Title: WG2 Consent Docket
Author: Debbie Anderson, Peter Constable, and Ken Whistler
Date: June 28, 2018
Action: For consideration by UTC

WG2 #67 (and SC2#23) met in London from June 18 to 22, 2018. During that meeting a number of recommendations were agreed to for the DAM2 of the 5th Edition. See WG2 N4954 (=L2/18-209) for the full details of all the recommendations.

As usual, in this consent docket, we summarize just the parts of the actions taken by WG2 which result in a different status between WG2 and the UTC regarding various character approvals. These are the differences where the UTC needs to make some decision regarding how to synchronize approvals (or to oppose a proposed change).

Note that the pipeline page: http://www.unicode.org/alloc/Pipeline.html has already been updated to reflect changes in approvals by WG2, and to highlight differences from the current approvals by the UTC, so that page can be useful in following the discussion below on the individual issues.

I. FDAM1 to the 5th edition

Amendment 1 to 5th edition will proceed to a FDAM ballot, with tentative starting date August 2018. There were a few technical changes to DAM1, but the UTC has already approved them. No action is required on UTC's part. The FDAM1 repertoire is posted as WG2 N5005.

II. DAM2 to the 5th Edition

Because the comments on PDAM 2.3 were resolved, Amendment 2 to the 5th Edition will proceed to an enquiry draft (DAM2). A chart showing the final repertoire for DAM 2 can be seen in WG2 N5000 = L2/18-212. The tentative start-date is August 2018. Note that UTC has planned for the repertoire of Unicode 12.0 to be aligned with Amendment 2, plus other emoji additions.

A. Cornish Punctuation (Supplemental Punctuation)

A ballot comment from Ireland recommended removing "MEDIEVAL" from U+2E4F MEDIEVAL CORNISH VERSE SEPARATOR. Since the character appears in manuscripts that don't fully align with the usual dates of Medieval Europe, WG2 recommended a name change (accepted by SC2) that removed "MEDIEVAL" and changed "SEPARATOR" to "DIVIDER" in the name. The name for U+2E4F is thus: CORNISH VERSE DIVIDER.

Recommendation: Approve the character (with the name changes).

B. Egyptian Hieroglyph Format Controls

Feedback from Andrew Glass (N5001), one of the authors for the Egyptian Hieroglyph control characters proposal, noted an error in the names of four of the format characters in the PDAM. He explained that "START" in the character names is not a verb, but a noun that identifies the leading edge, which may be right or left, depending upon the script's direction. The names should hence be modified to be aligned with their function, with the action being "insert." After discussion at WG2, the names proposed by Glass were agreed to, and "at" was inserted for added clarification (i.e., INSERT AT TOP START).

The following shows the current names and the names modified by WG2 (and accepted by SC2).

PDAM 2.3 names:	New names (accepted by SC2):
13432 EGYPTIAN HIEROGLYPH START AT TOP	EGYPTIAN HIEROGLYPH INSERT AT TOP START
13433 EGYPTIAN HIEROGLYPH START AT BOTTOM	EGYPTIAN HIEROGLYPH INSERT AT BOTTOM START
13434 EGYPTIAN HIEROGLYPH END AT TOP	EGYPTIAN HIEROGLYPH INSERT AT TOP END
13435 EGYPTIAN HIEROGLYPH END AT BOTTOM	EGYPTIAN HIEROGLYPH INSERT AT BOTTOM END

Recommendation: Accept the four name changes.

C. Nandinagari

The USNB submitted ballot comments on PDAM 2.2 for Nandinagari to be moved from U+11B90.. U+11BEF to U+119A0..U+119FF, and this move was incorporated in PDAM 2.3. However, the UTC never went on record as approving the change.

Recommendation: Accept the code point changes for Nandinagari.

D. Wancho

The UTC has already approved 55 Wancho characters, but four Wancho tone marks were not approved because evidence was lacking. A document demonstrating their use in printed text was provided for WG2 (N4976), and the four were recommended by WG2 (and accepted by SC2) as additions to the script. The four are:

> 1E2EC WANCHO TONE TUP
> 1E2ED WANCHO TONE TUP MANG
> 1E2EE WANCHO TONE OKOI
> 1E2EF WANCHO TONE OKOI MANG

Recommendation: Accept the four Wancho tone marks.

