

Some suggestions about the encoding of national flags as requested by the Emoji proposal (L2/08-081)

Karl Pentzlin – 2008-08-09

On 2007-08-10, the author posted on the Unicore mailing list and on the Unicode symbols mailing list:

Suggestion for encoding flags (as contained in emoji)

The emoji set as shown in L2/07-257 contains the flags of some countries.

To keep Unicode away from all political quarrels at this point, I suggest to encode flag symbols in a generic way.

- 1.) Declare a separate Flags block in the SMP (or elsewhere) containing at least 43 columns (688 code points), rather than to put them together with other emojis or symbols in a block.
- 2.) Define 676 (26 x 26) symbols for each possible combination of two letters A...Z, whose representative glyph as printed in the standard document is a flag showing these letters black on white.
I.e., the symbols are named like:
 - FLAG AA
 - FLAG AB
 - ...
 - FLAG AZ
 - FLAG BA
 - ...
 - FLAG BZ
 - ...
 - FLAG ZZ
- 3.) The semantics of these symbols are:
 - If the two letter combination is a country code in the actual version of ISO 3166, the actual flag of that political entity is a valid glyph of that symbol.
 - If the two letter combination is a valid Internet ccTLD, the actual flag of the associated political entity is also a valid glyph of that symbol. (E.g., "EU" is a valid ccTLD but no country code.)
- 4.) Thus, Unicode neither discriminates any political entity (as by encoding the flag symbol of important countries while declaring other countries unimportant by not encoding their flags), nor does it need to follow any changes in ISO 3166 or the ccTLD system, nor is it involved in any name quarrels (e.g. naming a flag for Palestine or Taiwan).
- 5.) Moreover, Unicode has not to care about the actual design of the flag (whether Italy adjusts the exact hues, or the US adds a 51st star).
- 6.) Flag symbols for codes defined in the actual version of ISO 3166 may be given an informative note in the printed version of the Unicode standard using "Flag of" + the English short country name officially used by the ISO 3166 Maintenance Agency (ISO 3166/MA),
e.g.:
 - FLAG AD
 - Flag of Andorra
 as these informative notes (unlike the character names) can be changed on every new printed edition of the Unicode standard.
- 7.) The suggested Flags block is restricted in a transparent way and by definition is not open for any flags of arbitrary regions, towns or organizations.
Thus, it does not invite everybody to propose their favorite local flag for encoding.
- 8.) If there is a need to distinguish waving flags (as shown in the emoji proposal) from strictly rectangular flag symbols, there two blocks of this kind following exactly the same pattern are needed.

On 2007-08-10, the author answered on the Unicore mailing list:

Doug Ewell 2007-08-10:

> *What happens when ISO 3166 withdraws a country code element, as they did last year for 'CS' when Serbia and Montenegro broke up?*

As long as the two letter combination is not assigned to another political entity, there is no problem. The definition of the character as suggested above does no longer explicitly declare the flag of the no longer existing country to be a valid glyph variant, but also does not explicitly exclude it. If there is a policy effective to reuse ISO 3166 codes neither as new ISO 3166 codes nor as ccTLDs (which I have not researched for at this time), then it makes sense to declare the now historic flag used for the political identity at the time of its dissolution staying a valid glyph variant. The informative note for such code points (see 6. of my original posting) then may appear like

FLAG SU

· Former Flag of Soviet Union

Recent addition:

http://www.iso.org/iso/customizing_iso_3166-1.htm#user-assigned-code-elements states:

User-assigned code elements

ISO 3166-1 specifies a series of code elements for user purposes which the ISO 3166/MA will never use in the updating process of the standard. To quote from ISO 3166-1:1997, clause 8.1.3 User-assigned code elements:

"If users need code elements to represent country names not included in this part of ISO 3166, the series of letters AA, QM to QZ, XA to XZ, and ZZ, and the series AAA to AAZ, QMA to QZZ, XAA to XZZ, and ZZA to ZZZ respectively and the series of numbers 900 to 999 are available."

These code elements are at the disposal of users who need to add further names of countries, territories or other geographical entities to their in-house application of ISO 3166-1. When exchanging data with users of ISO 3166-1 not connected to this particular in-house application the definition of these user-assigned code elements should be given. The incorporation of the complete ISO 3166-1 standard document in commercial products may be subject to a charge.

Thus, any code points assigned to AA, QM to QZ, XA to XZ, and ZZ, would in fact become private use code points.

Therefore, code points should only be assigned to AB...QL, RA...WZ, YA...ZY, restricting the character count to 634 and the required block size to 40 columns.