, H05, M25, M27, N07, O15, O17, P12, R10, R19, R21, S16, S18, S19, T17, U04, U10, U14, U20, V12, V13, V18, X01, X07, Z05, AA01-07, AA09-21, A23-1. It is unclear whether these are classifications in the sources whose characters are not encoded, or whether they were unified with others, whether the following characters come from any of them or whether these are just editorial errors. G17 is out of order between G07 and G08. AA10 and AA11 have names like "(Back of something?)" Consider whether these are appropriate inclusion, especially if no characters are encoded in these blocks.		Accepted in principle. As presented, the names list for Egyptian Hieroglyphs Extended-A was generated with a script that extracted the names from a spreadsheet which contains more signs, resulting sometimes in empty blocks. It is expected that all empty groups will be removed, and remaining groups will be renamed based on editorial input.
GB- 019	p256			te	U+25A38 K6-110D glyph is incorrect (it is the same as K6-110C glyph for U+25A31).		Accepted. Fixed with a temp patch, but requesting an update from ROK.

<sup>1</sup> MB = Member body / NC = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by \*\*)
2 Type of comment: ge = general te = technical ed = editorial