

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

Doc. Type: Disposition of comments

Title: Disposition of comments on PDAM2.3 to ISO/IEC 10646 5th edition

Source: Michel Suignard (project editor)

Project: JTC1 02.10646.00.02.00.05

Status: For review by WG2

Date: 2018-06-18

Distribution: WG2

Reference: SC2 N4608 WG2

Medium: Paper, PDF file

Comments were received from the following members: Ireland, Japan, and USA. The following document is the disposition of those comments. The disposition is organized per country.

Note – With some minor exceptions, the full content of the ballot comments has 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*.

The following items reflect the changes to Amendment 2 between Pdam2.3 and the planned Dam2 based on the ballot disposition:

Block moved

ELYMAIC block moved from U+10EC0..10EDF to U+10FE0..10FFF

Additions

@ Wancho Tone marks

UCS Name

1E2EC WANCHO TONE TUP

1E2ED WANCHO TONE TUP MANG

1E2EE WANCHO TONE OKOI

1E2EF WANCHO TONE OKOI MANG

Name changes

UCS Old name

New name

2E4F MEDIEVAL CORNISH VERSE SEPARATOR

CORNISH VERSE DIVIDER

Glyph changes through this ballot (pending receiving fonts with updated glyphs)

A7C2 LATIN CAPITAL LETTER ANGLICANA W

A7C3 LATIN SMALL LETTER ANGLICANA W

13430 EGYPTIAN HIEROGLYPH VERTICAL JOINER

13431 EGYPTIAN HIEROGLYPH HORIZONTAL JOINER

13432 EGYPTIAN HIEROGLYPH START AT TOP

13433 EGYPTIAN HIEROGLYPH START AT BOTTOM

13434 EGYPTIAN HIEROGLYPH END AT TOP

13435 EGYPTIAN HIEROGLYPH END AT BOTTOM

13436 EGYPTIAN HIEROGLYPH OVERLAY MIDDLE

13437 EGYPTIAN HIEROGLYPH BEGIN SEGMENT

13438 EGYPTIAN HIEROGLYPH END SEGMENT

1F995 SAUROPOD

1F996 T-REX

1F998 KANGAROO

1F99B HIPPOPOTAMUS

1FF9C PARROT

1FF9D RACCOON

1FF9E LOBSTER

1F9B0 EMOJI COMPONENT RED HAIR

1F9B1 EMOJI COMPONENT CURLY HAIR

1F9B2 EMOJI COMPONENT BALD

1F9B3 EMOJI COMPONENT WHITE HAIR

1F9B8 SUPERHERO

1F9B9 SUPERVILLAIN

1F9E7 RED GIFT ENVELOPE

Annotations, cross reference, source reference added/modified

A7BD LATIN SMALL LETTER GLOTTAL I (xref to A723 and A725)

Ireland: Negative

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 27, Row 2E0: Supplemental Punctuation

Ireland approves the proposed name change of 2E4F DOUBLE OBLIQUE HYPHEN WITH FALLING DOTS, but suggests that the name MEDIEVAL CORNISH VERSE SEPARATOR is not quite right. The medieval period in Europe is usually considered to range from the 5th to the 15th centuries. The stages of the Cornish language are generally divided into Old Cornish (800–1200), Middle Cornish (1200–1600), Late Cornish (1578–1800), and Modern or Neo-Cornish (1904–present). This character is found in a range of manuscripts from the 12th to 16th centuries and is a part of the unique Cornish scribal tradition. The name CORNISH VERSE SEPARATOR would be sufficient to identify the character. Note 2CFC COPTIC OLD NUBIAN VERSE DIVIDER (where Old Nubian describes a language stage) and 060E ARABIC POETIC VERSE SIGN.

Perhaps the name CORNISH VERSE DIVIDER would be best for consistency with the existing character name of 2CFC.

Accepted in principle

The proposed new name is CORNISH VERSE DIVIDER.

