ISO/IEC JTC1/SC2/WG2
Coded Character Set
Secretariat: Japan (JISC)

Doc. Type:	**Draft Disposition of comments**
Title:	**Draft Disposition of comments on SC2 N 3761 (PDAM text for Amendment 2 to ISO/IEC 10646:2003)**
Source:	**Michel Suignard (project editor)**
Project:	JTC1 02.10646.00.01
Status:	For review by WG2
Date:	2005-01-23
Distribution: WG2	
Reference:	SC2 N3761, N3779, WG2 N2097, N2765, N2829, N2869-2871
Medium:	Paper, PDF file

Comments were received from Canada, China, Germany, Ireland, Japan and USA. The following document is the proposed disposition of those comments. The disposition is organized per country.

Note – A positive vote without comment was also recorded from Iran but was received too late to be mentioned on N3779.

Note – The full content of the ballot comments (minus some character glyphs) have been included in this document to facilitate the reading. The dispositions are inserted in between these comments and are marked in **Underlined Bold Serif text,** *with explanatory text in italicized serif*.

Canada: Negative:

Technical Comments concerning NKO

1a) The English Block name contains an apostrophe while the character names do not. I believe the English names may not contain an apostrophe (I don't know why though), this is not true of the French names, to ensure coherence between Letter and Block names in the English version, the script should be called Nko inEnglish (N'ko in French).
Accepted
Also requested by Japan and the US in comment T.1. It should be noted that related to that issue of block name guidelines, the US has requested that block name follows the same guideline as for characters in its comment T2 to the 10646-2003 FPDAM1.

1b) Canada also requests that – to make it clear in the standard – that a paragraph / clause explaining the naming rules for Block Names be added. The naming rules should be identical to the rules for naming of characters except for allowing Upper and Lower case letters in the block names. This would also exclude the Apostrophe in the Block Name.
Accepted
See above

2) The explanations provided in Markham by the proposer (from Ireland) to explain the inclusion of the following characters ("they are used in historical documents to show the evolution of the script") is not wholly convincing:
07E8 NKO LETTER OLD JA
07E9 NKO LETTER OLD CHA
07EA NKO LETTER OLD RA
It is quite easy to find books on the history of Latin, Greek or let's say the Phoenician script, which will show old and new forms of the same characters. Hence, Canada does not understand why these characters are not treated as historical glyph variants. Except in treatise dealing with the history of the script, I believe these old forms are not used simultaneously with the new ones. These characters are just historical glyph variants of characters that are otherwise provided for in the proposal. Encoding such would be akin to encoding various historical forms of Latin letters such as Uncial or Fraktur, which is clearly not desirable. Canada is asking for their removal.
WG2 discussion

3) Some of the proposed nko combining marks have identical shape and canonical combining class to already encoded marks (07EB, 07ED, 07EE,...) - these combining marks would look and behave in the same way as already encoded combining characters. During the discussion in Markham, it was answered that these characters have a different origin from the common and generic combining marks (which may combine with any other base character) and may sometimes adopt different shapes. But this may also be true of umlauts or accents across languages: they don't always adopt the same shape (an umlaut may be represented inside the letter or an e above the letter).
http://fr.wikipedia.org/wiki/Umlaut
http://www.peter-doerling.de/Lese/Sutterlin0.htm
Canada finds the justification in document N2765 pretty poor. The statement about directionality issues is totally bogus. As for font-binding issues, they seem out of scope in a character-encoding standard. In Latin script, a simple acute accent can range from almost vertical (in Polish for example) to almost flat and will of course be set at a different height depending on the base. The usual fontlevel solution is to use ligatures, which allow all the desirable variability and would seem to be able to do so also for Nko.
WG2 discussion
It would help to know the exact list of problematic combining marks. The values 07EB, 07ED and 07EE are mentioned. In addition, 07F3 is probably implied by the ellipsis in the Canadian comment.

China: Negative

Technical comments:

Phags-pa:

Few agreements were achieved in WG2N2829 (Consensus on the encoding of the Phags-pa script in the PDAM code chart, Markham, Canada, 2004). It's true that names of A862 and A863 are reflected in the PDAM2, but more questions raised in WG2N2829 need further discussion by interested parties. China requests that her comments be fully reflected in next round of ballot on the bases of her comments being fully discussed by interested parties.

