

TO: UTC
FROM: Deborah Anderson, SEI, UC Berkeley
SUBJECT: Feedback from WG2 email discussion list on PDAM 2.2 and DAM1 for Unicode 11.0
DATE: 17 January 2018

Below is feedback received over the WG2 email list regarding characters in the ISO/IEC 10646:2017 (5th edition) [PDAM 2.2](#) and [DAM1](#). This collection of feedback fulfills resolution M 22-10 ([L2/17-350](#)) from the September 2017 SC2 meeting, as a way to relay WG2 feedback to the UTC before a ballot takes place and its results are available. (The PDAM 2.2 closes 21 February 2018, and the DAM 1 20 April 2018, according to INCITS.)

Sections 1 and 2 include comments on characters that have been [accepted for Unicode Version 11.0](#). Section 3 lists comments on other emoji characters, included here for any Emoji Subcommittee discussion on future repertoire.

The format used below is to provide a very short summary of the main point(s) in bold, with specific comments appended. Any requested character additions are noted in parentheses.

1. EMOJI COMMENTS (REGARDING CHARACTERS IN UNICODE VERSION 11.0)

General Comments

MICHAEL EVERSON

Some glyph improvements will be proposed on Irish ballot comments.

Note that in the technical comments below, some character additions are proposed. As I have reiterated many times, these additions are given in oversight of the inventions of the Emoji Subcommittee, which create gaps in the character repertoire and which are not uniformly subject to adherence to rigorous selection criteria. The ESC may decide to make some characters into emoji, but UCS characters can be used generally, so gaps created by such invention should be filled. Once again, we in SC2 are being asked to accept ESC invented characters, and some measure of reciprocity is expected.

Responses from the ESC to proposals along these lines have ALWAYS been “no emoji proposal has been received”. This is not acceptable. All pictographs are characters. ESC inventions are not systematic, and the ESC accepts characters on thin or little evidence while rejecting characters with much evidence.

Invention goes both ways, and SC2 NB appointed experts have “invented” far fewer things, in reaction to ESC inventions. To date, as well, the only discussion of this problem has been one-sided as well.

Once again, the encoding of these characters is as UCS characters for general use. The ESC can decide what it wants to give the colourful emoji bit.

1F974 FACE WITH UNEVEN EYES AND WAVY MOUTH: change name (L2/17-393 recommended INEBRIATED FACE)

MICHAEL EVERSON

I object to this character name. The intended meaning of this character is “intoxication” and “unevenness” of eyes and “waviness” of mouth are not the only ways to represent this concept. We have encoded a number of characters specifically referring to alcoholic beverages, so there seems to be no need to avoid the name “INTOXICATED FACE”. The suggestion that “INTOXICATED FACE” is a “loaded term” is silly.

1F9A0 MICROBE: name is in error, rename “AMOEBA” (Add BACTERIA and VIRUS

MICHAEL EVERSON

This character is simply a mistake, A “microbe” is just a “little thing”. Even from an emoji point of view, people routinely talk about (and distinguish) BACTERIA and VIRUS. MICROBE should be re-named AMOEBA and BACTERIA and VIRUS should be added to the standard.

There is no valid “inclusivity” of MICROBE to subsume BACTERIA and VIRUS and AMOEBA. Ask the CDC.

1F9B8 SUPERHERO: request to change name due to possible trademark issue (but other views expressed name as acceptable)

ANDREW WEST

There was some discussion on twitter recently about the fact that "super hero" is trademarked by DC Comics (see <https://twitter.com/FakeUnicode/status/926767529497739265> and https://en.wikipedia.org/wiki/Superhero#Trademark_status). Although it is far from clear that Unicode's character names infringe on the trademark, to be on the safe side it may be advisable to change SUPERHERO to COSTUMED HERO and SUPERVILLAIN to COSTUMED VILLAIN.

KEN WHISTLER (response to above)

On this topic, keep in mind that the publicly visible labels for these two, by a significant margin, will be whatever the "CLDR name" ends up being. And those are currently "superhero" (and "man superhero" and "woman superhero" for the gender-specific versions). See: <http://www.unicode.org/emoji/future/emoji-candidates.html>

Compared to the visibility of those labels, the formal Unicode/10646 name of the character(s) doesn't matter that much. It would be cleanest, however, to just keep the two aligned, where possible.

By the way, SUPERHERO is not the only new (or old) emoji with prior trademark claims on particular names. See also TEDDY BEAR, for example. But I don't think that names of individual pictographic characters like this, when the names also reflect very widespread generic usage, are problematical for us.

