

Title: ISO/IEC 10646 repertoire proposal for CDAM2.3

Source: Michel Suignard, Project editor

Status: Individual Contribution

Distribution: WG2

Reference: N5234, N5235R

Executive Summary:

This document describes in some details the content of CDAM2.3 in comparing the repertoire that was adopted during WG2 #70, and the further refinements done to the repertoire and CJK Horizontal Extensions. The result of these considerations can be seen in the updated version WG2 N5235R at <https://www.unicode.org/wg2/docs/n5235R-cdam2-3-drafchart.pdf>.

Amendment 2.3 repertoire

The document WG2 N5234 (WG2 #70 recommendations, dated 6-22-2023) made a set of recommendations defining the content for CDAM2.3 but left some latitude to refine the content based on additional feedback and further progress on some of the new repertoire. This document presents the updated repertoire and other content based on category.

CJK Unified Ideographs Extension I

The original proposal was done in <https://www.unicode.org/wg2/docs/n5214-23106-unc-extension-i.pdf>, slightly revised in <https://www.unicode.org/wg2/docs/n5214R-23114-cjk-unc-extension.pdf>, still with 603 proposed ideographs which was presented at the WG2 #70 meeting. After further refinement, the repertoire was modified by China to include 622 ideographs (block extended from U+2EE4F to U+2EE5F and the proposed range ending in U+2EE5D instead of U+2EE4A). The latest version of the proposal is located at <https://www.unicode.org/wg2/docs/n5214R2-23114-cjk-unc-extension.pdf> and endorsed by China. The new repertoire was also reviewed by the Unicode CJK & Unihan Group in <https://www.unicode.org/L2/L2023/23163-cjk-unihan-group-utc176.pdf> and endorsed by the Unicode Technical Committee during UTC#176 (July 25-27 2023), draft minutes at <https://www.unicode.org/L2/L2023/23157.htm>.

Tangut glyph changes (N5206 and N5237)

Document N5206 proposed 8 glyph changes, of which 7 were endorsed in recommendation M70.05. However, the additional 4 glyphs changes proposed in N5237 will not be implemented at this stage, based on additional feedback received by Chinese experts.

Name change for a character in Symbols for legacy computers supplement (U+1CE07)

Based on feedback received during review by Unicode, it was found that a name in that block could be improved because the proposed glyph for U+1CE07 is actually located in the top left of the character bounding box; therefore it is preferable to change the name of U+1CE07 TOP RIGHT BLACK LEFT-POINTING SMALL TRIANGLE to TOP LEFT BLACK LEFT-POINTING SMALL TRIANGLE.

IRG J source, new JMJ horizontal extension

Document N5221 describes 36422 new J source horizontal extensions, and these were accepted in WG2 recommendation M70.04. When implementing these changes, the editor was presented with a font to display these new 36,422 glyphs based on a previous JMJ collection, and the submission was accompanied with 44 changes from the original JMJ collection #391 described in ISO/IEC 10646 Annex A.5.11 “MOJI-JOHO-KIBAN IDEOGRAPHS-2018). These 44 changes are now documented in https://www.unicode.org/wg2/docs/n5221-JNB_contribution_2304.pdf and are reflected in the table in <https://www.unicode.org/wg2/docs/n5221-Proposed%20Horizontal%20Extension.pdf> (683MB). The 44 changes indicate differences between the original collection #391 and the new horizontal extensions. These may be simply a code point change of original JMJ sources, a disunification (in that a variant acquires its own code point for horizontal extension), or a move to just a variant (in such cases the JMJ source is a variant of another J source sharing the same code point). However, one code point change, requesting for MJ068081 to be moved from U+24138 to U+2413C presented a challenge:

44	MJ068081	燙	U+24138	24138 水 8519 UCS2003	燙 燙 TS-7042	U+2413C	2413C 水 8519 UCS2003	燙 燙 GX-098128
----	----------	---	---------	----------------------------	----------------	---------	----------------------------	------------------

The issue was that JMJ-068081 is the kIRG_JSource property value for U+24138 燙, and moving it to U+2413C 燙 resulted in the following IVS lacking a base character:

<24138,E0100> 燙 MJ039803

Therefore, the following additional mapping should be required, which will increase the horizontal extension to cover 36,423 ideographs:

U+24138 kIRG_JSource JMJ-039803

In other words, MJ068081 is moved from U+24138 to U+2413C, but JMJ-039803 should be allocated to the newly freed U+24138. This was reviewed and endorsed by the Japan NB.

Egyptian Hieroglyphs and related arrows

Recommendation M70.6 accepted in principle a repertoire based on N5215 (Encoding proposal for an extended Egyptian Hieroglyphs Repertoire at <https://www.unicode.org/wg2/docs/n5215-hieroglyphs.pdf>). Since then, the proposal was finalized as N5240 (<https://www.unicode.org/wg2/docs/n5240-hieroglyphs.pdf>) with 3994 characters proposed in a new block: Egyptian Hieroglyphs Extended-A (U+13460..U+143FF). It is related to a companion document requesting 2 arrows as N5239 at <https://www.unicode.org/wg2/docs/n5239-hieroglyphic-arrows.pdf>). These two documents have been reviewed by Unicode experts at UTC#176 and proposed for a future version of Unicode. Note that annotation in the the names list are preliminary and will be updated in the future.

Consideration on CJK Horizontal Extensions

The document N5235R reflects the repertoire addition as planned for CDAM2.3, but not glyph changes and the CJK Horizontal Extensions. The plan for CDAM2.3 is to show the Hanja extensions (K Sources) because there are only 134 new sources but adding 36,423 new glyphs to show the new J sources would make the amendment almost as big as a new edition. At this point, the plan is to show the full repertoire content as a separate online resource. In addition, during the development of this amendment it became possible to display all KP sources. While these are not new CJK Horizontal extensions, most of the glyphs were not available until this amendment.

-end of document-