T2. Page 29, Row A72: Latin Extended-D

Ireland affirms its strong support for the encoding of A7C0 LATIN CAPITAL LETTER THORN WITH DIAGONAL STROKE and ACC1 LATIN SMALL LETTER THORN WITH DIAGONAL STROKE. Ireland notes that other Latin letters, such as 0244 Ꝁ LATIN CAPITAL LETTER U WITH STROKE and 0289 ꝁ LATIN SMALL LETTER U BAR alongside A7B8 Ꝑ LATIN CAPITAL LETTER U WITH DIAGONAL STROKE and A7B9 ꝑ LATIN SMALL LETTER U WITH DIAGONAL STROKE, are similarly distinguished for use in various languages. Opentype features are not used to distinguish Ꝁ from Ꝑ. The Nordicist tradition is properly served with A7C4 LATIN CAPITAL LETTER THORN WITH STROKE and A7C5 LATIN SMALL LETTER THORN WITH STROKE, but the centuries-old Anglicist tradition (both in manuscript and set in type) would remain unsupported without these two characters. The Anglicist letters are *not* glyph variants of the Nordicist letters.

Þ ≠ Ꝑ
þ ≠ ꝑ

Ireland notes that ISO/IEC JTC1/SC2/WG2 N4836R (L2/17-236R) “Proposal to add LATIN LETTER THORN WITH DIAGONAL STROKE” is a 48-page document giving argument and examples of centuries of typographic tradition in which þ is used regularly in Anglicist contexts and ꝑ regularly in Nordicist contexts. The US NB comment requesting that this character be removed from PDAM 2.2 was a simple unsupported assertion that “they are glyph variants”. They are not, and the US NB did not even make a case, much less prove it.

Ireland does not believe that the editor had good grounds to remove the characters from PDAM 2.3. The characters must be put back in the ballot. Medievalists use specialized fonts, but variation selectors or Opentype features are not the way in which characters like this are differentiated. Compare also 1E9C LATIN SMALL LETTER LONG S WITH DIAGONAL STROKE and 1E9D LATIN SMALL LETTER LONG S WITH HIGH STROKE.

Noted

The aim of PDAM2.3 was to have a non-controversial repertoire, allowing its content to reach committee consensus and therefore be a good candidate for enquiry process (DAM). This is why these 2 characters were

removed from Amendment 2 (they were controversial). They can still be maintained in a future repertoire subject to committee level ballot like the CD for the 6th edition of ISO/IEC 10646.

T3. Page 55, Row 16FE: Ideographic Symbols and Punctuation

With regard to JTC1/SC2/WG2 N4943R (L2/18-121R) “Cluster Formation Model for Khitan Small Script”, Ireland believes that if the Small Khitan Script is returned to this Amendment (which we favour) and if it is decided that 034F COMBINING GRAPHEME JOINER should be replaced by a script-specific control, then that character should either go in this block, or it should be the first character in the KSS block, in order to facilitate the sorting behaviour described at the bottom of page 9 of N4943RI..

Noted

Concerning a script-specific control, there is no need to be in the same block. Many other scripts do otherwise. It is true that it would be convenient to have a code point value less than any other Khitan Small Scripts code point which could be achieved by using a code point in a block preceding the KSS block (for example the Ideographic Symbols and Punctuation at 16FE0-16FF as suggested by Ireland..

It is also not clear why the KSS Iteration Mark is not in that block like for Tangut and Nushu. It would make the KSS block content homogenous.

T4. Page 89, Row 1E2C: Wancho

Ireland affirms its support for the four combining characters proposed to be encoded at 1E2EC..1E2CF. These characters which were devised by the script’s inventor during the encoding and are currently being used to prepare lexicographical work for new publications. There would be no benefit to the UCS or to the users of the script to put off encoding these characters, and they should be returned to the Amendment. The UT and SC2 should trust the encoding process for developing writing systems, when the script inventor is involved with the development and encoding process.

Noted

Printed evidence of these 4 tone marks is provided in document N4976. Based on this, these four characters can be added to the Wancho repertoire in Amendment 2.

@	Tone marks
1E2EC	WANCHO TONE TUP
1E2ED	WANCHO TONE TUP MANG
1E2EE	WANCHO TONE OKOI
1E2EF	WANCHO TONE OKOI MANG.