MICHAEL EVERSON (response to Ken W above)

> On 10 Jan 2018, at 16:10, Ken Whistler via WG2 <wg2@unicode.org> wrote:
>

> [...] I don't think that names of individual pictographic characters like this, when the names also reflect very widespread generic usage, are problematical for us.

I agree. The term "superhero" occurs in pretty much every dictionary and has for a long time. OED:

c. In recent (often nonce) formations after superman, used to designate a person, animal, or thing which markedly surpasses all others, or the generality, of its class: e.g. "super-being, block [block n. 14], -boss [boss n.6], -brain, -brute, -car, -carrier [carrier 1m], -cinema, -city, -computer, -critic, -crook [crook n. 13], -dramatist, -goddess, -grid [grid 8a], -gun, -hero, -heroine, -journalist, -liner [liner2 8a], -magic, -male (see also sense 5f), -nation, -patriot, -port [port n.1 1], -profit, -race, rich [rich n. 11], -salesman, -salesmanship, -ship, -sleuth, speed, -spy, -stud [stud n.2 4d], -tanker [tanker1 1a], -tramp

Whether some comic company has trademarked "superhero" or "super-hero" or "super hero" I think is not something which would concern us. I mean... if DC has "super hero", what has Marvel got?

1F9C2 SALT SHAKER: (Add PEPPER POT)

MICHAEL EVERSON

The encoding of this character produces a gap in the UCS. PEPPER POT should be added.

1F9E7 RED GIFT ENVELOPE: change name to RED ENVELOPE and add annotation

MICHAEL EVERSON

This is not the standard name for this character. It should be reverted to RED ENVELOPE and an informative note should describe the character's use.

1F9EA TEST TUBE; 1F9EB PETRI DISH; 1F9EC DNA DOUBLE HELIX (Add ERLNMEYER FLASK)

MICHAEL EVERSON

These are useful characters, but ERLNMEYER FLASK is missing from the UCS to complete a laboratory set of characters. An ERLNMEYER FLASK is not a TEST TUBE. Both are required.

1F9F1 BRICKS: No good rationale for plural BRICKS provided, so change character to BRICK (Add BRICK WALL)

ANDREW WEST

My only comment is that I strongly object to encoding a BRICKS emoji.

No-one needs an emoji for a pile of bricks, and nobody has made a published submission for a pile of bricks emoji, so what is the rationale for encoding a BRICKS emoji? The original emoji submission was for a BRICK emoji (<http://www.unicode.org/L2/L2017/17172-brick-emoji.pdf>), which was later changed to BRICK WALL, and then finally to BRICKS with the note "Image to be changed to be pile of bricks". If a pile of bricks is intended to be a compromise between a brick and a brick wall, then it is a very bad compromise because it is useful neither as a brick nor a brick wall. I do not think that we should encode arbitrary emoji such as pile of bricks on the whim of the ESC without any formal submission or detailed rationale for changing BRICK to BRICKS.

As no submission for a pile of bricks emoji has been published and no rationale for encoding a pile of bricks emoji has been made by the ESC, either the BRICKS emoji should be removed

pending a proper submission for the character, or BRICKS should be changed to back BRICK per the original submission. My personal preference is to encode both a BRICK emoji and a BRICK WALL emoji.

MICHAEL EVERSON

BRICKS is a useless character. Nobody talks about piles of bricks. Images of piles of bricks aren't used to indicate anything. There is no reason for this to be plural. Moreover running into a brick wall seems to be a great metaphor for all of these discussions. The following are required:

BRICK

BRICK WALL

1F9F2 MAGNET: (Add HORSESHOE)

MICHAEL EVERSON

Reports are that some people propose using MAGNET on the basis of its shape alone rather than its true properties. This is a mistaken notion, and HORSESHOE should be lucky and get encoded.

2. COMMENTS ON NON-EMOJI CHARACTERS (REGARDING CHARACTERS IN UNICODE VERSION 11.0)

Latin Extended D: change names for 2 characters

MICHAEL EVERSON

A7B8 LATIN CAPITAL LETTER U WITH STROKE

A7B9 LATIN SMALL LETTER U WITH STROKE

These names should be changed to U WITH DIAGONAL STROKE. The UCS already has a character U WITH STROKE and in that the stroke is horizontal:

