Title: Emoji Ad-Hoc Meeting Report

Date: 2009-4-22
Source: Emoji Ad-hoc committee

An ad-hoc committee on Emoji encoding met in Dublin on April 21, 2009. The following were in attendance:

> Tero Aalto, Dae Hyuk Ahn, Deborah Anderson, Chen Zhuang, Peter Constable, Michael Everson, Gábor Hosszú, Kang Mi-Young, Kim Kyongsok, Tatsuo Kobayashi, Mike Ksar, Tamara Lopez, Yoshi Mikami, Karl Pentzlin, Masahiro Sekiguchi, Andreas Stötzner, Michel Suignard, André Szabolcs Szelp, V. S. (Uma) Umamaheswaran, Andrew West, Ken Whistler, Satoshi Yamamoto

Additionally, the following participated by telephone:

> Mark Davis, Peter Edberg, Katsuhiko Momoi, Markus Scherer

The ad-hoc committee meeting was chaired by Peter Constable.

The Ad-hoc committee had long deliberations on many issues related to the US/Unicode proposal in N3582/N3583 and to the suggested revisions to that proposal described by the Germany and Ireland in N3607. Through these deliberations, the Ad-hoc committee was able to arrive at consensus regarding revisions to the repertoire in N3583 that would yield a repertoire all found acceptable for addition to a PDAM document at this meeting. The Emoji Ad-hoc committee recommends to WG2 that the repertoire resulting from the agreed-upon revisions be added to a PDAM ballot during Mtg 54.

The following describes the major issues and consensus decisions of the Ad-hoc committee.

1. To begin, the chair presented an agenda for working through issues, and communicated an operational principle for the meeting: that the point of reference for discussions and proposed changes would be the Project Editor's document, N3580, which reflects the repertoire in N3583; and that agreed-upon changes would be recorded as changes in relation to that document. The operating principle was agreed upon.

2. The committee began discussions focused on determining a repertoire suitable for balloting. It was agreed that code position assignments of the final repertoire could follow the approach of N3607, and could be determined following the meeting between the US, Germany, Ireland, and the Project Editor. Thus, the remaining discussions in the meeting focused in name or glyph changes proposed in N3607, additions proposed in N3607, and status of certain controversial characters in N3583.

3. The committee identified three classes of characters proposed in N3583 that were controversial: five culturally iconic characters, ten region/language indicators and 66 Emoji Compatibility characters. Discussion of these 81 characters was postponed until after all other issues had been discussed.

4. The committee noted that there was a large degree of consensus already expressed between N3583 and N3607: for 382 of 674 characters, N3607 proposed no substantive change from N3583 and found no objectionable issues. The committee agreed that this class of characters was ready for inclusion in a ballot document.

5. The committee heard a summary assessment from US experts of the glyph and name changes proposed in N3607: some proposed glyph or name changes were seen as good, others were seen as neutral, but others were seen to be problematic in that they constituted a change in the

intended character. The US was agreeable to accepting 30 name changes that were seen to be good or neutral, and 116 glyph changes considered good or neutral. Altogether, this represented an additional 130 characters on which there was now a consensus in terms of definition, name and glyph. The committee agreed that these characters were, therefore, ready for inclusion in a ballot document.

6. This left 81 characters for which the US experts considered the name or glyph changes to be bad because they constituted a change in the intended character. There was consensus on a general principle to add these characters to a ballot document as proposed in N3583 without change. In subsequent discussion, a small number of cases among these 81 were found in which a name or glyph change was found to be appropriate; these cases are discussed further below.

7. Of the 81 characters for which N3607 name or glyph changes were considered problematic, many involved symbols for animals. N3583 had used generic names in some cases in which the character in current usage is, in fact, somewhat specific. For example, the name TIGER had been used in N3583 for a TIGER FACE symbol. The changes in N3607 had treated these as truly generic and proposed a different glyph accordingly, but this constituted a change from the character in current usage. It was agreed that the treatment of animal symbols in N3583 and N3607 should be re-analyzed to dis-unify in cases in which the character in current usage is distinct from a truly-generic animal symbol. It was agreed that this could be done following the ad-hoc meeting by experts from US, Germany and Ireland with the resulting repertoire considered ready for addition to a ballot document. It was also agreed that the repertoire would include generic animal symbols to cover zodiacal symbols and symbols for general food-menu categories (beef, pork, poultry).

8. N3607 had proposed addition of 143 characters. (This does not include any additions resulting from dis-unification of animals.) All of these were reviewed during the meeting. There was agreement to include 132 of the proposed additions into the repertoire to recommend for addition to a PDAM during WG2 meeting 54.

