

Irish comments on PDAM 7 for ISO/IEC 10646:2003

Reference: SC2 N4052

Closes: 2009-03-11

Date: 2009-03-01

Ireland **disapproves** the draft with the technical and editorial comments given below. Acceptance of these comments and appropriate changes to the text will change our vote to approval.

Technical comments

T1. Page 23, Clause 28, Character names and annotations. For “using their five hexadecimal digit value”, read “using their five-hexadecimal-digit value” (using hyphens to connect the adjectival modifier (cf. “a two-year-old horse”).

T2. Page 30-1348, Clause 34, Code Tables and list of character names. If T4 below is accepted, change “18B0-18D1, 18D3-18D7, 18D9-18DA” to “18B0-18C5, 18E0-18F2” in FPDAM 7.

T3. Page 1358, Annex B, List of combining characters. Replace 11 current entries as follows:

For 1BE7 BATAK VOWEL SIGN KEBERETEN read BATAK VOWEL SIGN E

For 1BE8 BATAK VOWEL SIGN KETOLONGEN read BATAK VOWEL SIGN PAKPAK E

For 1BE9 BATAK VOWEL SIGN TALINGA read BATAK VOWEL SIGN EE

For 1BEA BATAK VOWEL SIGN ULUA read BATAK VOWEL SIGN I

For 1BEB BATAK VOWEL SIGN HALUAN read BATAK VOWEL SIGN KARO I

For 1BEC BATAK VOWEL SIGN SIALA ULU read BATAK VOWEL SIGN O

For 1BED BATAK VOWEL SIGN HATULUNGAN read BATAK VOWEL SIGN KARO O

For 1BEE BATAK VOWEL SIGN BORUTA read BATAK VOWEL SIGN U

For 1BEF BATAK VOWEL SIGN HABORITAN FOR SIMALUNGUN SA read BATAK VOWEL SIGN U FOR SIMALUNGUN SA

For 1BF0 BATAK CONSONANT SIGN AMISARA read BATAK CONSONANT SIGN NG

For 1BF1 BATAK CONSONANT SIGN HAJORINGAN read BATAK CONSONANT SIGN H

T4. Page 15, Row 18B: Unified Canadian Aboriginal Syllabics Extended. With reference to ISO/IEC JTC1/SC2/WG2 N3533R “Proposed Revision for Unified Canadian Aboriginal Syllabics Extended Block”, Ireland requests that the characters shown in yellow in PDAM 7 be rearranged as shown in the third of the three charts in that document. Ireland is confident in the technical content of this repertoire and would not oppose the transfer of all of these characters to FDAM 6, but does not insist on such a transfer.

T5. Page 17, Row 1BC: Batak. Ireland reiterates its support for the repertoire and character names as presented in PDAM 7, which is based on the encoding model described in ISO/IEC JTC1/SC2/WG2 N3320R “Proposal for encoding the Batak script in the UCS”. Ireland would oppose a proposal to unify any of the letters here on the basis of superficial glyph resemblance. Ireland would not oppose improvements to the glyphs of U+1BFA and U+1BFB, but notes that these characters, whilst having long glyphs inconvenient for the UCS charts, are nevertheless proper characters, used to indicate the beginnings of texts. Other scripts, like Tibetan, also have characters which serve this function. Other large characters

which have been shrunk to fit the code charts include U+12031 CUNEIFORM AN PLUS NAGA SQUARED. Ireland will investigate the possibility of making the glyphs slightly smaller. We suggest that an editorial note such as “• can be rendered to fill the horizontal length of a column of text” be added. In short, however, Ireland believes that the encoding model and repertoire for Batak as presented is correct and that all of these characters are required for the proper representation of the languages which make use of the Batak script.

