Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type: **Working Group Document**
Title: **Future Additions to ISO/IEC 10646**
Source: **Irish National Body**
Status: **National Body Contribution**
Action: **For consideration by JTC1/SC2/WG2 and UTC**
Date: **2017-09-24**

The Irish National Body notes with regret that relations between ISO/IEC JTC1/SC2 and the Unicode Technical Committee have deteriorated in the last couple of years. We have made some of our concerns available in our ballot comments to both PDAM 1.2 (ISO/IEC JTC1/SC2 N4518) and PDAM 1.3 (ISO/IEC JTC1/SC2 N4546), and wish to reiterate some of those comments here in order to help encourage a re-set of the relationship between the two committees.

The UTC's recent formal publication of a number of graphic characters while still under technical ballot has, in our view, endangered the cooperation between our committees and presented difficulties with regard to synchronization. SC2 has been presented with a *fait accompli* because the UTC has essentially fast-tracked a number of characters to publication. But SC2 and UTC have no agreement that either committee can simply publish characters without the permission of the other. We have an agreement to fast-track only one kind of character: currency symbols. And the currency symbols which our two committees have agreed to fast-track have always (and only) been those devised by governmental authorities (such as Ukraine, Kazakhstan, India, Armenia, Georgia, and Russia).

SC2 has never agreed to fast-track graphic symbol characters invented for potential use. The UTC has never asked SC2 to do make such an agreement. Recently the Fitzpatrick symbols were fast-tracked, but only because the UTC representatives at the WG2 meeting in Sri Lanka made a special case for doing so—in order to solve a very serious public perception problem and to cater to the needs of many users who wished to use a number of graphic characters as a means of self-expression. The case was made, properly, and the characters were expedited to publication.

No such case was made for T-REX or SAUROPOD or MERPERSON or for the change of semantics made to BILLIARDS. The specifics of our objection to the name "T-REX" and the BILLIARDS issue are given in our ballot comments to PDAM 1.3. Our technical comments on the two dinosaur characters are relevant to this discussion, and we repeat a part of these here:

T6. **Page 162, Row 1F90: Supplemental Symbols and Pictographs.** In our comments to PDAM 1.2, Ireland objected to the encoding of 1F995 SAUROPOD and 1F996 T-REX as presented there. We understand that these have already been

published by the Unicode Consortium—without National Bodies having ever even seen a proposal for either—and that they are, with their names, an evident fait accompli. However, in terms of the processes which SC2 and the UTC make use of as we jointly administer a single character set, Ireland must point out that SC2 National Bodies never entered into an agreement with the UTC that we would rubberstamp everything—or anything—proposed by the UTC. In particular, we did not do so when the first large group of emoji were proposed for the standard on the basis of compatibility with existing Japanese character sets. SC2 treated the emoji repertoire as what they are—a collection of symbols—and National Bodies insisted on additions, deletions, and alterations. This application of SC2 National Body expertise was accepted by the UTC as a reasonable response to adventitious repertoire additions without a clear prior-usage rationale. The resulting additions, deletions, and alterations improved the UCS and were the product of effective and collegial collaboration between our two committees. Such collaboration must be encouraged. Currently the mood is more adversarial, to no good purpose.

Now that the Consortium has adopted special procedures for the *invention* of new symbol characters, it is clear that SC2 oversight is even more important than it has been in the past. The UTC has been permitting vendors to alter the semantics of symbols, rather than encouraging a uniform treatment of them (as with BILLIARDS, or with ANIMAL vs ANIMAL FACE mentioned above); the UTC has devised inconsistent methods for indicating gender and coloration, and has been devising symbol characters with little real rationale in terms of completeness of sets or even in terms of prediction of genuinely likely utility (given the way that text-frequency citations are prioritized over more semantic or metaphorical concerns. In our view, SC2's oversight and input is necessary in order to ensure the Emoji Subcommittee and the UTC do a better job than they have been doing. It should be pointed out that the guidelines on the Unicode website "Submitting Emoji Proposals" provides input to the Emoji Subcommittee. That Subcommittee is composed of experts at large, and has nothing to do with experts appointed by National Bodies to SC2 and SC2/WG2, and the procedures of that Subcommittee have neither been reviewed nor adopted by SC2. We are not bound by them. We have, in the past, augmented Emoji proposals by additions, deletions, and alterations, adding new symbol characters via ballot comments. This procedure was satisfactory and beneficial in the past. We apply it here per that precedent. We expect serious discussion and contributions from the UTC should they feel that this balance between our two committees be done away with. Accordingly, with regard to SAUROPOD and T-REX, we object that their encoding is a woefully inadequate representation of a class of animals that is both important and extremely popular.

