# Title: WG2 Consent Docket

Source: Peter Constable, Ken Whistler
Date: October 29, 2010
Action: For review by UTC

The WG2 Meeting #57 was held October 4 – 8 in Busan, South Korea. During that meeting a number of decisions were taken approving new repertoire for 10646, in addition to various other changes that were the result of disposition of comments for the CD of 10646 3rd edition.

This document summarizes changes in repertoire, character names, code points, or other changes which require some UTC action to bring the accepted repertoire back into synch. The complete set of actions taken at Busan can be found in the WG2 Meeting 57 resolutions, L2/10-390 (= WG2 N3904)

## 1 Changes related to 10646 3rd edition

Note: Details regarding code positions, glyphs, names, etc. related to the following items can be seen in the draft repertoire for the FCD, L2/10-387 (= WG2 N3945) and the final disposition of comments on the CD ballot, L2/10-385 (= WG2 N3936).

### 1.a CJK source references—"Idu" and "Hong Kong" characters

In Korea's ballot comment T4, they called out apparent errors in G source references for 92 Korean "Idu" characters and 58 "Hong Kong" characters: G1 (= GB 12345-90) is cited, but it appears not to include these characters.

In discussion, it turned out that, of the combined 150 characters, 91 are "Idu" characters, 58 are characters used in Hong Kong, and 1 is a Hakka character. The 58 Hong Kong characters have H source references in addition to G1 references.

It was found that the 58 Hong Kong characters and the 1 Hakka character are included in GB 15564-1995. This will be introduced as a new source, "GH", and the source references for these 59 characters will be revised to use GH instead of G1. The revised source-reference records for these 59 characters can be found on pages 13 – 14 of L2/10-385.

The 91 "Idu" characters were found to be included in GB 12052-89. This will be introduced as a new source, "GK", and the source references for these characters were will be revised to use GK instead of G1. The revised source-reference records for these 91 characters can be found on pages 15 – 18 of L2/10-385.

As a result of these changes, source-separation rules in Annex S will also be revised to exclude the two new sources from unification requirements in the case of these characters.

**Recommendation:** accept the new sources and the revisions to the source references for these 150 characters, as described in L2/10-385, in a future version.

## 1.b   CJK source references—characters without sources

In Korea's ballot comment T5, inconsistencies were noted between drafts for the 2<sup>nd</sup> and 3<sup>rd</sup> editions in the use of "KP1-0000" as a pseudo-source reference in the case of characters that have no actual known source. After discussion, it was decided that pseudo-source references will be included in all cases of ideographs with no reference, and that the following two IDs will be used:

- For unified ideographs: UI-0000
- For compatibility ideographs: CI-0000

See L2/10-385 for further details.

**Recommendation:** adopt the same conventions for documenting ideographs with no source reference.

## 1.c   Additions for Coptic

In L2/10-290 (= WG2 N3873), Michael Everson proposed various characters for use in relation to Coptic. UTC approved two of these characters (cf. UTC 124-C11):

- 2CF2 COPTIC CAPITAL LETTER BOHAIRIC KHEI
- 2CF3 COPTIC SMALL LETTER BOHAIRIC KHEI

There were other punctuation characters proposed some of which UTC considered acceptable and others that were questionable. In WG2 discussion, Ireland was agreeable to encoding only those punctuation characters that were not contentious. Thus, WG2 accepted the two letters listed above and also the following two punctuation characters, not yet approved by UTC:

- 2E33 RAISED DOT
- 2E34 RAISED COMMA

See L2/10-290 and L2/10-387 for details.

**Recommendation:** approve the two punctuation characters for addition to a future version.

## 1.d   Lao characters added

UK introduced a proposal (L2/10-335 = N3893) to encode two Lao characters used for the Khmu language:

- 0EDE LAO LETTER KHMU GO
- 0EDF LAO LETTER KHMU NYO

These are consonants that behave like other Lao consonants. Glyphs and character properties were shown in the proposal doc, and a font is available. An updated proposal that includes explicit details about collation is expected before UTC takes up this consent docket.