T6. Page 31, Row FB5: Arabic Presentation Forms-A. With reference to ISO/IEC JTC1/SC2/WG2 N3575 “Consistent naming and general category of ‘Arabic Pedagogical Symbols’”, Ireland requests that the following character names be used:

FBB2 ARABIC SYMBOL DOT ABOVE
 FBB3 ARABIC SYMBOL DOT BELOW
 FBB4 ARABIC SYMBOL TWO DOTS ABOVE
 FBB5 ARABIC SYMBOL TWO DOTS BELOW
 FBB6 ARABIC SYMBOL THREE DOTS ABOVE
 FBB7 ARABIC SYMBOL THREE DOTS BELOW
 FBB8 ARABIC SYMBOL THREE DOTS POINTING DOWNWARDS ABOVE
 FBB9 ARABIC SYMBOL THREE DOTS POINTING UPWARDS BELOW
 FBBA ARABIC SYMBOL FOUR DOTS ABOVE
 FBBA ARABIC SYMBOL FOUR DOTS BELOW
 FBBC ARABIC SYMBOL TWO DANDAS BELOW
 FBBD ARABIC SYMBOL TWO DOTS VERTICALLY ABOVE
 FBBD ARABIC SYMBOL TWO DOTS VERTICALLY BELOW
 FBBF ARABIC SYMBOL RING
 FBC0 ARABIC SYMBOL SMALL TAH ABOVE
 FBC1 ARABIC SYMBOL SMALL TAH BELOW

As always, Ireland strongly recommends the use of consistent naming conventions wherever possible.

T7. Page 48, Row 1700: Tangut. Ireland requests the removal of Tangut from PDAM 7. We believe that the script is not yet mature for progression to FPDAM. There are several reasons for this.

7.1. The character repertoire is inadequate. It does not take into account the additional characters and new glyphs in Lǐ Fànwén's *Tangut-Chinese Dictionary* published in June 2008 (《夏汉字典》增订版 / 李范文编著。— 北京: 中国社会科学出版社、2008年6月).

7.2. The current character repertoire has no explicit ordering principles, and many characters are clearly misordered. We consider explicit ordering principles to be essential so that users of the code chart can find characters within it. This is particularly important as Tangut glyphs are very similar and there are thousands of them.

7.3. The proposed encoding model for the Tangut repertoire on the ballot follows one Tangut scholar's prescriptive unifications of Tangut characters and expects the use of Variation Selector sequences to distinguish graphically-distinct characters that are used contrastively in the same source. Ireland rejects the proposal to use Variation Selector sequences to make character distinctions in Tangut.

7.4. With reference to ISO/IEC JTC1/SC2/WG2 N3948 “Expert Feedback on the proposed Tangut character set in PDAM 6.2”, it is clear that the user community of experts does not

believe that the character repertoire (which was on PDAM 6.2 and now unchanged on PDAM 7) script is mature for encoding and does not agree that its tacit encoding conventions (including the use of Variation Selector sequences) are sufficient for their needs.

7.5. WG2 made provision for progressing Tangut. In ISO/IEC JTC1/SC2/WG2 N3541, the “Report from the Ad Hoc on Tangut” from Meeting 53 in Hong Kong:

“China, Ireland, the UK and the US have agreed to come up with a common position before the next WG2 meeting. A resolution should be made to request these national bodies to work together for a documented common position on Tangut, and also to invite other interested national bodies to participate in this work. It is understood that if an agreed-upon position cannot be reached before the Dublin meeting, Tangut will be removed from Amendment 7.”

It does not seem that it will be possible to get a comprehensive agreed-upon position between all four of these National Bodies before the Dublin meeting, and therefore Tangut must be removed from Amendment 7 as agreed. However, Ireland proposes in T7.6 below a way forward.

7.6. With reference to document N3577 “Proposal for a revised Tangut character set for encoding in the SMP of the UCS”, Ireland believes that the character repertoire and encoding model proposed succeeds in meeting the requirements of the user community. We believe that the encoded character repertoire needs to be able to represent all of the graphically-distinct characters found in all the major modern Tangut dictionaries. This is a cornerstone of Tangutology. Accordingly, Ireland requests that the repertoire in N3577 be balloted in a new PDAM 8. We believe that the set requires two rounds of balloting, and would oppose placement of this large, reorganized repertoire in FPDAM 7.

Editorial comments

E1. Amendment text.

Delete “Tangut, ” from the title of the Amendment.

For “Page 1, Clause 1 Scope”, read “Page 1, Clause 1, Scope”.