E. Egyptological yod

The UTC had accepted the two characters U+A7BC LATIN CAPITAL LETTER EGYPTOLOGICAL YOD and U+A7BD LATIN SMALL LETTER EGYPTOLOGICAL YOD but the names were changed in PDAM 2.3 to U+A7BC LATIN CAPITAL LETTER GLOTTAL I and U+A7BD LATIN SMALL LETTER GLOTTAL I, so they would be consistent with other transliteration letters from U+A7BA… U+A7BF.

Recommendation: Accept the two name changes.

III. CD for the 6th Edition

SC2 will prepare a new project proposal for the Committee Draft (CD) of the 6th Edition. The following items were recommended by WG2 and accepted by SC2. The latest charts showing the content are found in N5006. Some content may be included in Unicode 13.0, which is scheduled for release in March 2020.

F. Bopomofo

In N4980, TCA proposed one letter, U+31BB BOPOMOFO FINAL LETTER G, be added to Bopomofo Extended block. (The character U+31B6 BOPOMOFO FINAL LETTER K was a mistake and should have been the proposed character.) WG2 recommended the addition of U+31BB. The addition seems uncontroversial.

Recommendation: Accept the Bopomofo character.

G. Extension G

WG2 recommended 4,938 characters be encoded in a new block, CJK Unified Ideographs Extension G, extending from U+30000.. U+3134F. The glyphs and sources are documented in N4982R and are based on N5007.

Recommendation: UTC should discuss.

H. Gongche characters

WG2 recommended encoding 7 ideographs U+2A6E0.. U+2A6E6 used in musical notation, based on document N4967. The discussion at WG2 focused on appropriate process for these ideographs and on their location. There was general consensus that these should be treated as unified CJK ideographs, but that they are exceptional: they should not be processed as part of Extension G or a future extension, and they are not UNC (urgently needed characters) but are valid candidates for encoding. There was general consensus to process them in SC2 as an independent, non-UNC proposal. Also, the basic elements of these ideographs are commonly-used characters, and there was concern about potential confusion for users if the characters were offered in input methods.

After considering adding them at the end of the Extension B block, WG2 recommended (and accepted by SC2) the creation of a new block at U+2A6E0.. U+2A6FF named CJK Unified Ideographs Supplement, which could be used for exceptional cases, and to encode these characters within that block.

Recommendation: UTC should discuss.

I. Khitan Small Script

The encoding model was the main concern of USNB ballot comments for PDAM 2.2, which prompted the removal of Khitan Small Script in PDAM 2.3. Alternate models were summarized in the WG2 discussion, including the options mentioned by Unicode script experts in N4977 (L2/18-210) along with an additional option, which is mentioned in the Khitan Small Script ad hoc report, N5002 (L2/18-213).

During an ad hoc, representatives of the user community, including experts from China, were favorable toward the model proposed in N4943 (L2/18-121), which requires use of a single format control character. It was agreed that the behavior of this format control should be completely documented in Unicode, including occurrence in unexpected contexts. Also, the importance of assigning properties defined in UAX #50 was noted. See N5002 (L2/18-213) for more details.

The outcomes of discussion at WG2 were:

- include Khitan Small Script in the CD
- accept one script-specific format character, U+16FE4 KHITAN SMALL SCRIPT FILLER, in the Ideographic Symbols and Punctuation block
- change the name of U+18B00 KHITAN SMALL SCRIPT ITERATION MARK to be KHITAN SMALL SCRIPT CHARACTER-18B00.

Recommendation: UTC should discuss.

J. Latin Letter Thorn with Diagonal Stroke

The USNB ballot comments for PDAM 2.2 recommended removal of the two Latin thorns with diagonal stroke, since they were considered glyph variants of U+A764 LATIN CAPITAL LETTER THORN WITH STROKE

and U+A765 LATIN SMALL LETTER THORN WITH STROKE that could be represented with OpenType features or different fonts. The revised proposal, N4836R, includes early attestations of the character with a diagonal stroke (found in Anglo-Saxon works), and indicated that this form appeared to be the predominant shape, compared to horizontal stroke used in Nordic traditions

During discussion at WG2, the proposal authors acknowledged that thorns with diagonal versus horizontal stroke can be considered glyph variants, but contended that disunification was nonetheless warranted. Regarding the alternative — maintaining the current unified encoding — the question of what the representative glyph should be was also discussed: the horizontal stroke has been shown in charts since Unicode 5.1 (2008), and implemented in some commonly-used fonts. The diagonal stroke is more widely used, but a change of representative glyph could lead to changes in fonts that might impact existing documents and, hence, not be welcomed by Nordic scholars.

