

Title:	Comments accompanying the US negative vote on the FPDAM2 to ISO/IEC 10646-1:2000
Source:	INCITS/L2
Action:	Forward to INCITS

The US National body is voting NO on the FPDAM2 to ISO/IEC 10646-1:2000. It will change its vote to YES if the technical comment T.8 is accepted.

Technical Comments:

T.1 Normative References

Update the reference to the Unicode Bidirectional Algorithm to version 3.2

T.2 Greek

Add the following characters:

GREEK CAPITAL LUNATE SIGMA (proposed code position: 03F9) as documented by INCITS L2 02/314R,
GREEK CAPITAL LETTER SAN (proposed code position: 03FA) as documented by INCITS L2/313R,
GREKK SMALL LETTER SAN (proposed code position: 03FB) as documented by INCITS L2/313R

T.3 Arabic

Change the character name for U+0658 to ARABIC MARK NOON GHUNNA

T.4 Symbols

Add the character VERTICAL LINE EXTENSION (proposed code position: 23D0) as documented by WG2 N2508. The character addition would complement the existing HORIZONTAL LINE EXTENSION at code position 23AF.

T.5 Annex A.1 Collections of coded graphic characters

a) Change the entry 105 to read:

105 RTL ALPHABETIC PRESENTATION FORMS

b) The code range for collection 110 should read 2B00-2B7F.

c) To complete the change for collection 63, the first entry for collection 63 in annex A.1 (between 62 and 64) should read:

63 (Collection specified as union of other collections)

d) Collection 251 SCRIPT-SPECIFIC FORMAT CHARACTERS should not contain the collection 207

e) After the collection 381, a new collection should be added as follows:

382 CJK COMPATIBILITY IDEOGRAPHS-2003 * described as Collection 381 with the addition of U+FA70-FAE9. This is only required if the addition of the range FA70-FAE9 is maintained in this proposed amendment. Note that the US is asking for its removal in comment T8.

f) In the alphabetically list in Note 3, add collections 98 and 99 to the 'Arrows' entry and 380, 381, 382 (as appropriate, see above) to the 'Ideographs' entry.

T.6 Annex A.2 Block lists

Modify the code position range for MISCELLANEOUS SYMBOLS AND ARROWS to 2B00-2B7F

T.7 Annex Q and R

In the context of merging the two parts of ISO 10646, the US is asking for an overhaul of Annex Q and R of Part 1 by making the code mapping for Hangul Syllables (Annex Q) and the Names of Hangul Syllables (Annex R) available as separate text files. This will make the information much more usable.

T.8 DPRK Compatibility characters

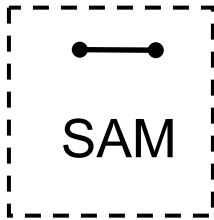
The US asks for removal of characters from FA70 to FAE9.

Rationale:

The US has found many discrepancies in the list of characters (incorrect mapping). For example, the reference mapping for characters FA9D to FAE9 is shifted by one position, leaving the proposed code position FA9D without reference mapping to a CJK Unified character. Some characters like FA8C and FA90 and many others seem to have incorrect mappings. Furthermore, it looks like not enough research has been done to look for existing variants in the CJK repertoire that could have been used for encoding the KP1 characters. Given these issues, the US does not see these proposed additions as mature enough for encoding and is therefore asking for their removal pending further study.

T.9 Syriac

Change the glyph for 070F SYRIAC ABBREVIATION MARK to include a default rendering in addition to the current picture, suggested new glyph:



T.10 Annex F, Soft Hyphen format character

The SOFT HYPHEN entry in clause F.1.1 should be reworded as follows:

SOFT HYPHEN (00AD): SOFT HYPHEN (SHY) is a format character that indicates a preferred intra-word line-break opportunity. If the line is broken at that point, then whatever mechanism is appropriate for intra-word line-breaks should be invoked, just as if the line break had been triggered by another mechanism, such as a dictionary lookup. Depending on the language and the word, that may produce different visible results, such as:

- inserting a graphic symbol indicating the hyphenation and breaking the line after it,
- inserting a graphic symbol indicating the hyphenation, breaking the line after the symbol and changing spelling in the divided word parts,
- not showing any visible change and simply breaking the line at that point.

The inserted graphic symbol, if any, can be take a wide variety of shapes, such as HYPHEN (2010), ARMENIAN HYPHEN (058A), MONGOLIAN TODO SOFT HYPHEN (1806), as appropriate for the situation.

When encoding text that includes explicit line breaking opportunities, including actual hyphenations, characters such as HYPHEN, ARMENIAN HYPHEN, and MONGOLIAN TODO SOFT HYPHEN may be used, depending on the language.

When a SOFT HYPHEN is used to represent a possible hyphenation point, the spelling is that of the word without hyphenation (for example: "tug<00AD>gumi"). When encoding text that includes hard line breaks, including actual hyphenations, the spelling must reflect the changes due to hyphenation (for example: "tugg<2010>" / "gumi").

Editorial Comments:

E.1 Clause 25.3 Alternate coded representations

In the 6th line add a 's' at the word 'representation' (after multiple).

E.2 Clause 27 CJK Unified Ideographs

Add a mention that the clause number is now 28 (following the insertion of clause 24).

End of US comments