For “Page 2, Clause 3 Normative references”, read “Page 2, Clause 3, Normative references”.

For “Page 2, Clause 4 Terms and definitions”, read “Page 2, Clause 4, Terms and definitions”.

For “Page 4, Clause 5 General structure of the UCS”, read “Page 4, Clause 5, General structure of the UCS”.

For “Page 8, Sub-clause 6.4 Naming of characters”, read “Page 8, Sub-clause 6.4, Naming of characters”.

For “Page 14, Sub-clause 20.3 Format characters”, read “Page 14, Sub-clause, 20.3 Format characters”.

For “Page 23, Clause 28 Character names and annotations”, read “Page 23, Clause 28, Character names and annotations”.

For “Page 25, Clause 29 Named UCS Sequence Identifiers”, read “Page 25, Clause 29, Named UCS Sequence Identifiers”.

For “Page 1351, annex A.1”, read “Page 1351, Annex A.1”.

For “Page 1358, Annex B List of combining characters”, read “Page 1358, Annex B, List of combining characters”.

For “Page 1379, Annex F.2 Script-specific format characters”, read “Page 1379, Annex F.2, Script-specific format characters”.

E2. **Page 7, Row 050: Cyrillic Supplement.** Ireland suggests that that the following notes be used for consistency:

Letters for Enets and Khanty

0510 CYRILLIC CAPITAL LETTER REVERSED ZE

..

Letters for Chukchi, Itelman, and Khanty

0512 CYRILLIC CAPITAL LETTER EL WITH HOOK

..

Letters for Aleut

051E CYRILLIC CAPITAL LETTER ALEUT KA

..

Letters for Abkhaz

0524 CYRILLIC CAPITAL LETTER PE WITH DESCENDER

0525 CYRILLIC SMALL LETTER PE WITH DESCENDER

- used in modern Abkhaz orthography
- 04A7 cyrillic small letter pe with hook



E3. **Page 12, Row 0D0: Malayalam.** Ireland recommends the replacement of the rather inconsistent font used for Malayalam with the table shown below.

	0D0	0D1	0D2	0D3	0D4	0D5	0D6	0D7
0		ഐ 0D10	ഠ 0D20	ര 0D30	ീ 0D40		ജ 0D60	ധ 0D70
1			ഡ 0D21	റ 0D31	ൂ 0D41		ഞ 0D61	ന 0D71
2	ം 0D02	ഒ 0D12	ഌ 0D22	ല 0D32	ു 0D42		ഊ 0D62	നം 0D72
3	ഃ 0D03	ഓ 0D13	ണ 0D23	ള 0D33	്യ 0D43		ൌ 0D63	ഠ 0D73
4		ഔ 0D14	ത 0D24	ഴ 0D34	് 0D44			ൽ 0D74
5	അ 0D05	ക 0D15	ഥ 0D25	വ 0D35				ൻ 0D75
6	ആ 0D06	ഖ 0D16	ഭ 0D26	ശ 0D36	െ 0D46		ഠ 0D66	
7	ഇ 0D07	ഗ 0D17	ധ 0D27	ഷ 0D37	േ 0D47	ൗ 0D57	റ 0D67	
8	ഈ 0D08	ഘ 0D18	ന 0D28	സ 0D38	ൈ 0D48		റ 0D68	
9	ഉ 0D09	ങ 0D19	ണ 0D29	ഹ 0D39			ന 0D69	ന 0D79
A	ഊ 0D0A	ച 0D1A	പ 0D2A	റ്റി 0D3A	ൊ 0D4A		ർ 0D6A	ൻ 0D7A
B	ഋ 0D0B	ഘ 0D1B	ഫ 0D2B		ോ 0D4B		രി 0D6B	ൻ 0D7B
C	ൠ 0D0C	ജ 0D1C	ബ 0D2C		ൌ 0D4C		നു 0D6C	ർ 0D7C
D		ത്വ 0D1D	ഭ 0D2D	് 0D3D	് 0D4D		യ 0D6D	ൽ 0D7D
E	എ 0D0E	ണ 0D1E	മ 0D2E	ാ 0D3E			വ 0D6E	ൾ 0D7E
F	ഏ 0D0F	ട 0D1F	യ 0D2F	ി 0D3F			ൻ 0D6F	ക 0D7F

