

Title: UTC Liaison Report from SC2 #22, WG2 #66

Date: 2017-9-30

Source: Peter Constable, Unicode Liaison to WG2

Action: For review by UTC, Unicode officers

SC2 and SC2/WG2 meetings were held in Huhhot, China, September 25 – 29, 2017. This document reports on select topics arising from these meetings that will be of interest for the Unicode Consortium.

For full details of WG2 recommendations to SC2 at WG2 #66, see [L2/17-349](#) (= WG2/N4874). For resolutions of SC2 #22, see [L2/17-350](#) (= SC2/4566).

Progress on ISO/IEC 10646 amendments and new editions

Development of the 5th edition of ISO/IEC 10646 has essentially been completed. Some changing requirements within ISO delayed publication of the 5th edition, but those were resolved at the SC2 meeting. See below for more details on publication issues for ISO/IEC 10646.

Ballot comments on a third committee draft for Amendment 1 (PDAM1.3) were discussed and resolved. Amendment 1 will progress to enquiry draft (DAM1) this fall. A successful ballot is expected, in which case Amendment 1 will proceed directly to publication in mid-2018.

Ballot comments on the first committee draft for Amendment 2 were discussed and resolved. Two more committee-level drafts are expected, in November 2017 and March 2018. Publication of Amendment 2 is expected to be in Q1 2019.

Synchronization of repertoire between Unicode versions and ISO/IEC 10646 5th edition is roughly as follows:

- Unicode 10.0 repertoire is the same as ISO/IEC 10646 5th edition plus some additions from Amendment 1
- Unicode 11.0 (June 2018) is expected to be the same as Amendment 1 plus some additions from Amendment 2.
- Unicode 12.0 (April 2019) is expected to be the same as Amendment 2 plus maybe some additions from the 6th edition.

The following table summarizes the anticipated schedule for progress on ISO/IEC 10646:

SC2 Meeting	5 th Edn	5 th Edn, Amd 1	5 th Edn, Amd 2	6 th Edn
September 2017 (Huhhot)	Publication issues resolved.	PDAM1.3 ballot comments resolved. Preparation of DAM1 approved.	PDAM2.1 ballot comments discussed. Preparation of additional PDAM2.x ballots approved.	n/a
June 2018 (London)	n/a	Resolve DAM1 ballot comments, if required.	Resolve PDAM2.3 ballot comments. Authorize DAM2 preparation.	Initiate project subdivision ballot (?)
June 2019 (US)	n/a	n/a	Resolve DAM2 ballot comments, if required. (May have already proceeded to publication.)	Resolve CD ballot comments; authorize DIS preparation.

Progress on ISO/IEC 14651

Amendment 1 to ISO/IEC 14651:2016 (4th edition) has just been published. The CD ballot for the 5th edition has just completed without comments, and a DIS ballot will be issued. If the DIS ballot passes without comments, it will proceed directly to publication.

No new work for 14651 is anticipated in the coming year. It is anticipated that Amendment 1 of the 5th edition will be initiated at the SC2 meeting in June 2018, and that it will be synchronized with the repertoire for Unicode 12 (April 2019).

During discussion, there was mention of a potential future need for a collation tailoring that treats variation selectors in ideographic variation sequences with non-zero weights, and a question of whether anything currently in ISO/IEC 14651 might prevent such tailoring. No specific proposals were made at this meeting, however. The Japanese delegation was invited to submit a document on this topic, if needed, at a future SC2 meeting.

Publication of ISO/IEC 10646

Recent editions and amendments have been published as a pair of zip files: one containing multiple PDF files (main text, charts), and one containing data files. Some months ago, ISO raised some concerns with this: they prefer to have an organization more consistent with other ISO standards:

- A single PDF file with conventional ISO file naming and watermark.
- Data files located loose (not zipped) on an ISO server.

Because the size of the code charts would make a single PDF for the text cumbersome, they proposed that a second PDF with the code charts be kept on an ISO server. In intervening months leading up to the Huhhot meeting, the project editor was working with ITTF to find a solution that respected Unicode's rights in relation to the code charts (developed by Unicode for use by ISO), by having PDFs for the charts located on Unicode.org.

In parallel, ISO noted that requirements for the text of clauses 2 and 3 had been updated, and that the FDIS text for the 5th edition did not reflect these changes. The project editor was prepared to make necessary changes. However, the SC2 secretariat was concerned that these changes might warrant re-balloting.

Due to these happenings, the publication of the 5th edition was delayed, and there was no consensus among the SC2 secretariat, the project editor and ISO on how to deal with publication of the 5th edition.

In the Huhhot meeting, SC2 resolved to complete publication of the 5th edition with limited changes, and to postpone implementation of a new approach until the 6th edition. The following SC2 resolution captures decisions regarding the 5th edition:

RESOLUTION M 22-06: Publication of ISO/IEC 10646 (Ed.5)

SC 2 requests ITTF to publish the final text of 5th edition of ISO/IEC 10646 as contained in SC 2 N 4511 with the following changes:

- editorial updates of Foreword, Clause 2 and 3 to apply the new boilerplate text in accordance with the ISO/IEC Directives Part 2.
- combine the pdf files of the main part and code chart part into one file and remove the file links on the main part to the code chart part so that the pages are numbered sequentially without skipping any page numbers
- change the file links to the electronic inserts (text files) from the main text to hyperlink in accordance with ISO TMB Resolution 38/2017

SC 2 instructs its Secretariat to submit the updated text to ITTF.

Details for how code charts and data files will be handled in the 6th edition will require further discussion and negotiation between Unicode, ISO and the SC2 secretariat.