Technical details:
1. Two vowel letters OE and UE and four consonant letters expressing initial consonants of "审, 影, 匣, 非" in Chinese language should be at least added to the list of nominal characters.
2. A series of variants of letters should be added.
3. If information processing is taken into consideration, a syllable delimiter, a joiner and its variant are needed.
4. Khubilai Khan font should be used.
5. Three variant selectors are needed.
6. Inputting rules for single variant presentation forms should be specified.

For further details, read China-Mongolia joint proposal WG2Nxxxx please.

WG2 discussion

The document WG2Nxxxx is assumed to be WG2 N2870 (Summary of the Revised User's Agreement related to Phags-pa Script), although China has submitted two other documents in the meantime:
1. *N2869: Proposal to encode the Phags-pa script*
2. *N2871: Some problems on the Encoding of Phags-pa Script.*

All these documents do not comment on the original proposal as documented in PDAM2 but instead are describing a new proposal which is similar but has many differences in names and character contents with the previous version. These differences are not documented by the new documents. Furthermore the document N2829 which was the result of the ad-hoc committee at the last WG2 meeting is apparently not the base for further process and is not considered or answered by the new documents. For example, it was agreed that A86F should be called PHAGS-PA LETTER CANDRABINDU. In N2869, it is called simply CANDARABINDU and located at 2874.

In light of these considerations, it is necessary to conduct a new ad-hoc meeting in Xiamen, China and try to converge the various proposals (PDAM2 and WG2N2869) before any progress can be made concerning this script.

Germany: Negative

Technical Comments:

Remove the Phoenician block
Reason:
Encoding Phoenician is redundant, and needlessly proliferates Canaanite diascripts. It is in stark contrast to both the Unicode/ISO 10646 script unification model and the character/glyph model.
We are convinced that the script used in writing Phoenician, Old Hebrew, Old Aramaic, Moabite, and Ammonite is one and the same script. The problem is if we encode all of Hebrew and Aramaic with one Unicode block, the Hebrew block, and then introduce a Phoenician block, it will become quite impossible to draw well-defined lines. Will we also encode Elephantine letters, the Mesha stele, the Dead Sea scrolls in Phoenician, but Rabbinic and Modern Hebrew in Hebrew? The Phoenician block needlessly complicates text processing and goes against current practice where all these scripts are unified in one encoding, in either (a misnamed) Hebrew encoding or in Roman transliteration.

The German vote of disapproval and the proposal to remove the Phoenician block from ISO/IEC 10646/PDAM 2 is based on expertise obtained from the University of Tubingen. See also document SC2/WG2 N2097.

WG2 discussion

It should be noted that WG2 N2097 is vastly anterior to the recent discussion concerning the Phoenician script and by no mean can be categorized as supporting the position expressed above. Excerpt:

*"It can be sensible to generate, for a limited field of application, **standardized** character forms that can be used in print. This is the case, e.g. for the Ugaritic cuneiform, the Old Persian cuneiform, for **Phoenician** and for the Old South Arabian alphabet."*

The point presented by Germany has been abundantly discussed in previous documents. WG2 already addressed these issues and approved publication of PDAM2. There is no reason to go over this again.

Ireland: Negative

Technical comments

T.1. Hebrew Characters. With reference to ISO/IEC JTC1/SC2/WG2 N2840 "Proposal to add HEBREW POINT HOLAM HASER FOR VAV to the BMP of the UCS", Ireland requests the addition of HEBREW POINT HOLAM HASER FOR VAV at position U+05BA, which would entail moving HEBREW POINT QAMATS QATAN (under ballot in FPDAM-1) to position U+05C7.

WG2 discussion

Also requested by US comment T3.
Propose acceptance.

T.2. Kannada Characters. With reference to ISO/IEC JTC1/SC2/WG2 N2860 "Proposal to add six characters for Kannada to the BMP of the UCS", Ireland requests the addition of the characters
KANNADA VOWEL SIGN VOCALIC L at position U+0CE2,
KANNADA VOWEL SIGN VOCALIC LL at position U+0CE3,
KANNADA DANDA at position U+0CE4,
KANNADA DOUBLE DANDA at position U+0CE5,
KANNADA SIGN JIHVAMULIYA at position U+0CF1, and
KANNADA SIGN UPADHMANIYA at position U+0CF2.