Based on ISCII 1988**Various signs**

0D02 െ MALAYALAM SIGN ANUSVARA
 0D03 െ MALAYALAM SIGN VISARGA

Independent vowels

0D05 അ MALAYALAM LETTER A
 0D06 ആ MALAYALAM LETTER AA
 0D07 ഇ MALAYALAM LETTER I
 0D08 ഇു MALAYALAM LETTER II
 0D09 ഉ MALAYALAM LETTER U
 0D0A ഉു MALAYALAM LETTER UU
 0D0B ഋ MALAYALAM LETTER VOCALIC R
 0D0C ൠ MALAYALAM LETTER VOCALIC L
 0D0D  <reserved>
 0D0E എ MALAYALAM LETTER E
 0D0F ഐ MALAYALAM LETTER EE
 0D10 ഐ MALAYALAM LETTER AI
 0D11  <reserved>
 0D12 ഒ MALAYALAM LETTER O
 0D13 ഓ MALAYALAM LETTER OO
 0D14 ഔ MALAYALAM LETTER AU

Consonants

Alternate romanizations are shown as aliases for some letters to clarify their identity.


0D15 ക MALAYALAM LETTER KA
 0D16 ഖ MALAYALAM LETTER KHA
 0D17 ഗ MALAYALAM LETTER GA
 0D18 ഘ MALAYALAM LETTER GHGA
 0D19 ങ MALAYALAM LETTER NGA
 0D1A ച MALAYALAM LETTER CA
 = cha
 0D1B ഛ MALAYALAM LETTER CHA
 = chha
 0D1C ജ MALAYALAM LETTER JA
 0D1D ഝ MALAYALAM LETTER JHA
 0D1E ഞ MALAYALAM LETTER NYA
 = nha
 0D1F ട MALAYALAM LETTER TTA
 = ta
 0D20 ഡ MALAYALAM LETTER TTHA
 = tta
 0D21 ഢ MALAYALAM LETTER DDA
 = hard da
 0D22 ണ MALAYALAM LETTER DDHA
 = hard dda
 0D23 ഹ MALAYALAM LETTER NNA
 = hard na
 0D24 ത MALAYALAM LETTER TA
 = tha
 0D25 ഡ MALAYALAM LETTER THA
 = ttha
 0D26 ഡ MALAYALAM LETTER DA
 = soft da
 0D27 ഡ MALAYALAM LETTER DHA
 = soft dda
 0D28 ന MALAYALAM LETTER NA
 0D29 ഹ MALAYALAM LETTER NNA
 0D2A പ MALAYALAM LETTER PA

0D2B ഫ MALAYALAM LETTER PHA
 0D2C ബ MALAYALAM LETTER BA
 0D2D ഭ MALAYALAM LETTER BHA
 0D2E മ MALAYALAM LETTER MA
 0D2F യ MALAYALAM LETTER YA
 0D30 ര MALAYALAM LETTER RA
 0D31 റ MALAYALAM LETTER RRA
 0D32 ല MALAYALAM LETTER LA
 0D33 ള MALAYALAM LETTER LLA
 0D34 ഴ MALAYALAM LETTER LLLA
 = zha
 0D35 വ MALAYALAM LETTER VA
 0D36 ശ MALAYALAM LETTER SHA
 = soft sha
 0D37 ഷ MALAYALAM LETTER SSA
 = sha
 0D38 സ MALAYALAM LETTER SA
 0D39 ഹ MALAYALAM LETTER HA
 0D3A ഴ MALAYALAM LETTER TTTA

