

Title: UTC Liaison Report from WG2 #59

Date: 2012-02-20

Source: Peter Constable, Unicode Liaison to SC2

Action: For review by UTC, Unicode officers

A WG2 meeting was held in Mountain View, CA from February 13 – 17, 2012. This document reports on select topics arising from those meeting that will be of interest for the Unicode Consortium.

For the complete WG2 resolutions, see [L2/12-098](#) (= N4254).

Progress on ISO/IEC 10646 amendments and new editions

During WG2 #59, a report was received from the SC2 secretariat indicating that the FDIS text for the third edition had been circulated for final balloting.

Processed ballot comments on PDAM text for Amendment 1 of the third edition were processed at this meeting. For a significant portion of the additions in Amendment 1, there was sufficient consensus that those portions could be considered ready to progress to enquiry stage. However, other portions of Amendment 1 were not considered mature enough to progress. Accordingly, it was decided to split off the latter out of Amendment 1 and for Amendment 1, with the more limited and mature content, to progress the enquiry stage. Thus, the editor will be preparing a DAM draft for circulation and balloting. The target dates for Amendment 1 are: DAM 2012-03, FDAM 2012-11.

Action was taken at WG2 #59 to start Amendment 2. A proposal will be circulated for SC2 ballot shortly. (Working group process for Amendment 2 will be a bit different from past amendments; this is discussed in detail below.) The target dates for this amendment are: PDAM 2012-03, DAM 2012-11 and FDAM 2013-07. Initial repertoire for Amendment 2 includes all of the characters removed from Amendment 1, as discussed above, plus a number of characters and scripts approved at WG2 #59.

The following is a tentative schedule for future WG2 meetings and potential timetable for progress on amendments. (This was presented in the liaison report after the Helsinki meeting; the projected timetable is unchanged from that report.)

WG2 Meeting	Amd 1	Amd 2	4 th Edition ??
June 2011 (Helsinki)	Progress on PDAM	n/a	n/a
Feb 2012 (Mountain View)	Authorize 5-month DAM ballot	Initiate Amd 2; authorize PDAM ballot(s)	n/a
Oct 2012 (Chiang Mai)	DAM ballot disposition; authorize FDAM ballot	Dispose of PDAM comments; authorize 5-month DAM ballot	Initiate 4 th edition & authorize CD ballot(s)??
June? 2013 (Berlin? Vilnius?)	n/a	DAM ballot disposition; authorize FDAM ballot	Dispose of CD comments & authorize DIS ballot??

This is, of course, speculative: there is no current plan of record for a 4th edition, and the rate at which projects progress could be other than suggested here. This is just one possible course of events.

WG2 Working Methods

Changes in JTC1 procedures that went into effect in 2011 entailed changing schedules for SC2 ballots and WG2 meetings. This, in turn, led Unicode officers and UTC to consider ways that WG2 process could be made more agile. At the Helsinki meeting (WG2 #58), the US presented a document (WG2/N4061 = [L2/11-192](#)) proposing that WG2 adopt more agile working methods. Specifically, the proposal was to authorize the editor to resolve ballot comments independently and to prepare new committee-stage drafts for circulation and balloting, potentially with new repertoire, without requiring WG2 to meet and take explicit actions authorizing these changes. (This is all permissible under the JTC1 procedures.)

This proposal was generally well received at the Helsinki meeting, though reservations were voiced about allowing the editor to introduce new scripts in PDAM drafts without a WG2 resolution. WG2 took a resolution allowing the editor, on a trial basis, to introduce limited changes on his own in the course of disposing of ballot comments, to circulate additional PDAM drafts between WG2 meetings, and to adopt use of electronic means of collaboration with working group experts.

At WG2 #59, that trial of revised working methods was reviewed. The project editor indicated that he had benefited since the draft disposition of ballot comments could be discussed by email, and as a result the disposition was better worked out before the WG2 meeting. On the other hand, the editor also expressed a level of frustration at not being able to add new characters or scripts without WG2 action. Initially, he had anticipated disposing of comments and issuing a second PDAM ballot before WG2 #59, but because he was constrained from adding any new repertoire, he determined that the limited benefits of a second PDAM ballot were not worth the work it would take. Yet there were proposals that were sufficiently mature at that point and that could have made a second PDAM ballot worthwhile.

As a result of the editor's feedback, WG2 re-considered earlier reservations about allowing the editor to add new characters at the PDAM stage without explicit WG2 action to authorize those additions. It was noted that WG2 could word a new project / project-subdivision proposal so as to give the project editor authorization in this regard and that, provided this was constrained to changes at the committee (CD / PDAM) stage, then WG2 would still be given opportunity to evaluate resulting drafts before agreeing to progress drafts to the enquiry (DIS / DAM) stage.

WG2 was already anticipating initiating a project-subdivision proposal for Amendment 2 at this meeting. After discussing the above issues, there was consensus (effectively) to expand the trial of new working methods for Amendment 2, authorizing the project editor in advance to add new repertoire to the amendment while still at the committee stage. Thus, WG2 adopted a resolution (M59.17) to have a project subdivision proposal crafted in this way. The project subdivision proposal ([WG2/N4248](#) = L2/12-xxx) includes the following text (emphasis added):

Scope: Incorporate an additional 848 characters as identified below *and possibly additional characters while this amendment is at the committee draft stage...*

Future additions: *While the amendment is still in development at the committee draft stage, the project editor is authorized to add other new scripts or characters that are evaluated as mature proposals with adequate review and consensus from participating NB experts. Any such additions at the committee draft stage would undergo at least one PDAM ballot before the draft is advanced to enquiry (DAM) stage.*

As a result, all of the key recommendations made in L2/11-192 have been adopted by WG2 at least on a trial basis.

