Subject: **Alternative Model for Subdivision Flags**
Date: **2016-08-04**
From: **Mark Davis**
Draft: **link**

There is great demand for the flags for subregions (aka "subdivisions") within countries, such as in document 16-180. While we want to provide for these, we don't want to encode additional flags one at a time. Some subregions are claimed to be exceptional, but that depends on which features are cited. Every subregion is exceptional in its own way (even the happy ones; my apologies to Tolstoy). We thus want to follow the same principles as for the current emoji flag sequences, providing a generative mechanism with clear validity criteria and identification.

We started along that path in the draft UTS #52, but suspended work on that to focus on gender. We are now looking a reactivating that (document 16-226), but since flags would be the main focus, have discussed some alternative approaches. This presents one of those alternatives.

An **tag extended regional indicator sequence** (TERIS) is defined to be a sequence of the form <RI TAG+ TAG_TERM>, where

1. RI is U+1F1E6..1F1FF (REGIONAL INDICATOR SYMBOL LETTER A - Z)
2. TAG+ is a sequence of one or more U+E0020..E007a (TAG SPACE - TILDE)
3. TAG_TERM is U+E007F CANCEL TAG
4. The initial RI must not be part of a grapheme cluster with a previous RI. *(reword this for clarity)*

A **emoji extended flag sequence** is defined to be a TERIS that satisfies the following criteria:

1. Let SD be the string resulting from<RI TAG+>, where
    a. the RI is mapped to [a-z] by subtracting 0x1F185
    b. each TAG is mapped to [0-9a-z] by subtracting 0xE0000
2. SD must then be a specification as per [CLDR] of either a Unicode subdivision_id or a 3-digit unicode_region_subtag, *and*
3. SD must have CLDR idStatus equal to "regular" or "deprecated".

Notes:

1. The deprecated values are only included for compatibility, and should not be used.
2. The syntax allows the 3-digit forms for future compatibility, but none are currently valid.
3. Note that there is no hyphen, unlike ISO subdivisions like "GB-SCT".
4. The core definition allows characters like TAG PLUS for future-proofing; these are not currently valid.
5. Like the emoji flag sequences, these can request an image for whatever is currently the flag of the specified subregion. They are not intended to provide a mechanism for versioned representations of any particular flag image.
6. The choice of which emoji extended flag sequences to support is entirely up to platforms and programs. There is no requirement that any be supported.
    a. *Review Note: To promote interoperability, we expect to catalog those commonly supported on platforms, as is done with ZWJ sequences.*
7. However, it is recommended that unsupported and invalid emoji extended flag sequences be handled with "missing flag glyphs" as in http://unicode.org/reports/tr51/#Flag_Presentation.

Examples:

✦ represents the TAG_TERM

a-z represent the REGIONAL INDICATOR SYMBOL LETTER A - Z
0-9, A-Z represent TAG DIGIT ZERO - NINE, TAG LATIN SMALL LETTER A - Z

Name	Image	Encoding
Flag for England		gbeng✦
Flag for Scotland		gbsct✦
Flag for Wales		gbwls✦

Behavior

The fallback behavior on older systems will be to the appearance of a single RI character. This is sufficient indication that there is something odd, without appearing to be malformed text.

The segmentation behavior of the TAG characters was already adjusted in Unicode 9.0 for use in sequences, so that they don't break from previous RI or TAG characters:

```
GB: E0020..E007F  ; Extend # Cf  [96] TAG SPACE..CANCEL TAG
WB: E0020..E007F  ; Extend # Cf  [96] TAG SPACE..CANCEL TAG
LB: E0020..E007F  ; CM     # Cf  [96] TAG SPACE..CANCEL TAG
```