The following were recommended by WG2 (and accepted by SC2):

> A7C0 LATIN CAPITAL LETTER THORN WITH DIAGONAL STROKE
> A7C1 LATIN SMALL LETTER THORN WITH DIAGONAL STROKE

Recommendation: UTC should discuss.

K. Latin Small Letters with Overcurl

The original proposal N4907 (L2/17-358) by Michael Everson requested a combining diacritic for the overcurl in Medieval Cornish. WG2 discussion focused on two points: (i) whether base letters with this overcurl should be deemed distinct text elements, or merely presentation forms of existing base-diacritic combinations; and (ii) if distinct encoding were to be approved, whether a productive combining-mark or specific, atomic characters would be preferable. WG2 recommended (and SC2 accepted) the addition of 10 atomic characters to the CD, in the Latin Extended-D block.

The following were recommended by WG2 (and accepted by SC2):

> A7D0 LATIN SMALL LETTER A WITH OVERCURL
> A7D1 LATIN SMALL LETTER E WITH OVERCURL
> A7D2 LATIN SMALL LETTER I WITH OVERCURL
> A7D3 LATIN SMALL LETTER M WITH OVERCURL
> A7D4 LATIN SMALL LETTER N WITH OVERCURL
> A7D5 LATIN SMALL LETTER R WITH OVERCURL
> A7D6 LATIN SMALL LETTER S WITH OVERCURL
> A7D7 LATIN SMALL LETTER T WITH OVERCURL
> A7D8 LATIN SMALL LETTER U WITH OVERCURL
> A7D9 LATIN SMALL LETTER Y WITH OVERCURL

Recommendation: UTC should discuss.

L. Tangut components

Andrew West and colleagues discovered seven additional Tangut components and proposed these additions in N4957 (L2/18-194). The Tangut Components block has adequate space for them, and they do not appear to be controversial.

The seven components recommended by WG2 (and accepted by SC2) are:

18AF3 TANGUT COMPONENT-756
18AF4 TANGUT COMPONENT-757
18AF5 TANGUT COMPONENT-758
18AF6 TANGUT COMPONENT-759
18AF7 TANGUT COMPONENT-760
18AF8 TANGUT COMPONENT-761
18AF9 TANGUT COMPONENT-762

Recommendation: Accept the seven Tangut components.

M. Zanabazar Square additions

Two Zanabazar Square characters were proposed by Andrew West (N4945 = L2/18-132) and were discussed by the script experts in the N4977 (L2/18-210). Andrew West agreed to revise the proposal, adding guidance on how to handle cases that fall outside the "typical" examples (such as those shown in figures 3, 5, 7, and 11). He also agreed to fix the representative glyphs (changing the dotted circles to dotted squares). (Note: The revised update, incorporating new text and the glyph fixes, is now posted as N4945R [L2/18-132R].)

The following two characters were recommended by WG2 (and accepted by SC2):

11A48 ZANABAZAR SQUARE CLUSTER-INITIAL LETTER LA
11A49 ZANABAZAR SQUARE CLUSTER-INITIAL LETTER SA

Recommendation: UTC should discuss.

N. Name changes in Geometric Shapes Extended

Twelve characters in the CD for the 6th edition (N5006) have "LARGE" prepended to their names, in order to make them similar to U+2B1B BLACK LARGE SQUARE (etc.) in the Miscellaneous Symbols and Arrows block. These include:

1F7E0 LARGE ORANGE CIRCLE
1F7E1 LARGE YELLOW CIRCLE
1F7E2 LARGE GREEN CIRCLE
1F7E3 LARGE PURPLE CIRCLE
1F7E4 LARGE BROWN CIRCLE
1F7E5 LARGE RED SQUARE
1F7E6 LARGE BLUE SQUARE
1F7E7 LARGE ORANGE SQUARE
1F7E8 LARGE YELLOW SQUARE
1F7E9 LARGE GREEN SQUARE
1F7EA LARGE PURPLE SQUARE
1F7EB LARGE BROWN SQUARE

Recommendation: UTC should discuss.

O. Three pictograph additions to Supplemental Symbols and Pictographs|

The CD (N5006) includes the following three characters:

1F979 TROLL
1F9A3 MAMMOTH
1F9A4 DODO

The three characters were in PDAM 2.2, and were removed in PDAM 2.3 based on a USNB comment, but have been added back into the CD (N5006).

Recommendation: UTC should discuss.

P. Cypro-Minoan

This script has been added to the CD (N5006), though no new additional information has been received.

Recommendation: UTC should discuss.