Date: 2008-4-21

ISO/IEC JTC 1/SC 2/WG 2
Universal Multiple-Octet Coded Character Set (UCS) - ISO/IEC 10646
Secretariat: ANSI

Doc Type	**Liaison Report**
Title:	**Unicode Liaison Report — WG 2 meeting 52**
Source:	**Peter Constable—Unicode Liaison**
Action:	**Review by WG2 and SC2 member NBs; WG2 action requested in relation to publication of IDS**
Distribution:	**SC2/WG2**

The Unicode Consortium is pleased to report on-going progress in development of the Universal Character Set resulting from collaboration with SC2, as well as progress on the Unicode Standard and related standards and technologies.

Release of Unicode 5.1

Unicode 5.1 was released on April 4, 2008. This release of Unicode is synchronized with ISO/IEC 10646:2003 plus Amendments 1 through 4. The previous version, Unicode 5.0, was closely aligned with Amendments 1 and 2. (It also included four Devanagari characters added in Amendment 3.)

This release includes additional characters added to the UCS in Amendments 3 and 4 (1,624 additional characters). The extensive set of character properties in the *Unicode Character Database* (UCD) were updated to cover all newly-added characters. Among the UCD changes were new properties related to text segmentation, and documentation of U-source ideographs.

A number of standardized named sequences were added for Tamil. The Unicode Consortium suggests to WG2 that these be added to the list of named sequences in ISO/IEC 10646. A separate document has been submitted from UTC and the US: please see N3407.

A significant change related to the UCD is publication of the UCD in XML format. The schema for this form of the UCD is documented in UAX #42: Unicode Character Database in XML. See UAX #44: Unicode Character Database (UCD) for details.

Changes were made to the Unicode Stablity Policy to strengthen normalization stability, to extend constraints on property values, and to add a stability policy regarding case pairs. Please see Unicode Character Encoding Stability Policy for details.

The Unicode Technical Committee continues to update the specification of UTF-8 to address security concerns as they are identified. In Unicode 5.1, the conformance clauses were updated adding additional constraints on conversion of ill-formed UTF-8 sequences.

Several of the Unicode Standard Annexes that describe various character properties or text processing guidelines were also updated.

For complete details regarding changes in Unicode 5.1, please see Unicode 5.1.0.

Unicode Technical Committee Review of WG2 #51 (Hangzhou)

Various actions were taken by WG2 in Hangzhou in relation to FPDAM 5 and PDAM 6 that impacted prior decisions made by UTC in relation to Unicode 5.1 or future versions of the Unicode Standard. WG2 also took actions adding new characters to one of these amendments which had not previously approved by UTC for addition to Unicode. UTC reviewed the WG2 actions and, except as noted below, took actions to implement corresponding changes in Unicode so as to maintain synchronization.

As mentioned in the Liaison report submitted in Hangzhou (N3361), UTC had approved the encoding in Unicode of Avestan script as described in WG2/N3197 with the exception of one character: 10B38 AVESTAN SEPARATION POINT. UTC experts considered this character further after WG2 #50 (Frankfurt) with concern over the proliferation of script-specific word-separator-dot characters that differ from script to script only in minor differences in appearance. As mentioned in the previous Liaison report, UTC concluded that it would be better to encode one generic (non-script-specific) word-separator dot character that could be used for multiple scripts with the assumption that the range of acceptable glyph representations would allow for the differences in appearance in different scripts. This position was presented by the Liaison at WG2#51, and by the US in ballot comments on PDAM 5. These comments were accommodated by WG2#51 only to the extent of approving the encoding of U+2E31 WORD SEPARATOR MIDDLE DOT; the AVESTAN SEPARATION POINT was not withdrawn from FPDAM 5. UTC continues to recommend to WG2 that 10B38 AVESTAN SEPARATION POINT be withdrawn from Amendment 5.

In all other respects, UTC endorses the actions taken by WG2 #51 in relation to FPDAM 5 and PDAM 6.

New Edition of ISO/IEC 10646 (WG2/N3274, N3275, N3276, N3360, N3362, N3364, SC2/N3985)

In view of the approval by SC2 of the *Project Subdivision Proposal for Consolidating ISO/IEC 10646: 2003 and Amendment 1 - 5 into a New Edition of ISO/IEC 10646* (N3364), the Unicode Consortium is beginning to plan for a minor release, Unicode 5.2, that would tentatively synchronize with Amendments 5 and 6 in order to facilitate synchronization of a new edition of ISO/IEC 10646 with that version of the Unicode Standard.

Publication of Ideographic Description Sequences—*WG2 action requested*

The Unicode Consortium would like to request that the Ideographic Description Sequences for the encoded CJK ideographs be included as part of the 10646 data files.

Common Locale Data Repository (CLDR) Version 1.5.1

An update release of the CLDR, Version 1.5.1, was released on December 21, 2007. This release included significant additions to the data structure and process for computing time-zone names, and

additional data for finding default script or country given a language, or the converse. The structure has also been updated for the latest version of BCP 47 (IETF Language Tags), and new currency codes.

Version 1.6 of CLDR is scheduled for July 2008.

The Unicode Consortium feels confident that National Bodies and experts represented in WG2 will find the CLDR offers useful benefits in enabling support in software products for languages and cultures from across the world. As always, experts in WG2 are invited to participate in the on-going development of CLDR. Current information on CLDR can be found on the Unicode Web site at http://unicode.org/cldr/.