Our extensive ballot comments on dinosaur symbols from PDAM 1.2 have discussed this previously and remain valid. We have repeated many of them here, as they seem to us to have been dismissed without due consideration, due to the imminent publication of Unicode 10.0 at the time PDAM 1.2 was balloted. To put it frankly, if the UTC is going to invent symbols and encode them, there must be a price to pay. That price is simple: accept reasonable additions, deletions, and alterations of the things which they propose. We do not insist or expect that the emoji property will be given to any of the symbols we propose. We do think that the UTC have invented a drastically incomplete set of symbols, without any rationale as to its incompleteness and so, to receive our support, the repertoire must be expanded somewhat to be

complete. (In his disposition of comments to PDAM 1.2, the editor suggested that "Both characters are to be used in an Emoji context, not as part of a Dinosaur taxonomy'. This cannot be accepted. All characters in the UCS are available for use by everyone, and all symbol characters which are valid symbol characters whether or not the UTC applies the emoji property to some of them.)

We know that some members of the UTC recognize that they "dropped the ball" with regard to approving and publishing characters without allowing SC2 sufficient time and without permitting changes requested by more than one National Body in ballot comments. And we believe that the desire on the part of (at least most) members of both committees is to work well together. But this failure is part of a structural problem. Admittedly, publication scheduling could well make expedience a tempting thing for companies wishing to take advantage of public enthusiasm for a class of graphic characters. But this does not mean that SC2 National Bodies have ceded, or should cede, or will cede their responsibility in evaluating and maintaining the Universal Character Set. And, well, this means that the UTC cannot just do what they want. The right way forward is for them to expect, and to welcome, constructive oversight by SC2 National Bodies on the characters invented by their Emoji Subcommittee. That includes accepting that decisions made by that Subcommittee might not be the right or the best ones.

We know of a proposal to encode an OSTRICH emoji was submitted with over 20 pages of citations from people who were specifically requesting such an emoji for the "head-in-the-sand" metaphor. This should surely be considered to be a reasonable request for an emoji-type character. The proposer was told to follow the Emoji Subcommittee rules involving search-engine word counts, and was told never to submit so many examples of requests again. On the other hand, the CRICKET emoji seems to have been approved despite the lack of any frequency of requests for such a character; the proposal (L2/16-300; there is no WG2 document), the statement is made explicitly:

> No mention of "cricket" is included in the any of the Unicode meeting minutes of proposed emoji from the last year. However anecdotal this example may be, several friends and colleagues have expressed excitement over the idea of a cricket emoji.

This does not inspire confidence in the process for accepting or rejecting this class of graphic symbol characters.

Do we object to the encoding of CRICKET? Not as such. But we recognize that the Emoji Subcommittee's process is both inconsistent and imperfect. It certainly was inconsistent and imperfect in terms of T-REX and SAUROPOD. The input of SC2 expertise, however, can, should, and must supply the needed oversight on this process.

Again:

> To put it frankly, if the UTC is going to invent symbols and encode them, there must be a price to pay. That price is simple: accept reasonable additions, deletions, and alterations of the things which they propose. We do not insist or expect that the emoji property will be given to any of the symbols we propose.

This is not rhetoric. The UTC is inventing characters and this is not a part of our normal process. SC2 must therefore insist that we exercise additional powers of scrutiny and emenda-

tion in evaluating them. If the UTC works with us as equal partners and accepts our input, both our working relationship and the content of the standard will be improved.

Ireland also believes that it is necessary that ISO/IEC 10646 contain a new normative annex describing the sequences established for combinations together with representative glyphs.

In our comments on PDAM 1.3, Ireland requested the addition of number of characters on the next available PDAM, in the Supplemental Symbols and Pictographs bloc. These address some of the shortcomings in the set of characters invented by the Unicode Technical Committee's Emoji Subcommittee. We present below a list of these characters, and follow them with the proposal summary form for formality's sake.