**Recommendation:** UTC should discuss and decide what to do.

### 1.e   Latin characters added

In L2/10-336 (=WG2 N3896), Lorna Priest, Laurentiu Iancu and Michael Everson propose two Latin characters; this proposal has not yet been reviewed by UTC, but has been approved by WG2 with revised code positions. The following two characters were approved by WG2:

- A792 LATIN CAPITAL LETTER C WITH BAR
- A793 LATIN SMALL LETTER C WITH BAR

Character properties are provided in L2/10-336.

**Recommendation:** approve the two Latin characters at code positions as revised by WG2 for addition to a future version.

### 1.f   Named character sequence identifiers for ISO/IEC 9995-7

ISO/IEC JTC 1/SC 35 submitted a proposal (L2/10-351 = WG2 N3897) to make various changes in glyphs and character annotations for existing UCS characters for keyboard-related symbols used in SC 35 standards, and also to add new characters and named character sequences for additional keyboard-related symbols. In discussion, it was decided to adopt changes involving only already-existing characters and to leave the other requests for national body review. Accordingly, the SC 35 liaison submitted a second document during the meeting (L2/10-391 = WG2 N3943) that specifies the changes to be adopted at the Busan meeting.

Thus, WG2 approved the following named UCS sequence identifiers for addition in the 3$^{rd}$ edition:

- <21F3, 20E2> KEYBOARD SYMBOL SCROLLING
- <2139, 20E2> KEYBOARD SYMBOL HELP

**Recommendation:** approve the two named character sequences for addition in a future version.

## 2   Changes related to PDAM1 of 10646 3rd edition

Note: Details regarding code positions, glyphs, names, etc. related to the following items can be seen in the draft repertoire for PDAM1, L2/10-386 (= WG2 N3946)

### 2.a   Armenian eternity signs

In Armenian ballot comments on the CD for the 3$^{rd}$ edition, Armenia requested the addition of a new character for the Armenian eternity sign. In L2/10-354 (= WG2 N3924), Michael Everson noted that there are two orientations for the eternity sign (comparable to alternate orientations for the svasti sign). Accordingly, Ireland proposed two characters.

In Everson's proposal, code positions in the Armenian block were proposed; Armenia had proposed a code position in the Dingbats block. In discussion, it was decided that code positions suitable for dingbats / symbols were appropriate. Thus, the following two characters were approved:

- 1F53E RIGHT-FACING ARMENIAN ETERNITY SIGN

- 1F53F LEFT-FACING ARMENIAN ETERNITY SIGN

Character properties are provided in L2/10-354; glyphs and other details can be seen in L2/10-386.

**Recommendation:** Consider the evidence for a second symbol and decide what to do.

## 2.b   Coptic numerals

At UTC #123, a proposal to encode Coptic numbers (L2/10-114) was reviewed and accepted (cf. UTC 123-C34). WG2 reviewed a subsequent version of the proposal, L2/10-206 (= WG2 N3843) and approved the characters for addition to 10646, though with a change in name and code position for one character:

- 102FC COPTIC NUMBER MARK

was changed to

- 0605 ARABIC COPTIC NUMBER MARK

**Recommendation:** UTC should discuss and decide what to do.

## 2.c   Sinhala numerals

At UTC #110 (February 2007), a proposal from Michael Everson (L2/07-002) to encode Sinhala archaic numbers was reviewed, and a decision was taken to encode 20 characters at 0DE7..0DEF and 0DF5..0DFF (cf. UTC 100-C35). Subsequently, a contribution (L2/08-007) was received from ICTA requesting that encoding of Sinhala numerals be put on hold pending further research. Accordingly, UTC decided to put these on hold (cf. UTC 114-C26) and the characters were removed from Amendment 5 of 10646.

This year, at UTC #124, new proposals from Michael Everson (L2/10-301) and from the Government of Sri Lanka (L2/10-312) were reviewed. Certain name changes were discussed, and it was agreed that the Government of Sri Lanka would submit a revised proposal to WG2.