T5. Page 100, Row 1F90: Supplemental Symbols and Pictographs

Ireland affirms its support for the encoding of BILLIARD GAMES at 1F93F, since (nearly) every vendor has changed the glyph of 1F3B1 BILLIARDS to that of the divination toy EIGHT BALL. It would be wrongly inconsistent for the black and white glyph for that character to differ from the form used by vendors for emoji purposes. It may be the case that a vendor has not followed through with the unfortunate innovation of other vendors, but all should be encouraged to harmonize with one another.



1F93F BILLIARD GAMES

→ 1F3B1  billiards.

Noted

A separated proposal should be made to disunify a symbol which is currently ambiguous.

T6. Page 100, Row 1F90: Supplemental Symbols and Pictographs

Ireland affirms its strong support for the encoding of 1F98B TROLL. It would be inappropriate to unify the European troll (also used metaphorically for “internet trolls”) with the 1F479 JAPANESE OGRE (oni or namahage) and 1F47A JAPANESE GOBLIN (tengu) which have no such metaphorical use, and whose semantics are explicitly

defined by the word JAPANESE. The semantics of pictographic characters are NOT irrelevant. Moreover, the statement in the US NB's ballot comments on PDAM 2.2, that "emoji size" should be taken into consideration, is irrelevant. Emoji is one usage scenario for pictographic characters and dingbats. It is not the only usage scenario. Moreover, the UCS and ISO/IEC 10646 has no defined concept of "emoji". Pictographic characters may be used as emojis, but they are still pictographic characters. Standardizers may take note that the use of some characters as emoji we enjoy today may not persist in the decades and centuries to come. The universality of the UCS must not be forgotten.



1F97B TROLL

Noted

This character is not part of Pdam 2.3 but can be discussed in the context of a future amendment.

T7. Page 100, Row 1F90: Supplemental Symbols and Pictographs

Ireland affirms its strong support for the encoding of 1F9A3 MAMMOTH, 1F9A4 DODO, and 1F9A5 SQUIRREL. In particular, the attempt to unify squirrel with CHIPMUNK is an inappropriate over-extension of squirrel for the source character CHIPMUNK which was encoded for compatibility with the Windows webdings character set. The suggestion by the US NB that DODO and MAMMOTH "do not meet the selection requirements" for emoji is false. The Emoji Subcommittee's internal procedures may assist them in their work; they are not binding on SC2 National Bodies.



1F9A3 MAMMOTH

- indicates great size



1F9A4 DODO

- indicates extinction



1F9A5 SQUIRREL

Noted

These characters are not part of Pdam2.3. The cases for encoding MAMMOTH and DODO are reasonable and these proposal can be entertained for future additions. However the case for SQUIRREL is weaker as current vendor implementation show more often a Squirrel for the code point attributed to the Chipmunk.

T8. PDAM 2.3

Ireland continues to object to the UTC's publication of the pictographic characters on this PDAM while those characters are under technical ballot. This effectively invalidates the vote of every National Body participating in SC2 with regard to the technical content of ISO/IEC 10646. SC2 has not been asked, and has not agreed, to rubber-stamp whatever characters the UTC's Emoji Subcommittee decides to invent in order to help vendors sell their telephones and other devices. That Subcommittee keeps making blunders and mistakes, creating gaps in the set of pictographic characters encoded in the UCS. Oversight is required. We have suggested many times that a fair price for the UTC to pay for the dumping of 50 or 60 characters a year into the UCS for the purposes of emoji would be to allow the addition of a smaller set of characters, like dinosaurs, pepper mills, animal faces, and so on, to make the set of UCS pictographs generally useful for everyone, not just those exchanging colourful emoji characters in text messages and the like. The Emoji Subcom - mittee may be satisfied with using Google frequency counts and other heuristics as a "scientific" tool to predict what symbols might be popular as emojis, but in effect they are simply using a magic 8-ball to do it. Vendors do not own the UCS. National Bodies have their own expertise as well, and the UTC is effectively preventing National Bodies from having a say in the population of the UCS with symbols and pictographs.

Ireland very much regrets the tension which has resulted between the two committees due to the lack of agreement on this issue. Cooperation goes both ways. It would be preferable to "duelling ballot comments".