WG2 discussion

Related to similar request in FPDAM1 Irish comment T5 concerning Bengali Danda.

T.3. Invisible letter. With reference to ISO/IEC JTC1/SC2/WG2 N2822 "Proposal to add INVISIBLE LETTER to the UCS", Ireland requests the addition of the character INVISIBLE LETTER at position U+2427.

WG2 discussion

T.4. 0238 LATIN SMALL LETTER TH WITH STRIKETHROUGH: Ireland would support the removal of this character from the PDAM pending further study of it and similar phonetic ligatures. We do not, however, insist on this point. At the very least it ought to be moved to 1D7A with other phonetic characters.

Accepted (move)

Moving it to 1D7A as also requested by the USA (ballot comment T.1).

Editorial comments

E.1 Title. Change "Phags-Pa" to "Phags-pa" in the title of the Amendment.

Accepted

E.2 Page 8 - Row 09: Phoenician. We would like clarification: is this "Row 09" or is it "Row 109"?

Accepted in principle

By definition a Row is defined within a plane and cannot exceed 255 (decimal) and is always expressed in hexadecimal notation (00-FF). Row 09: Phoenician is the Row 9 within Plane 01 (SMP) and can never be expressed as Row 109. The notation has been used in this form for all documents addressing ISO/IEC 10646 supplementary planes.

E.3 Page 10 - Row 20: Cuneiform. The glyphs at U+12031 and U+12050 should be reduced to fit into their boxes. In addition, we would like clarification: is this "Row 20" or is it "Row 120"?
Accepted in principle
See above for Row value.

E.4 Page 12 - Rows 20-21: Cuneiform. We would like clarification: is this "Rows 20-21" or is it "Rows 120-121"?
Accepted in principle
See above for Row value.

E.5 Page 14 - Rows 21-22: Cuneiform. We would like clarification: is this "Rows 21-22" or is it "Rows 121-122"?
Accepted in principle
See above for Row value.

E.6 Page 16 - Row 22: Cuneiform. The glyph at U+1221D should be reduced to fit into its box. In addition, we would like clarification: is this "Row 22" or is it "Row 122"?
Accepted in principle
See above for Row value.

E.7 Page 18 - Rows 22-23: Cuneiform. We would like clarification: is this "Rows 22-23" or is it "Rows 122-123"?
Accepted in principle
See above for Row value.

E.8 Page 20 - Row 23: Cuneiform. We would like clarification: is this "Row 23" or is it "Row 123"?
Accepted in principle
See above for Row value.

E.9 Page 22 - Row 20: Cuneiform. The characters at U+12407, U+1241D, U+12427, U+12428, U+12429, U+1242A, U+1242B, U+1242E, and U+12431 should be reduced to fit into their boxes. In addition, we would like clarification: is this "Row 24" or is it "Row 124"?
Accepted in principle
See above for Row value.

Japan, Negative

Technical Comments
(Collection and block names for N'Ko script)
In updates to "Page 1349, Annex A.1", change
128 N'KO 07C0-07FF to
128 NKO 07C0-07FF
In updates to "Page 1351, annex A.1", change
N'Ko 128 to
NKo 128
In updates to "Page 1352, Annex A.2.1", change
N'KO 07C0-07FF to
NKO 07C0-07FF
Rationale:
Current draft uses APOSTROPHE in the middle of collection name and block name for N'Ko script.
APOSTROPHE is new to UCS collection/block names. Unlike character names, we have no guidelines for

collection/block names, but introduction of a new punctuation character in such names break some existing practices. For example, XML Schema specification (developed by W3C) has a notion of character class, used in its regular expression syntax, to explicitly restrict allowed character repertoire in some context. One of the methods to specify a character class is the use of a block name, called "block escape", and the current syntax doesn't allow APOSTROPHE.

Accepted

See disposition of Canadian comment 1a) and 1b).

USA: Yes with comments:

Technical comments:

T.1 N'KO script

The N'KO script should be renamed NKO (removing the apostrophe) to be consistent with de facto block name naming convention. This also aligns the block name with the character names used within the block.

Accepted

See disposition of Canadian comment 1a) and 1b).