In Busan, WG2 took up the revised proposal (L2/10-337 = WG2 N3888) and took action to approve characters as follows:

- 10 Sinhala numerals in code positions 0DE6 to 0DEF in the Sinhala block
- 20 Sinhala archaic numerals in code positions 111E1 to 111F4, in a new block named Sinhala Archaic Numbers located at 111E0..111FF

Details can be found in L2/10-337 and in L2/10-386. Certain name changes discussed at UTC #124; in particular, it was agreed that names of digits in the BMP be changed from SINHALA LITH DIGIT etc. to SINHALA ASTROLOGICAL DIGIT etc. Between UTC #124 and WG2 57, this was discussed within the Sri Lanka national body, and they decided against this change since it was felt that the revised names would be confusing to the target user community.

The proposal documents provided thus far from ICTA do not include details on character properties. Properties are included in L2/10-301; these should be reviewed.

**Recommendation:** Rescind the previously-approved 20 characters and approve the new block and the 30 characters, as described, for addition in a future version; review the character properties and determine appropriateness.

## 2.d   Myanmar additions for Shan Pali

In L2/10-345 (= WG2 N3906), Michael Everson proposes the addition of seven additional Myanmar characters used for Shan Pali. The characters include five consonant letters, a letter modifier and a combining mark. A revised version of L2/10-345 is expected before UTC takes up this consent docket that makes explicit the collation details and that provides clarification on interaction of the mark with other marks.

WG2 approved the seven Myanmar characters for addition in a new block, Myanmar Extended-B, located at A9E0..A9FF. The glyphs and other details for the seven characters can be seen in L2/10-345 and in L2/10-386.

**Recommendation:** Review and decide what to do.

## 2.e   Bassa Vah script

At UTC #122 and #123, preliminary SEI proposals to encode the Bassa Vah script were discussed by UTC. After the last meeting, the main outstanding UTC concern was whether there was adequate evidence of casing. Since then, SEI was able to determine that the script does not have case. During the Busan meeting, SEI submitted a revised proposal (L2/10-382 =N3941) reflecting the final repertoire.

WG2 approved the script for encoding in a new block, Bassa Vah, at 16AD0..16AFF. Details can be seen in L2/10-382 and in L2/10-386.

**Recommendation:** approve the Bassa Vah script, with block and characters as described, for addition to a future version.

## 2.f   Mro script

A proposal to encoded Mro script was submitted to UTC in 2009 (L2/09-097 = WG2 N3589), though it has not yet been discussed by UTC. This proposal was taken up in Busan, and WG2 approved the script for encoding in a new block, Mro, at 16A40..16A6F. Details can be seen in L2/09-097 and in L2/10-386.

**Recommendation:** review the Mro script proposal and decide what to do.

## 2.g   Sindhi script

Sindhi script has been approved by UTC in a Sindhi block at 11640..1164F (cf. UTC 124-C34). WG2 approved these characters, but in a different character range, 112B0..112FF. See L2/10-386 for details.

**Recommendation:** approve the revised range for the Sindhi block and revised code positions for Sindhi characters.

## 3   WG2 decisions that bring 10646 into sync with Unicode (informative)

WG2 took various actions to add characters, change names, etc. that bring 10646 into sync with Unicode. These are summarized here; no UTC action is required.

Changes related to the FCD for 10646 3rd edition (see L2/10-386 = WG2 N3946 for the draft repertoire of the FCD, and L2/10-385 = WG2 N3936 for the final disposition of comments on the CD ballot):