SC2 process and cooperation with UTC

Prior to the Huhhot meeting, documents were submitted to SC2 from Japan and from Ireland voicing concerns with process in relation to certain character additions that were added in committee drafts for Amendment 1 and Amendment 2. SC2 has in past resolutions agreed that the project editor be free to add characters into committee drafts without prior action from WG2 or SC2 provided that proposals for adding such characters are sufficiently mature. The project editor has sometimes made such evaluation on the basis of discussion and actions taken within UTC.

In the Japan and Ireland contributions, concerns are voiced that this process has not worked to their satisfaction. The specific concerns of each national body were:

- Small kana characters were added in PDAM1.2 on the basis of UTC discussion. Japan ballot comments requested that these be removed; the disposition taken was to move them from Amendment 1 into PDAM2. Japan's position, however, was that the proposal for addition of these characters was not sufficiently mature for them to be added to any amendment.

Note: While some national bodies, such as US, felt that these characters were sufficiently mature for committee-level balloting, Japan has an underlying concern that these characters are of particular interest for Japan, and hence that input from Japan should have been sought before they were added to a committee ballot.

- Ireland voiced concern that Unicode has, in effect, “fast-tracked” new symbol characters used for emoji to publication without adequate time for technical comments and balloting within SC2.

Note: During discussion, the Ireland and UK representatives pointed out that the emoji additions were published in UTS #51 version 5.0, which was released in May, prior to the release of Unicode 10.0; and that these new characters were included in PDAM1.3, but this was not distributed until June, hence there was no opportunity for national bodies to submit PDAM1.3 ballot comments prior to when these characters were released by Unicode. Also note, though, that the characters were also included in the PDAM1.2 ballot, which national bodies were able to comment on earlier. However, the PDAM1.2 ballot was not closed until after the Q1 UTC meeting, and UTC did not consider the PDAM1.2 ballot comments until the May meeting, at which point it was deemed too late to entertain comments such as proposed name changes.

An ad hoc meeting was held to discuss process issues. This meeting was attended by representatives from Ireland, Japan and UK, as well as by US / UTC representatives, Debbie Anderson, Lisa Moore, Michel Suignard, and Peter Constable. Agreement was reached on various ways in which process and collaboration between UTC and WG2 could be improved. These points are captured in [L2/17-352](#) (= WG2/N4903).

The results of the ad hoc meeting are reflected in the following SC2 resolution:

RESOLUTION M 22-10: Process for Synchronization of ISO/IEC 10646 and Unicode standard

In response to the contributions of Japan and Ireland contained in Document SC 2 N 4549 and SC 2 N 4551, SC 2 requests WG 2 to improve the synchronization process between SC 2/WG 2 and UTC as reported by WG 2 N 4903. If the project editor concludes to add new proposals that were not discussed in prior SC 2/WG 2 meetings to a CD/PDAM ballot text, SC 2 also requests the project editor to add editor's notes to the text to be balloted so that the SC 2 national bodies can be aware of it.

SC 2 further requests that the WG 2 Convenor post the relevant documents of such additions to the WG 2 document registry in advance of the start of the ballot.

Rules for IVD registration

UTS #37, Unicode Ideographic Variation Database, currently requires that registered ideographic variation sequences are permitted only for CJK ideographic variants that can be unified with the base character under current CJK unification rules. During the WG2 meeting, the IRG Rapporteur presented a preliminary proposal ([L2/17-370](#) = WG2/N4829, section 12) to allow this rule to be relaxed for certain cases in which characters could be shown to be variants yet did not satisfy the rules for unification.

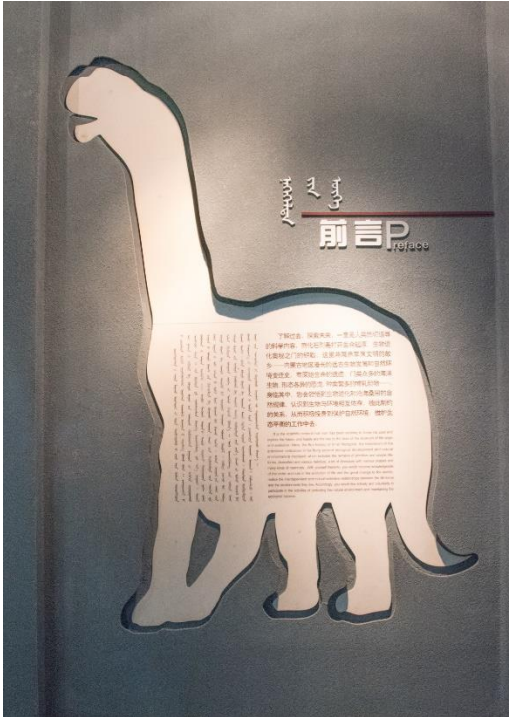
This proposal was discussed by WG2. The cases identified by IRG were broadly of two types, according to whether the variation was solely in regard to the structural arrangement of identical elements. It was noted by Michel Suignard that, in ICANN processes, there is a need to identify several kinds of variants, including orthographic, lexemic and semantic variants as well as visual variants.

IRG will be revisiting this before the next UTC meeting. A request was made for IRG to provide a proposed revision to the rule that would give a clear constraint on what non-unifiable glyphs could be treated as variants within IVD registrations.

It is anticipated that this topic and a corresponding IRG contribution will be brought for discussion at the October UTC meeting.

On a lighter note... dinosaurs in Inner Mongolia

National body comments on PDAM1.3 included discussion of dinosaur symbols. Hohhot was a fitting venue for such discussion since Inner Mongolia is an important location for dinosaur fossils. The Inner Mongolia hosts treated SC2 and WG2 participants to a visit to the Inner Mongolia Museum, which has a large exhibit hall dedicated to dinosaurs.



Entrance to the dinosaur exhibit hall at the Inner Mongolia Museum.