

To: UTC

From: Deborah Anderson, IR and Head of Delegation

Re: Brief Report from WG2 Meeting #61, Vilnius, Lithuania

Date: 28 June 2013

WG2 meeting #61 took place from 10 – 14 June 2013 in Vilnius, Lithuania, very capably hosted by the Lithuanian Standards Board. Nine national bodies were present, as were 4 liaison members, with a total of 32 participants.

I. Ballots

Two ballots were discussed at the meeting:

1. Draft Amendment 2 (DAM2) to the 3rd edition of ISO/IEC 10646: 2012
2. Committee Draft (CD) for the 4th edition of ISO/IEC 10646

1. DAM2

For the DAM2, a number of changes were agreed to in the meeting, the most notable being:

- the addition of one new currency symbol, the Azerbaijani MANAT SIGN
- removal of the “Hungarian” script from DAM2 (moving instead into the 4th edition, see 2. below)
- adding an annotation to FF1F that indicates it as an “undescribed component” in Ideographic Description Sequences.

Further details are contained in the Consent Docket, L2/13-154.

The amendment was progressed to an FDAM ballot, a two-month ballot on which no technical comments are allowed. The Editor of 10646 noted that AMD 2 will not likely be published separately, but instead will be incorporated into the 4th edition that *will* be published.

Unicode 7.0 currently is aimed at being synchronized with the repertoire of Amendments 1 and 2 of the 3rd edition. (It was mentioned during the meeting that it is theoretically possible for Unicode 7.0 to include additional characters beyond those in Amendments 1 and 2, if the UTC approved such additions.)

2. CD

For the CD, a number of changes to fonts, formatting, and repertoire were discussed and approved. The following changes in the CD are worthy of mention:

- the CJK data files will be reformatted so that the CJK_U, CJK_C, and IICORE.txt files will be merged into a single file
- the STIX font will be used for standardized math variants glyphs, other fonts (if available) may be used to represent variants of other scripts, and for the standardized variants of CJK compatibility ideographs, a single nominal glyph of the corresponding CJK Unified Ideograph will be shown
- fonts based on a horizontal orientation will be used to generate the Mongolian and Phags-pa code charts

- the “Old Hungarian” script was added (from DAM2, with a name change), so it will get another round of review but will miss inclusion in Unicode 7.0
- a LATIN LETTER SINOLOGICAL DOT was approved, thereby bringing to an end a long-time disagreement about this middle dot (formerly called LATIN LETTER MIDDLE DOT, but now represented with an enlarged middle dot glyph)
- six Siddham variant characters requested by the Japanese NB were approved, as were 14 Siddham section marks (with name changes).

Further details are contained in the Consent Docket, L2/13-154.

The CD will progress to a Draft International Standard (DIS) ballot, a three-month ballot that is targeted to begin August 2013.

II. Characters and Scripts Not in Current Ballots

A number of characters were accepted for a future amendment to the 4th edition, including:

- OLD ITALIC LETTER TTE (which will bring to a head the question of whether Old Italic and North Italic should be considered one script or two, another long-standing disagreement)
- TELUGU LETTER RRRA
- characters for other scripts.

III. Other Topics

Other discussions took place during the WG2 meeting:

- A Nüshu ad hoc agreed to rely on the book *Nüshu Du Ben* as the source reference, though Michael Everson mentioned that the order of some of the characters as they appear in the book will need to be re-ordered. The Nüshu expert at the meeting, Zhao Liming, reported that rare characters that had been left out of the 2012 proposal from China (N4341), but agreed they could be added to the proposal. She mentioned that her font may be missing the glyphs, but agreed that Michael could make the additions to the font. Debbie Anderson and Andrew West will review any revised proposal, trying to arrive at a mature proposal over the summer so it could eventually be included in a PDAM. The ad hoc report is N4461.
- An ad hoc meeting met to discuss the cedilla/comma below issue that affects Latvian, Livonian, and Marshallese and was described in N4449. The meeting was attended by WG2 members and a representative from Latvia, Roberts Rozis. The result of the discussion, chronicled in the ad hoc report N4456, was to propose 4 characters for Marshallese that would *not* have decompositions. This solution was deemed less disruptive to current data than other options.
- A short discussion on Tangut took place. China reported on a new series of fonts being developed, with glyphs based on the original texts; a report on the font project is N4455. Work on these fonts may affect the Tangut repertoire in the proposal authored by Andrew West, et al. The work on the fonts is expected to be completed in late November

2013. In order to try to progress the Tangut proposal, SEI is organizing a meeting in Beijing in December 2013, with experts from China, the US, GB, and Ireland. The goal is to have an agreed-upon proposal ready for a PDAM by early 2014.

- IRG's Rapporteur, Dr. Lu, asked for the removal of 6 characters from Extension E, currently on the CD. She also reported that Extension F, now containing over 8K characters, is too big for review, and has been broken into two parts, "F1" and "F2". The review of "F1" may be completed in a year's time, and may be ready for balloting by the time of the WG2 meeting in Sri Lanka in fall 2014. The UNC process is still being refined; the current UNC request from the Unicode Consortium/US for 19 characters is still under review by IRG members.

IV. Future Meetings

The next WG2 meeting will be held Feb. 24-28, 2014 in San Jose, CA, and will be hosted by the Unicode Consortium. WG2 meeting #63 is slated to take place from 29 September – 3 October, 2014, in Sri Lanka, collocated with SC2. Meeting #64 is tentatively dated to late April or early June 2015 and is still looking for a host, with Europe as a possible location. The IRG meetings will take place from Nov. 18-22, 2013 in Japan, May 19-23, 2014 in Qingdao, China, and November 17-21, 2014 in U.S. (San Jose, CA).

V. OWG-SORT

OWG-SORT group met briefly on Tuesday, 11 June, to discuss Lithuanian issues with Virginijus Dadurketičius. One of the main problems described is searching on Lithuanian composed sequences. OWG-SORT members recommended Lithuania participate in CLDR, as their issues chiefly concern implementations.

The regular OWG-SORT meeting was held on Thursday, 13 June, where members agreed to make various corrections and updates to the CTT so it will be in synch with the changes in UCA. The 14651 PDAM2 ballot will be sent to the SC2 Secretariat by the end of October, so voting and results will be available by the next OWG-SORT meeting in February 2014. The repertoire of PDAM2 should be in synch with Amendment 2 of 10646. However, if Amendment 2 of 10646 does not get published, additions can still be made to 14651 Amendment 2.

The convener, Alain LaBonté, also presented the draft of an informative annex to 14651, entitled "Searching and Fuzzy Matches," which was discussed.