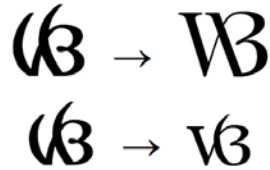
Noted

Ballot comment may be not the optimal place to make such comment.

Editorial comments:

E1. Page 29, Row A72: Latin Extended-D

After consultation with expert typographers about the appropriate design of the Anglicana W, Ireland requests the following glyph changes for A7C2 and A7C3 respectively. The Anglicana W itself has a number of glyph variants in the manuscript tradition, and the typographic forms shown on the right below are both acceptable within that tradition and offer a modern look which harmonizes well in text.



Accepted in principle

Same comment was made in ballot for Pdam2.2 (WG2 N4940). It will be fixed, based on receiving an updated font from Ireland.

E2. Page 32, Row A72: Latin Extended-D

Ireland agrees with the name change of both A7BC and A7BD for consistency with the other letters for the spiritus lenis diacritical mark, but proposes the change of the annotation of A7BD from “also used in Egyptology as yod” to the following:

A7BD LATIN SMALL LETTER GLOTTAL I

= egyptological yod

x A723 latin small letter egyptological alef

x A725 latin small letter egyptological ayin.

Accepted

E3. Page 51, Row 1343: Egyptian Hieroglyphs Format Controls

Ireland remains convinced that the glyphs for these characters must be revised. Control characters in the UCS typically have identifying letter abbreviations in their dotted-box glyphs. The exception to this is the ideographic Description Characters, which are intended to be “visibly displayed graphic characters, not invisible composition controls” as the annotation to the code chart states. The glyphs of the Egyptian characters in this PDAM are hardly distinguishable from those of the Ideographic Description Characters, and we believe this could be confusing to users and implementers. After consultation with expert Andrew Glass, we propose that the glyphs below be used, as they conform well to the general principles used in the UCS for control characters. We are pleased to note the editor’s agreement that confusion with Ideographic Description characters should be avoided, but believe the best way to achieve this is to incorporate letters into the glyph design as is common even if not universal for control characters of this kind. We have heard no genuine reason why this should not be done.

EVJ ⋮	13430 EGYPTIAN HIEROGLYPH VERTICAL JOINER
EHJ ✱	13431 EGYPTIAN HIEROGLYPH HORIZONTAL JOINER
EST ⌋	13432 EGYPTIAN HIEROGLYPH START AT TOP
ESB ⌋	13433 EGYPTIAN HIEROGLYPH START AT BOTTOM
EET ⌋	13434 EGYPTIAN HIEROGLYPH END AT TOP
EEB ⌋	13435 EGYPTIAN HIEROGLYPH END AT BOTTOM
EOM ⊕	13436 EGYPTIAN HIEROGLYPH OVERLAY MIDDLE
EBS (13437 EGYPTIAN HIEROGLYPH BEGIN SEGMENT
EES)	13438 EGYPTIAN HIEROGLYPH END SEGMENT

Accepted in principle

There is no requirement to have letter abbreviations in glyphs used for control characters, in fact many do not, and the only typical requirement is to use a dotted rectangular box. Therefore there is no general principle in the UCS to have such abbreviations. In this case, the main information is to convey the type of layout adjustment related to a given control character. The acronyms brings very little practical information and also impose to make the embedded pictographs much smaller. It is preferable to make these pictographs use as much of the content box as possible. However, the editor is open to improved glyphs.

For reference, the current glyphs are shown here:

13430	⋮	EGYPTIAN HIEROGLYPH VERTICAL JOINER
13431	✱	EGYPTIAN HIEROGLYPH HORIZONTAL JOINER
13432	⌋	EGYPTIAN HIEROGLYPH START AT TOP
13433	⌋	EGYPTIAN HIEROGLYPH START AT BOTTOM
13434	⌋	EGYPTIAN HIEROGLYPH END AT TOP
13435	⌋	EGYPTIAN HIEROGLYPH END AT BOTTOM
13436	⊕	EGYPTIAN HIEROGLYPH OVERLAY MIDDLE
13437	(EGYPTIAN HIEROGLYPH BEGIN SEGMENT
13438)	EGYPTIAN HIEROGLYPH END SEGMENT

E4. Page 115, Row 1F90: Supplemental Symbols and Pictographs

Ireland recommends that the following glyphs be used for the reference glyphs (basically white outlined rather than black silhouettes), for consistency with other glyphs in the standard:



1F995 SAUROPOD



1F996 T-REX



1F998 KANGAROO



1F99B HIPPOPOTAMUS



1F99C PARROT



1F99D RACCOON



1F99E LOBSTER

Accepted in principle

Based on receiving a font with the updated glyphs.