T.2 Seventeen Planes restriction

To improve the synchronization between the Unicode Standard and ISO/IEC 10646 and to make UTF-16 and UCS-4 equivalent in repertoire representation, the U.S. is asking to make all code positions in plane 11 (hexadecimal notation) and above permanently reserved, by doing the following changes:

- Remove the note part of Figure 1 (page 6) and the 2nd note part of Figure 2 (page 7)
- Make the 2nd note in Clause 7 (General requirements for the UCS) normative. It becomes requirement 'b.', pushing the following requirements to 'c.' and 'd.'.
- Replace the text in sub-clause 9.2 (Other Planes reserved for future standardization) by the following:
  Planes 11 to FF in Group 00 and all planes in any other groups (i.e. Planes 00 to FF in Groups 01 to 7F) are permanently reserved.
  Code positions in these planes do not have a mapping to the UTF-16 form (see Annex C).
- Replace 'character' by 'code position' in annex D (UCS Transformation Format 8 (UTF-8)) when the UCS range description goes beyond plane 10.

WG2 discussion

Propose acceptance

T.3 Hebrew script

The US is in favor of adding the following Hebrew character as proposed by document WG2 N2840.
05BA HEBREW POINT HOLAM HASER FOR VAV
This is related to US comment 10646:2003 FPDAM1 T.11 requesting the move of the currently proposed Qamats Qatan from 05BA to 05C7.

WG2 discussion

Also requested by Irish comment T1.
Propose acceptance.

T.4 Mathematical characters

a) The US is in favor of adding the following Mathematical character as proposed by document L2/04-329 (WG2 document number TBD) with modified code positions to take into account characters proposed in the FPDam1:
Miscellaneous Technical:

23DC ⏜ TOP PARENTHESIS
23DD ⏝ BOTTOM PARENTHESIS
23DE ⏞ TOP CURLY BRACKET
23DF ⏟ BOTTOM CURLY BRACKET
23E0 ⏠ TOP TORTOISE SHELL BRACKET

b) The US is in favor of adding the following Mathematical character as proposed by document L2/04-410 (WG2 document number TBD):

Combining Diacritical Marks for Symbols:
20EC   COMBINING RIGHTWARDS HARPOON WITH BARB DOWNWARDS
20ED   COMBINING LEFTWARDS HARPOON WITH BARB DOWNWARDS
20EE   COMBINING LEFT ARROW BELOW
20EF   COMBINING RIGHT ARROW BELOW

Miscellaneous Technical:
23E1   ELECTRICAL INTERSECTION
23E2   WHITE TRAPEZIUM
23E3   BENZENE RING WITH CIRCLE
23E4   STRAIGHTNESS
23E5   FLATNESS
23E6   AC CURRENT

Miscellaneous Symbols:
26B2   NEUTER

Miscellaneous Mathematical Symbols-A
27C7   OR WITH DOT INSIDE
27C8   REVERSE SOLIDUS PRECEDING SUBSET
27C9   SUPERSET PRECEDING SOLIDUS

Miscellaneous Symbols and Arrows:
2B14   SQUARE WITH UPPER RIGHT DIAGONAL HALF BLACK
2B15   SQUARE WITH LOWER LEFT DIAGONAL HALF BLACK
2B16   DIAMOND WITH LEFT HALF BLACK
2B17   DIAMOND WITH RIGHT HALF BLACK
2B18   DIAMOND WITH TOP HALF BLACK
2B19   DIAMOND WITH BOTTOM HALF BLACK
2B20   WHITE PENTAGON
2B21   WHITE HEXAGON
2B22   BLACK HEXAGON
2B23   HORIZONTAL BLACK HEXAGON

Mathematical Alphanumeric Symbols:
1D7CA   MATHEMATICAL BOLD CAPITAL DIGAMMA
1D7CB   MATHEMATICAL BOLD SMALL DIGAMMA
WG2 discussion

T.5 Letterlike Symbols
The US is in favor of adding the following Letterlike symbol as proposed by document L2/04-394 (WG2 document number TBD):
214D   A/S   AKTIESELSKAB
WG2 discussion