Addition for Sanskrit

0D3D ഴ MALAYALAM SIGN AVAGRAHA
 = praslesham

Dependent vowel signs




0D3E െ MALAYALAM VOWEL SIGN AA
 0D3F െ MALAYALAM VOWEL SIGN I
 0D40 െ MALAYALAM VOWEL SIGN II
 0D41 െ MALAYALAM VOWEL SIGN U
 0D42 െ MALAYALAM VOWEL SIGN UU
 0D43 െ MALAYALAM VOWEL SIGN VOCALIC R
 0D44 െ MALAYALAM VOWEL SIGN VOCALIC RR
 0D45  <reserved>
 0D46 െ MALAYALAM VOWEL SIGN E
 • stands to the left of the consonant
 0D47 െ MALAYALAM VOWEL SIGN EE
 • stands to the left of the consonant
 0D48 െ MALAYALAM VOWEL SIGN AI
 • stands to the left of the consonant

Two-part dependent vowel signs

These vowel signs have glyph pieces which stand on both sides of the consonant; they follow the consonant in logical order, and should be handled as a unit for most processing.

0D4A െ MALAYALAM VOWEL SIGN O
 ≡ 0D46 െ 0D3E െ
 0D4B െ MALAYALAM VOWEL SIGN OO
 ≡ 0D47 െ 0D3E െ
 0D4C െ MALAYALAM VOWEL SIGN AU
 • archaic form of the /au/ dependent vowel
 → 0D57 െ malayalam au length mark
 ≡ 0D46 െ 0D57 െ

Various signs

0D4D െ MALAYALAM SIGN VIRAMA
 = chandrakkala (the preferred name)
 = vowel half-u
 0D4E  <reserved>
 0D4F  <reserved>
 0D50  <reserved>

0D51	☐	<reserved>
0D52	☐	<reserved>
0D53	☐	<reserved>
0D54	☐	<reserved>
0D55	☐	<reserved>
0D56	☐	<reserved>
0D57	ഠ	MALAYALAM AU LENGTH MARK
		• used alone to write the /au/ dependent vowel in modern texts
		→ 0D4Cഠ malayalam vowel sign au

0D7D	ൽ	MALAYALAM LETTER CHILLU L
		• historically derived from the full letter ta
		• used for chillu t and chillu d
0D7E	ൾ	MALAYALAM LETTER CHILLU LL
0D7F	ക	MALAYALAM LETTER CHILLU K

Additional vowels for Sanskrit

0D60	ഠ	MALAYALAM LETTER VOCALIC RR
0D61	ഠ	MALAYALAM LETTER VOCALIC LL

Dependent vowels

0D62	ഠ	MALAYALAM VOWEL SIGN VOCALIC L
0D63	ഠ	MALAYALAM VOWEL SIGN VOCALIC LL

Reserved

For viram punctuation, use the generic Indic 0964 and 0965.

0D64	☐	<reserved>
		→ 0964 devanagari danda
0D65	☐	<reserved>
		→ 0965 devanagari double danda

Digits

0D66	൦	MALAYALAM DIGIT ZERO
0D67	൧	MALAYALAM DIGIT ONE
0D68	൨	MALAYALAM DIGIT TWO
0D69	൩	MALAYALAM DIGIT THREE
0D6A	൪	MALAYALAM DIGIT FOUR
0D6B	൫	MALAYALAM DIGIT FIVE
0D6C	൬	MALAYALAM DIGIT SIX
0D6D	൭	MALAYALAM DIGIT SEVEN
0D6E	൮	MALAYALAM DIGIT EIGHT
0D6F	൯	MALAYALAM DIGIT NINE

Malayalam numerics

0D70	൧൦	MALAYALAM NUMBER TEN
0D71	൧൦൦	MALAYALAM NUMBER ONE HUNDRED
0D72	൧൦൦൦	MALAYALAM NUMBER ONE THOUSAND

Fractions

0D73	൧൦൦	MALAYALAM FRACTION ONE QUARTER
0D74	൧൦൦	MALAYALAM FRACTION ONE HALF
0D75	൧൦൦	MALAYALAM FRACTION THREE QUARTERS

Date mark

0D79	൧൦൦	MALAYALAM DATE MARK
------	-----	---------------------

Chillu letters

0D7A	൧൦൦	MALAYALAM LETTER CHILLU NN
0D7B	൧൦൦	MALAYALAM LETTER CHILLU N
0D7C	൧൦൦	MALAYALAM LETTER CHILLU RR
		• historically derived from the full letter ra
		• also used for chillu r