It should be noted that the action proposed is still pending ballot by SC2 members. In the event that it is approved, UTC and L2 should plan to allow for two (or, perhaps, even three) PDAM2 ballots before the

WG2 #60 meeting in October 2012, and that there would be opportunity for new script or character additions to be incorporated in the second (or third) PDAM—provided there are sufficiently mature proposals that have been reviewed by enough experts with a stake in the issues to indicate a good level of consensus.

CJK Unified Ideographs Extension E

Lu Qin reported on IRG’s progress on Extension E. They anticipate that Extension E will be mature and ready for submission to WG2 by June 2012. Thus, it is conceivable for Extension E to be added into Amendment 2.

Items from liaison report to WG2 #59

UTC requested that various items be covered in the liaison report submitted to WG2 prior to the Mountain View meeting. The following is a summary of outcomes:

- Emoji variation sequences: WG2 agreed (M59.04) to add these to Amendment 2.
- Change to syntax for ideographic description sequences: WG2 agreed (M59.03) to revise the text in Annex I as requested. The recorded resolutions do not make clear whether this change would be made in Amendment 1 or Amendment 2; my understanding is that it will be changed in PDAM2.
- Alternate IRG process for “urgently-needed” ideographs: No WG2 action was taken. Annex C of the IRG *Principles and Procedures* document discusses urgently-needed characters and allows for NBs to submit proposals for such characters directly to WG2, by-passing IRG; but it does not discuss a regular process for IRG. IRG will consider the US/UTC proposal ([L2/11-442R2](#) = N4230).
- Lithuanian text processing issues: This will be discussed in a separate section, below.

Use of variation sequences for CJK compatibility characters

A proposal ([L2/11-361R](#) = N4165) was submitted by Ken Lunde on behalf of UTC to register variation sequences for already-encoded CJK compatibility ideographs, beginning with a set of 89 compatibility characters considered critical for Japanese use.

Two specific details of L2/11-361R are important to note:

- It proposed that 92 variation sequences for these 89 compatibility characters—in the case of three characters, it allowed for distinct glyphs corresponding to distinct JIS X 0213 and IBM sources.
- It proposed that sequences be registered in the IVD (as opposed to standardized variants).

There was some lively discussion of this proposal. One key question raised was what the intent of the proposal was in terms of what distinctions should be made by variation sequences:

- Capture only the information lost by normalization—viz., the distinction between a compatibility ideograph and the unified ideograph that is its canonical decomposition?

- Capture distinctions between different sources for the compatibility character (as done in L2/11-361R)?
- Capture distinctions between some set of glyph variants (assuming some analysis of glyph variants that cuts across all sources)?

An ad hoc committee met to consider these issues, and the outcome was a *different* proposal document, prepared by Ken Lunde during the Mountain View meeting: N4246R = [L2/12-095](#). This alternate proposal differs from L2/11-361R in important ways:

- It proposes use of standardized variants rather than IVD.
- It captures only the information lost by normalization (e.g., there would not be two standardized variant sequences corresponding to distinctions between JIS X 0213 and IBM sources / glyphs).
- Because (per the previous item) the set of variants is determined by existing character properties, it proposes sequences for all 1002 CJK compatibility ideographs at once, rather than only proposing sequences for a subset with others to follow over time.

WG2 has not taken any action on this topic other than to invite input from NBs.

Lithuanian text processing issues

Lithuanian issues were considered at UTC #130 as a result of various documents that had been submitted prior to the meeting. These included some documents (e.g. [L2/12-048](#), [L2/12-067](#)) that describe problems with font and keyboard implementations. In response, UTC took the following action:

[[130-C2](#)] Consensus: The UTC encourages its member companies to review their implementations to ensure the correct input and display of all Lithuanian characters.

However, further input received from Lithuanian experts (including some late-breaking documents such as [N4191A](#)—an addendum to N4191 = L2/12-026, as well as input from Lithuanian delegates present at the WG2 meeting) clarified that text input and display were not the only concerns, or even the most significant concerns. Rather, the issues that, reportedly, are presenting the biggest obstacles for Lithuanian users pertain to sorting and related text processes such as search and copy/paste.

Noting that the same problems do not arise in current implementations for other pre-composed Latin characters used in European languages, this has led Lithuanian experts to believe that adding other pre-composed characters for Lithuanian accented letters would provide a solution.

These issues were discussed in an ad hoc committee meeting on Lithuanian. A more complete perspective on the likely efficacy of various potential solutions was provided. The outcome of the ad hoc did not include any specific recommendations for WG2 action. However, the ad hoc committee did arrive at some recommendations for the Lithuanian national body and Lithuanian users, including suggestions involving CLDR and interaction with the CLDR project. See the report of the Lithuanian ad-hoc committee meeting ([L2/12-097](#) = N4242) for details.

Given the broader scope of concerns discussed during the WG2 meeting than had been discussed at UTC #130, UTC may want to revisit this topic at their next meeting.