T.6 Latin characters
a) The US is in favor of adding the following Latin characters in the Latin Extended-B block as proposed by document WG2 N2847:
0242   LATIN CAPITAL LETTER B WITH STROKE
0243   LATIN CAPITAL LETTER U BAR
0244   LATIN CAPITAL LETTER TURNED V
0245   LATIN CAPITAL LETTER E WITH STROKE
0246   LATIN SMALL LETTER E WITH STROKE
0247   LATIN CAPITAL LETTER J WITH STROKE
0248   LATIN SMALL LETTER J WITH STROKE
0249   LATIN CAPITAL LETTER SMALL Q WITH HOOK TAIL
024A   LATIN SMALL LETTER Q WITH HOOK TAIL
024B   LATIN CAPITAL LETTER R WITH STROKE
024C   LATIN SMALL LETTER R WITH STROKE
024D   LATIN CAPITAL LETTER Y WITH STROKE

024E   LATIN SMALL LETTER Y WITH STROKE
024F   LATIN SMALL LETTER V WITH CURL
This would make the Latin Extended-B collection fixed.
WG2 discussion

b) The US is in favor of creating a new Latin Extended-C block and corresponding collection at the range 2C60-2C7F and insert the following characters as proposed by document L2/04-372 (WG2 document number TBD):
2C60   LATIN CAPITAL LETTER L WITH DOUBLE BAR
2C61   LATIN SMALL LETTER L WITH DOUBLE BAR
2C62   LATIN CAPITAL LETTER L WITH MIDDLE TILDE
2C63   LATIN CAPITAL LETTER P WITH STROKE
2C64   LATIN CAPITAL LETTER R WITH TAIL
WG2 discussion

T.7 Combining Diacritical Marks Supplement
The US is in favor of adding the following Contour tone marks in the Combining Diacritical Marks Supplement block as proposed by document WG2 N2847:
1DC4   COMBINING MACRON-ACUTE
1DC5   COMBINING GRAVE-MACRON
1DC6   COMBINING MACRON-GRAVE
1DC7   COMBINING ACUTE-MACRON
1DC8   COMBINING GRAVE-ACUTE-GRAVE
1DC9   COMBINING ACUTE-GRAVE-ACUTE
1DCA   COMBINING LATIN SMALL LETTER R BELOW
WG2 discussion

T.8 Counting Rod Numerals
The US is in favor of creating a new Counting Rod Numerals block and corresponding collection at the range 1D360-1D37F as proposed by document WG2 N2816:
1D360   COUNTING ROD UNIT DIGIT ONE
1D361   COUNTING ROD UNIT DIGIT TWO
1D362   COUNTING ROD UNIT DIGIT THREE
1D363   COUNTING ROD UNIT DIGIT FOUR
1D364   COUNTING ROD UNIT DIGIT FIVE
1D365   COUNTING ROD UNIT DIGIT SIX
1D366   COUNTING ROD UNIT DIGIT SEVEN
1D367   COUNTING ROD UNIT DIGIT EIGHT
1D368   COUNTING ROD UNIT DIGIT NINE
1D369   COUNTING ROD TENS DIGIT ONE
1D36A   COUNTING ROD TENS DIGIT TWO
1D36B   COUNTING ROD TENS DIGIT THREE
1D36C   COUNTING ROD TENS DIGIT FOUR
1D36D   COUNTING ROD TENS DIGIT FIVE
1D36E   COUNTING ROD TENS DIGIT SIX
1D36F   COUNTING ROD TENS DIGIT SEVEN
1D370   COUNTING ROD TENS DIGIT EIGHT
1D371   COUNTING ROD TENS DIGIT NINE
WG2 discussion

T.9 Modifier Tone Letters
The US is in favor of adding the following characters in the Modifier Tone Letters block as proposed by document L2/04-349 (WG2 document number TBD):
A717   MODIFIER LETTER DOT VERTICAL BAR
A718   MODIFIER LETTER DOT SLASH
A719   MODIFIER LETTER DOT HORIZONTAL BAR
A71A   MODIFIER LETTER LOWER RIGHT CORNER ANGLE
WG2 discussion

T.10 Phags-Pa
The US wants to reaffirm its support for the Phags-Pa as currently documented in PDAM2 and would not welcome significant modification without consensus of all of interested parties .

WG2 discussion
See Chinese comment

Editorial Comments:

E.1 Annex R, Names of Hangul syllables
The information in the linked file (HangulSy.txt) and the tables R.1 to R.4 is redundant. Because the linked file is more complete, the U.S. is asking for the removal of the tables.
WG2 discussion