9. EARTH GLOBE (proposed in N3583 at U+1F322) was one of the 81 cases in which name or glyph changes proposed in N3607 was considered problematic by US experts. In review of proposed additions, it was found that EARTH GLOBE could be unified with one of the additions proposed in N3607, EARTH GLOBE ASIA-AUSTRALIA. It was agreed that the name EARTH GLOBE would be changed to EARTH GLOBE ASIA-AUSTRALIA and the use the glyph used in N3607 (U+1F30D in that document).

10. CACTUS (proposed in N3583 at U+1F330) was one of the 81 cases in which the name or glyph changes proposed in N3607 was considered problematic by US experts. In review of proposed additions, it was noted that N3607 proposed a dis-unification to two distinct cacti symbols, and that the additional character proposed in N3607 was, in fact, the character intended in the US proposal. As a result, it was agreed not to add a second cactus symbol but to change the glyph for CACTUS.

11. WESTERN PERSON (proposed in N3583 at U+1F358) was one of the 81 cases in which name or glyph changes proposed in N3607 was considered problematic by US experts. N3607 had proposed a male/female pair of symbols—entailing a name change for the already-proposed character plus the addition of another character. In review of that proposed addition, US experts clarified that the character in current usage was gender neutral, displayed with either male or female images depending on the vendor. It was agreed not to add an additional character, and

for the already-proposed character to keep the name WESTERN PERSON but to change the glyph to a gender-neutral image.

12. EXPRESSIONLESS FACE (proposed in N3583 at U+1F39D) was a character for which a glyph change was proposed in N3607 that was considered a good change by US experts. In review of proposed additions, details of this character were clarified by US experts, and it was found that one of the proposed additions, DEADPAN FACE (1F629 in N3607) was, in fact, the same character. As a result, it was agreed not to add another character but to change the glyph of EXPRESSIONLESS FACE to the glyph used in N3607 for the proposed addition DEADPAN FACE.

13. The above notes capture all decisions taken during the meeting related to Emoji characters other then the three classes of controversial characters noted earlier. The result at this point was a consensus that 725 characters are ready for addition to a PDAM ballot document. This set includes the 674 characters proposed in N3583 (with agreed-upon name or glyph changes) less 81 controversial characters plus 132 agreed-upon additions. This does not include further additions resulting from re-analysis of animal symbols.

14. Discussion then turned to the three classes of controversial characters noted above. Throughout this discussion, there was broadly-held sentiment that none of these proposed characters were ones that people liked to encode, but there was also general understanding of the need for interoperability with existing usage in Japan.

15. The first class of controversial characters discussed was five culturally-iconic symbols: MOUNT FUJI (U+1F3E1), TOKYO TOWER (U+1F3E2), STATUE OF LIBERTY (U+1F3E3), SKETCH OF JAPAN (U+1F3E4) and MOYAI (U+1F3E5). A consensus was reached to include these characters in the repertoire to be recommended for addition to a PDAM ballot document, but with names changed as follows:

Code position in N3583	Name in N3583	New name
U+1F3E1	MOUNT FUJI	EMOJI COMPATIBILITY SYMBOL-1
U+1F3E2	TOKYO TOWER	EMOJI COMPATIBILITY SYMBOL-2
U+1F3E3	STATUE OF LIBERTY	EMOJI COMPATIBILITY SYMBOL-3
U+1F3E4	SKETCH OF JAPAN	EMOJI COMPATIBILITY SYMBOL-4
U+1F3E5	MOYAI	EMOJI COMPATIBILITY SYMBOL-5

(Note: per the consensus decision noted in point 2, the code positions shown are not final.)

16. The second class of controversial characters discussed was the 66 characters proposed in N3583 in an Emoji Compatibility Symbols block. The US withdrew its request to encode these characters at this time.

17. The last class of controversial characters discussed was the ten Emoji regional indicator symbols proposed in N3583 in the range U+1F401 to U+1F40A. In N3607, Germany and Ireland had proposed as an alternative to encode 676 symbols for two-letter codes. After long discussion, the ad-hoc committee was not able to reach consensus on some compromise. The committee was left with three options:

   a. Do not encode any of these characters at this time.

b. Include ten characters in the repertoire to be added to a PDAM ballot document with names changed to the form EMOJI COMPATIBILITY SYMBOL-n.

c. Include 676 symbols for two-letter codes.