E5. Page 100, Row 1F90: Supplemental Symbols and Pictographs

Ireland believes that the glyphs for 1F9B8 SUPERHERO and 1F9B9 SUPERVILLAIN should be replaced with more anthropomorphic figures. The glyphs below are based on public-domain heroes; specific attributes of costuming can be altered, but the use on the PDAM of “caped smiley faces” makes no sense, so we recommend caped isotype figures as shown below.



1F9B8 SUPERHERO



1F9B9 SUPERVILLAIN

Accepted in principle

The current glyphs could definitively be improved. However it is not clear that the new glyphs are distinctive enough.

E6. Page 100, Row 1F90: Supplemental Symbols and Pictographs

Ireland believes that the glyph for 1F9E7 RED GIFT ENVELOPE should have a closer vertical hatching as other red-coloured glyphs in the standard do.



1F9E7 RED GIFT ENVELOPE

Accepted in principle

Based on receiving a font with the updated glyph..

E7. Page 100, Row 1F90: Supplemental Symbols and Pictographs

Ireland believes that the glyphs for the hair colour swatches should have the same structure as the EMOJI MODIFIER FITZPATRICK TYPE characters which serve a similar function. At present these are essentially images of human scalps, for which there is no reasonable use scenario that isn't rather gruesome. The glyphs below have the same wiggly box that the Fitzpatrick modifiers do, as well as person images which harmonize with the reference glyphs of human pictographs.



1F9B0 EMOJI COMPONENT RED HAIR



1F9B1 EMOJI COMPONENT CURLY HAIR



1F9B2 EMOJI COMPONENT BALD



1F9B3 EMOJI COMPONENT WHITE HAIR

Accepted in principle

These characters are not modifiers and should not have frames in their representative glyphs. However the glyph can be improved, based on receiving a font with the updated glyphs.

As a result of these dispositions, Ireland changed its vote to YES

Japan: Positive with comments

Technical comment:

TE.1. Page 29 23.01 List of source references

No CJK ideographs are found in this standard having “GHR”, “T13” and “UK” source references. It seems these entries had been added for the characters on CJK Unified Ideograph Extension-G at pdam 2.2..

Proposed change by Japan:

Delete the description for “GHR”, “T13” and “UK” source references.

Accepted

TE.2. Page 33 Table 5: Format details of the tags used in the source reference file for CJK ideographs

Same as TE.1 comment, no CJK ideographs are found in this standard having “GHR”, “T13” and “UK” source references on this table.

Proposed change by Japan:

Delete “GHR-dddddd.dd” “T13-hhhh” and “UK-ddddd” on “Third field format” column.

Accepted

USA: Positive with comments

Comments (GE for General, TE for Technical, ED for Editorial):

TE.1. Elymaic

The USNB requests Elymaic be moved from its current location, U+10EC0..U+10EDF, to U+10FE0..U+10FFF in order to be in sync with the latest Roadmap.

Proposed change by USA:

We recommend Elymaic be moved to U+10FE0..U+10FFF.

Accepted

ED.2. Latin Extended-D

The subhead above U+A7C4 is currently “Letters used in early pinyin romanization” It should be “Letters used in early Pinyin romanization”.

Proposed change by USA:

Capitalize “Pinyin”.

Accepted

ED.3. Page 3

On page 3, *Whole blocks* should precede Khitan Small Script, Shuishu Logograms, and Shuishu Radicals.

Proposed change by USA:

Move “Whole blocks” to the appropriate place before Khitan Small Script, etc.

Accepted

Although all the Editor’s note section will disappear in DAM2 anyway.