- 20B9 INDIAN RUPEE SIGN added (cf. UTC 124-C8)
- Named UCS sequence identifiers added for Sinhala per US ballot comment T1
- Miao block extended, Miao characters re-ordered, names changed and 5 new characters added per US ballot comments T2a – T2f.
- Names changes to 1F540..1F542, POMMY to POMMEE, per US ballot comment T3
- Addition of the formal name alias "WEIERSTRASS ELLIPTIC FUNCTION" to 2118 SCRIPT CAPITAL P, per US ballot comment T4.
- A7AA LATIN CAPITAL LETTER H WITH HOOK added (cf. UTC 123-C17)
- 4 Vedic characters (1CF3..1CF6) added (cf. UTC 123-C30, 123-C31, 124-C29):
- 2 IPA characters (A7F8..A7F9) added (cf. UTC 123-C21)
- 2 letter-like symbols for Canadian legal use (1F16A..1F16B) added (cf. UTC 123-M1)
- 2 Tifinagh characters (2D66..2D67) added (cf. UTC 123-C18)
- 35 Arabic characters for African and Asian languages added (cf. UTC 124-C21)
- 1 CJK ideograph for URO (9FCC) added (cf. UTC 124-C14)

Additions to PDAM1 (see L2/10-386 = WG2 N3946 for the draft repertoire of PDAM1):

- Palmyrene script added (cf. UTC 124-C10)
- Old North Arabian script added (cf. UTC 123-C26)

## 4   Outstanding points of non-synchronization (informative)

There are outstanding points of non-synchronization between Unicode and 10646—characters / scripts added to one but not the other. These are described here for informative purpose; there are no recommendations for UTC action, though UTC may wish to discuss these items and consider possible actions.

### 4.a   Duployan

UTC has approved Duployan characters (cf. UTC 124-C31, 124-C32, 124-C33). Because the proposal was brought before WG2 for the first time and there were a point of concern raised, WG2 preferred to wait for national bodies to review first. The concern had to do with the layout of characters in the Duployan block: Michael Everson is not convinced that the current organization in sub-groups is useful and prefers the binary order to match the collation order.

## 4.b   Linear A

Linear A was proposed by SEI and approved by UTC (cf. UTC 123-C25, 124-C7). Just prior to the WG2 meeting, SEI received new information indicating that their proposal was not yet stable and mature, and so they requested that WG2 take no action.

## 4.c   Batak symbols

Two Batak symbols were approved by WG2 but have not been approved by UTC:

- 1BFA BATAK SYMBOL BINDU GODANG
- 1BFB BATAK SYMBOL BINDU PINARJOLMA

These characters had initially been in amendments to the 1st edition of 10646, and they were pulled out and moved into the CD for the 3rd edition to accommodate US ballot comments. On the CD ballot for the 3rd edition, there were no US comments on these characters; this was an oversight rather than intentional. Hence, these characters will be included in the FCD for the 3rd edition (cf. L2/10-387).

## 4.d   Latin middle / glottal dot

The following character had been approved by WG2 for addition to 10646 but has not been approved by UTC:

- A78F LATIN LETTER GLOTTAL DOT

Note: the name for A78F as initially accepted by WG2 was LATIN LETTER MIDDLE DOT. At WG2 57, the name was changed in response to ballot comments from Ireland.

This character had initially been in amendments to the 1st edition of 10646, and it was pulled out and moved into the CD for the 3rd edition to accommodate US ballot comments. On the CD ballot for the 3rd edition, there were no US comments on this character; this was an oversight rather than intentional. Hence, it will be included in the FCD for the 3rd edition (cf. L2/10-387).

## 4.e   Additional emoticons

In the development of Amendment 8, the Irish national body proposed 13 emoticon symbols in addition to those being added to support emoji used in Japanese mobile devices. These were approved by WG2, but at WG2 #56 they were moved from Amendment 8 into the CD for the 3rd edition. These were cited in the consent docket following WG2 #56 (L2/10-149, section L), but minutes of the subsequent UTC meeting, UTC #123 (L2/10-108) indicate that no UTC action was taken regarding these characters when the consent docket from WG2 #56 was taken up. These also were not added to the agenda for UTC #124. Hence, these 13 characters remain as an outstanding point of disunification.

The 13 characters approved by WG2 are located in the Emoticons block, 1F600..1F64F. They are listed, with their names and glyphs, on pages 48 – 49 of L2/10-142 (= WG2 N3834).