A straw poll was conducted to determine how many national bodies could live with each of the three options. (Ten national bodies were represented in the ad-hoc committee at this point.) Results were as follows:

Can live with option a (do not encode at this time): 7

Can live with option b (include ten characters in a PDAM ballot): 6

Can live with option c (include 676 symbols for two-letter codes): 3

In accordance with this, the decision of the committee was not to include any characters from this class in the repertoire to be recommended for addition to a PDAM ballot document during WG2 meeting 54.

18. This concluded the agenda for the meeting. The result was a consensus that 730 characters are ready for addition to a PDAM ballot document. This set includes the 674 characters proposed in N3583 (with agreed-upon name or glyph changes) less 76 controversial characters plus 132 agreed-upon additions. This does not include further additions resulting from re-analysis of animal symbols.

Per the consensus decision recorded in point 7, individual experts from Germany, Ireland and US met after the ad-hoc meeting to resolve the open issues related to animal symbols. The outcome was that names for fifteen characters in N3583 should be changed and that 18 additional characters should be included in the repertoire.

In reviewing the agreed-upon changes, it was found that ten additional name changes were needed to reconcile name conflicts resulting from agreed-upon decisions, to create a consistent set of names for eight moon phases, or to align names with other existing naming conventions. Nine of these cases pertain to characters among the set of 81 for which US experts had expressed concern with name or glyph changes. To resolve the conflict, US experts agreed to name changes proposed in N3607. These changes do not impact the definition of characters. (E.g., BUS is changed to BUS-2.)

In summary, the final result includes changes from N3583 as follows:

- Remove 76 characters

- Add 150 characters

- 61 name changes

- 119 glyph changes

The final repertoire includes 748 characters. Per the consensus decisions of the Emoji Ad-hoc committee, it is recommended that these be added to a PDAM ballot document during WG2 meeting 54. The final repertoire is reflected in N3626.

To accommodate the proposed additions, the following new block allocations are suggested:

- Playing Cards, U+1F0A0 – U+1F0FF

- Miscellaneous Pictographic Symbols, U+1F300 – U+1F5FF

- Emoticons, U+1F600 – U+1F64F

- Transport and Map Symbols, U+1F680 – U+1F6FF

In summary, proposed code point allocations are as follows:

- Miscellaneous Technical (U+2300 – U+23FF): 11 characters

    o U+23E9 – U+23F3

- Miscellaneous Symbols (U+2600 – U+26FF): 1 character

    o U+26CE

- Dingbats (U+2700 – U+27FF): 12 characters

    o U+2705

    o U+270A – U+270B

    o U+2728

    o U+274C

    o U+274E

    o U+2753 – U+2755

    o U+2795 – U+2797

- Supplemental Punctuation (U+2E00 – U+2EFF): 1 character

    o U+2E32

- Playing Cards (U+1F0A0 – U+1F0FF): 59 characters

    o U+1F0A0 – U+1F0AE

    o U+1F0B1 – U+1F0BE

    o U+1F0C1 – U+1F0CF

    o U+1F0D1 – U+1F0DF

- Enclosed Alphanumeric Supplement (U+1F100 – U+1F1FF): 4 characters

    o U+1F170 – U+1F171

    o U+1F17E

    o U+1F18E

- Enclosed Ideographic Supplement (U+1F200 – U+1F2FF): 13 characters

    o U+1F201 – U+1F202

    o U+1F232 – U+1F23A

    o U+1F250 – U+1F251

- Miscellaneous Pictographic Symbols (U+1F300 – U+1F5FF): 517 characters

    o U+1F300 – U+1F30C

    o U+1F30F

    o U+1F311 – U+1F320

- o U+1F330 – U+1F335
- o U+1F337 – U+1F37C
- o U+1F380 – U+1F393
- o U+1F3A0 – U+1F3C4
- o U+1F3C6 – U+1F3CA
- o U+1F3E0 – U+1F3F0
- o U+1F400 – U+1F43D
- o U+1F440
- o U+1F442 – U+1F46F
- o U+1F471 – U+1F49E
- o U+1F4A0 – U+1F4B6
- o U+1F4B8 – U+1F4E1
- o U+1F4E3 – U+1F518
- o U+1F51C – U+1F528
- o U+1F52A – U+1F53D
- o U+1F540 – U+1F543
- o U+1F54B – U+1F54E
- o U+1F550 – U+1F55B
- o U+1F560 – U+1F564
- Emoticons (U+1F600 – U+1F64F): 61 characters
  - o U+1F600 – U+1F628
  - o U+1F62A – U+1F63D
- Transport and Map Symbols (U+1F680 – U+1F6FF): 69 characters
  - o U+1F680 – U+1F6C4

Complete details of the proposed repertoire are available in N3626.