Portrait and role symbols

1F93F   BILLIARD GAMES

→ 1F3B1 ⑧ billiards

Dinosaurs and other prehistoric reptiles

1F9A1   ANKYLOSAUR

• a suborder of the order Ornithischia

1F9A2   ARCHAEOPTERYX

• a genus of the suborder Theropoda of the order Saurischia

1F9A3   CERATOPSIAN

• a suborder of the order Ornithischia

• includes Protoceratops, Styracosaurus, Triceratops

1F9A4   DROMAEOSAURID

• a family of the suborder Theropoda of the order Saurischia

• includes Deinonychus, Utahraptor, Velociraptor

1F9A5   HARDROSAURID

• a family of the suborder Ornithopoda of the order Ornithischia

• includes Edmontosaurus, Hadrosaurus, Parasaurolophus

1F9A6   ICHTHYOSAUR

• a member of the order Ichthyosauria

1F9A7   IGUANODON

• a genus of the suborder Ornithopoda of the order Ornithischia

1F9A8   ORNITHOMIMID

• a family of the suborder Theropoda of the order Saurischia

• includes Gallimimus, Ornithomimus, Struthiomimus

1F9A9   PACHYCEPHALOSAUR

• a family of the suborder Pachycephalosauria of the order Ornithischia

• includes Pachycephalosaurus, Stegoceras

	1F9AA	PLESIOSAUR

• a member of the order Plesiosauria

• includes Elasmosaurus, Liopleurodon, Plesiosaurus

• represents Nessie, the Loch Ness Monster

	1F9AB	PTEROSAUR

• a member of the order Pterosauria

• includes Pteranodon, Pterodactylus, Quetzalcoatlus

	1F9AC	SPHENACODONTID

• a genus of the family Sphenacodontidae

• includes Ctenospondylus, Dimetrodon, Secodontosaurus, Sphenacodon

	1F9AD	STEGOSAURIAN

• a suborder of the order Ornithischia

• includes Huayangosaurus, Kentrosaurus, Stegosaurus

Extinct creatures

	1F9AE	SABRE-TOOTHED CAT

= Smilodon

	1F9AF	MAMMOTH

• indicates great size

Animal symbols

	1F9C9	DEER IN PROFILE

	1F9CA	GORILLA IN PROFILE

	1F9CB	RHINOCEROS IN PROFILE

Extinct creature

	1F9CC	DODO

• indicates extinction

Animal symbols

	1F9CD	BADGER

	1F9CE	SQUIRREL

Fantasy being

	1F9CF	TROLL

A. Administrative

1. Title
Future additions to the UCS
2. Requester's name
Irish National Body
3. Requester type (Member body/Liaison/Individual contribution)
Member body contribution.
4. Submission date
2017-09-24
5. Requester's reference (if applicable)
6. Choose one of the following:
6a. This is a complete proposal
Yes.
6b. More information will be provided later
No.

B. Technical – General

1. Choose one of the following:
1a. This proposal is for a new script (set of characters)
No.
1b. Proposed name of script
1c. The proposal is for addition of character(s) to an existing block
Yes
1d. Name of the existing block
Supplemental Symbols and Pictographs
2. Number of characters in proposal
22.
3. Proposed category (A-Contemporary; B.1-Specialized (small collection); B.2-Specialized (large collection); C-Major extinct; D-Attested extinct; E-Minor extinct; F-Archaic Hieroglyphic or Ideographic; G-Obscure or questionable usage symbols)
Category A.
4a. Is a repertoire including character names provided?
Yes.
4b. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?
Yes.
4c. Are the character shapes attached in a legible form suitable for review?
Yes.
5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?
Michael Everson.
5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:
Michael Everson, Fontographer.
6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?
No.
6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?
No.
7. Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?
No.
8. Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see Unicode Character Database http://www.unicode.org/ Public/UNIDATA/UnicodeCharacterDatabase.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.
See above.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.
Yes.
2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?
No.
2b. If YES, with whom?
General use.
2c. If YES, available relevant documents
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?

General use symbols.
4a. The context of use for the proposed characters (type of use; common or rare)
General use.
4b. Reference
5a. Are the proposed characters in current use by the user community?
Yes.
5b. If YES, where?
Fonts and images for all of these can be found easily.
6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?
No.
6b. If YES, is a rationale provided?
6c. If YES, reference
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?
No.
8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?
No.
8b. If YES, is a rationale for its inclusion provided?
8c. If YES, reference
9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?
No.
9b. If YES, is a rationale for its inclusion provided?
No.
9c. If YES, reference
10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?
No.
10b. If YES, is a rationale for its inclusion provided?
10c. If YES, reference
11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?
No.
11b. If YES, is a rationale for such use provided?
11c. If YES, reference
11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
No.
11e. If YES, reference
12a. Does the proposal contain characters with any special properties such as control function or similar semantics?
No.
12b. If YES, describe in detail (include attachment if necessary)
13a. Does the proposal contain any Ideographic compatibility character(s)?
No.
13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?