

TO: UTC
FROM: Deborah Anderson, IR
DATE: 30 April 2009
RE: WG2 Report, Meeting 54

Meeting 54 of WG2 met at Dublin City University, and was hosted by DCU's Centre for Next Generation Localisation. Eleven national body representatives were present, as well as 3 liaison organizations and 1 invited expert, totaling 37 attendees.

The weather was moderate, with one day being outright warm (for Dublin). In general, discussions at the meeting were constructive, with Amendment 6 and 7 getting resolved. A new Amendment 8 was started, to cover significant new repertoire additions (Emoji, alchemical symbols, CJK Extension D, etc.) which will require two rounds of technical review. This allowed the remaining content of Amendment 7 (after the removal of Tangut) to proceed directly to FPDAM, with the likelihood that it will not get caught up in the controversy and complicated ballot disposition of comments sure to attend the Emoji and other new content in Amendment 8. The meeting was notable for lengthy ad hocs on Emoji and Hungarian, as well as shorter meetings on Tangut, Nüshu, and Jurchen, and brief informal meetings on Batak and Old South Arabian.

The highlights of the meeting are listed below.

The outstanding issues relating to Amendment 6 were resolved, and Amendment 6 now progresses to FDAM. Various characters were modified or added (see Consent Docket, L2/09-177). The Canadian Syllabics under ballot in Amendment 7 were consolidated into the block under ballot in Amendment 6 and reordered (as had been requested by the USNB and other NBs).

For Amendment 7, a few characters were modified (or dropped) from the current ballot, and one character, KATAKANA ARCHAIC E, was added. The Arabic Pedagogical symbols had their names changed, and two Batak characters were removed. The historic letter HIRAGANA LETTER ARCHAIC YE was moved (with its new name) to a new block Historic Kana, alongside KATAKANA ARCHAIC E. These changes were consistent with USNB positions (though the USNB had no firm position on the re-naming of the Arabic Pedagogical symbols).

Tangut was removed from the Amendment 7 ballot. The Tangut ad hoc recommended the authors of the latest Tangut proposal (N3577 = L2/09-115, 116, 117) explicitly identify the new 311 additions to the repertoire, provide source information for these characters, describe the radical system in more detail, and describe the encoding principles. Fuller details are included in the Tangut ad hoc report N3629 (L2/09-169). This outcome was a compromise that differed from the USNB position going into the Dublin meeting.

Amendment 8 was begun, and will include CJK Unified Ideographs Extension D (formerly called “Urgently Needed Characters”), the Bamum supplement, alchemical symbols, Ethiopic characters, and miscellaneous other characters (including a LATIN LETTER MIDDLE DOT for transliteration of Phags-Pa); see the Consent Docket for details, L2/09-177.

A total of 748 Emoji symbols were also approved for Amendment 8, thanks to the deft handling of discussion by Peter Constable, ad hoc chair, who juggled comments from participants in Dublin with input over the phone from the proposal authors in the US. The net result of the discussion, documented in the Emoji ad hoc report L2/09-153 (N3636), was that all the Emoji characters requested by the US went to ballot, with the exception of the 66 controversial Emoji compatibility characters (for logos, etc.), which the USNB agreed to withdraw, and the 10 regional indicators, which the US will need to pursue further. The five culturally-iconic characters were agreed to, but were re-named “EMOJI COMPATIBILITY SYMBOL-1,” “EMOJI COMPATIBILITY SYMBOL-2,” etc. A total of 150 characters were added to the set proposed by the US in N3582/N3583 (L2/09-25 and L2/09-26R); the new additions derive from the proposal from the Irish and German NBs (N3607=L2/09-114). The final Emoji set in Amendment 8 includes 61 name changes and 119 glyph changes to those characters proposed by the US; these changes were based on N3607 and subsequent discussions with the Irish and German NBs. Detailed ballot comments on names and glyphs of Emoji characters are expected from various national bodies.

Two historic Asian character proposals, Jurchen and Nüshu, were discussed in separate ad hocs. Based on the discussion, the Chinese NB will provide revised versions of these proposals before the next WG2 meeting. Note that Jurchen is a large set: 1425 characters (1376 plus 49 radicals). The ad hoc reports summarize the outcome of the discussion, see L2/09-167 (=N3639 for Jurchen) and L2/09-170 (=N3635 for Nüshu).

Outstanding issues on “Hungarian,” listed in L2/09-165 (N3637), were discussed in an ad hoc meeting. Some agreement was reached on a few issues: the script will be located in a single block in the SMP; the script has casing; two punctuation marks, six numbers, and seven “bug” characters are eligible for encoding; historic ligatures need not be separately encoded. The name of the script remains a stickler, however, as does the question of whether to encode several historical characters. The authors of the proposal from Ireland and Austria (L2/09-142=N3615) agreed to revise their proposal in accordance with the agreements made at the ad hoc and will re-submit it. Hungary expressed willingness to withdraw its request for historic ligatures in return for meeting its other strong requirements. It is hoped the parties can continue to work towards the eventual approval of this script. The ad hoc report, L2/09-168 (N3640), summarizes the outcome.

IRG has a draft Principles and Procedures document available for comment (N3562), and a major revision to Annex S is under review by the IRG.

Jamo filler text was changed in Clause 22.1 of the working draft for the second edition of 10646 (=Clause 26.1 of the current standard); the new text answers questions raised by Unicode/USNB (N3624=L2/09-149) and was acceptable to the Unicode liaison and USNB representatives. The Republic of Korea also requested a note be added to Clause 26.1 of 10646, stating that Hangeul text can be represented in different ways in the standard, and KS X 1026-1 “provides guidelines on how to ensure interoperability in information exchange.” The note was approved for inclusion in Amendment 6.

No special discussion took place on the IVD, but NB representatives (and Japan in particular) were reminded to carefully review the documents N3590 (L2/09-080R, “Handling Glyph Shapes for Government Use in WG2/N3530 via Variation Sequences”) and N3591 (regarding payment of the registration fee for IVD submissions).

At the OWG-SORT meeting, the informative text about preprocessing Hangeul submitted by the US and the longer text from Republic of Korea were removed from annex sub-clause C.4.1 in the FPDAM 1 ballot of ISO/IEC 14651: 2003. The Amendment progressed to FDAM balloting. The US and Republic of Korea agreed to work further on developing acceptable text for an annex on preprocessing Hangeul.

The next WG2 meeting will be held in Tokyo October 26-30. The spring 2009 meeting is slated to take place in Mountain View from 19-23 April 2010. The USNB needs to take action to verify its offer to host the April 2010 meeting.