1D7E ⚡ LATIN SMALL CAPITAL LETTER U WITH STROKE

Hanifi Rohingya: change names for 5 characters

MICHAEL EVERSON

10D1D HANIFI ROHINGYA VOWEL MARK A

10D1E HANIFI ROHINGYA VOWEL MARK I

10D1F HANIFI ROHINGYA VOWEL MARK U

10D20 HANIFI ROHINGYA VOWEL MARK E

10D21 HANIFI ROHINGYA VOWEL MARK O

In the UCS we do not use the term VOWEL MARK for this kind of letter. There is only one "VOWEL MARK" in the standard and that is 111CC SHARADA EXTRA SHORT VOWEL MARK which is not like these. The term VOWEL SIGN is used in the Standard, and this term was also used in the proposal for Hanifi Rohingya (L2/16-311R); there is no indication of any discussion where SIGN was changed to MARK. However, the term VOWEL SIGN usually refers to Brahmic dependent vowel signs. The Hanifi Rohingya characters function like other letters in the script (they have shaping behaviour), so they are really LETTERS. They also have traditional names. There are three options we could choose (not given in any order of preference):

10D1D HANIFI ROHINGYA VOWEL AA [sic]

10D1E HANIFI ROHINGYA VOWEL I

10D1F HANIFI ROHINGYA VOWEL U

10D20 HANIFI ROHINGYA VOWEL E
10D21 HANIFI ROHINGYA VOWEL O

10D1D HANIFI ROHINGYA LETTER AA [sic]
10D1E HANIFI ROHINGYA LETTER I
10D1F HANIFI ROHINGYA LETTER U
10D20 HANIFI ROHINGYA LETTER E
10D21 HANIFI ROHINGYA LETTER O

10D1D HANIFI ROHINGYA LETTER AA-FOR
10D1E HANIFI ROHINGYA LETTER I-FOR
10D1F HANIFI ROHINGYA LETTER U-FOR
10D20 HANIFI ROHINGYA LETTER E-FOR
10D21 HANIFI ROHINGYA LETTER O-FOR

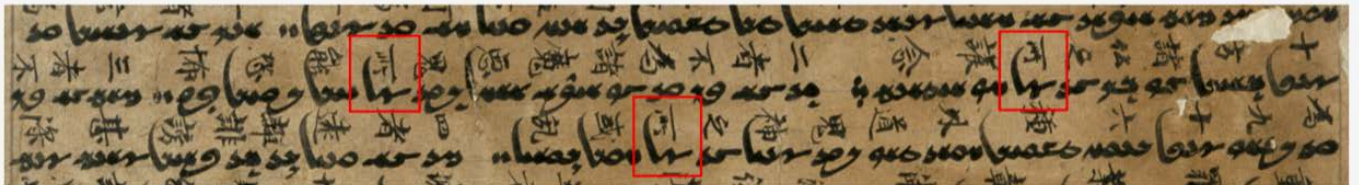
But the current names are inconsistent with the rest of the standard.

Sogdian: change name or remove one character

MICHAEL EVERSON

10F45 SOGDIAN PHONOGRAM SHIN

The term “phonogram” is not used elsewhere in the standard and should not be introduced. The character is a variant of the letter SHIN used to represent the Chinese word 所 ‘suǒ’. We could call it “LOGOGRAM SUO”, but I don’t know how this thing is to be read. We should get a proper reading if possible of the Sogdian text (as presumably pronounced). This is a DAM. Perhaps it would be best to leave this character out for further study. Giving it a misnomer would not be a good idea, and it is doubtful that immediate encoding is urgent. At a minimum if we can’t decide on the status of it, this one character should be moved to PDAM 2.2.






Three instances of  PHONOGRAM SHIN used in place of the normative isolated form  of SHIN in So 14830. The  is used here for transcribing Chinese 所 *suǒ*, and reflects the syntactic function or the enclitic nature of 所, which is a type of relative marker (Yoshida, personal communication, 2016).

Figure 22: Specimens of *shin*.

3. EMOJI COMMENTS (REGARDING POST-UNICODE 11.0 REPERTOIRE)

1F93F BILLIARD GAMES

MICHAEL EVERSON

I would like to reiterate my very strong support for the encoding of this gaming character, particularly as vendors co-opted BILLIARDS for the magic 8-ball.

1F97B TROLL

MICHAEL EVERSON

I would like to reiterate my very strong support for the encoding of this fictional-being character.

1F9A3 MAMMOTH; 1F9A4 DODO

MICHAEL EVERSON

I would like to reiterate my very strong support for the encoding of these characters.

1F9A5 SQUIRREL

MICHAEL EVERSON

I would like to reiterate my very strong support for the encoding of this character. In particular I note that some vendors have been violating the character identity of 1F43F CHIPMUNK by drawing it as a squirrel. Not only are chipmunks and squirrels very different creatures, but the chipmunk encoded was for compatibility with a Windows character set, and turning it into a squirrel invalidates the mapping.