

# Reactivate UTS 52 mechanism in reduced form (v2)

Date: 2016-11-01

To: UTC

From: Mark Davis, Peter Edberg

Draft: [link](#)

This is a proposal to reactivate a portion of the tag mechanism originally proposed in [draft UTS 52](#), but in modified form and with reduced scope. The proposal here is to integrate the mechanism into UTR51 rather than have a separate UTS52, so that there is a single document that specifies all of the emoji sequences.

The initial support is only for subregion flags, although the mechanism is general enough to expand later. The reduction in scope is motivated by the fact that a different mechanism (with better fallback) is proposed for gender variants, while the direction and hair color tags in the original proposal were deemed of lower priority than the other proposed tags. Moreover, with the ZWJ sequence mechanism now widely supported, cases where that works (like hair color) should use ZWJ sequences instead.

*Note: there is a proposal to have a “reduced” format for just flags that should be discussed at the same time as this document [L2/16-234](#). That reduced format has the disadvantage that it is more complicated for segmentation.*

Proposed Changes to: [http://unicode.org/reports/tr51/#Emoji\\_Sequences](http://unicode.org/reports/tr51/#Emoji_Sequences)

**Renumber old [ED-15a](#). emoji zwj element to [ED-15b](#).**

**Add new definition**

## ED-15a. emoji tag sequence (ETS)

A sequence consisting of an emoji core sequence followed by one or more non-terminating TAG characters, followed by tag\_term.

```
emoji_tag_sequence := tag_base tag_key_value_pair+ tag_term
tag_base           := emoji_character | emoji_modifier_sequence
tag_key_value_pair := tag_key tag_value
tag_term           := \x{E007F} // CANCEL TAG
tag_key            := [\x{E0041}-\x{E005A}] + // tag_A..tag_Z
tag_value          := [\x{E0020}-\x{E0040}\x{E005B}-\x{E007E}] +
                    // TAGs - tag_A..tag_Z - tag_term
```

There is an additional constraint on well-formedness: each of the tag\_keys in the sequence must be in alphabetical order. So <base><tag\_B><tag\_x><tag\_A><tag\_y><term> is ill-formed.

The above defines a well-formed emoji tag sequence. Each tag\_key\_value\_pair defines a particular visual variant to be applied to the tag\_base character(s).

Note that the TAG characters in tag\_term, tag\_key, and tag\_value are disjoint.

**Review Note:** We could allow the tag\_key to be empty. That <tag\_key, tag\_value> would be required to be first in the emoji\_tag\_sequence. The tag\_value can never be empty, however.

The meaning and validity criteria for emoji\_tag\_sequences and expected visual variants for tag\_key\_value\_pairs are determined by the documentation in *Annex C. Emoji Tag Sequences*. Whenever an emoji\_tag\_sequence is not well-formed according to ED15a or not valid according to Annex C, it should be shown as the base character with an overlay, as in [Flag Presentation](#).

**Note to committee:** the previous version of this document used a visible existing emoji for tag\_term. However, that has the problem that it would require further surgery to segmentation. Thus the approach here is to instead specify that invalid or ill-formed sequences have a fallback appearance.

---

**Add emoji\_tag\_sequence to:** [ED-17](#). emoji sequence

---

**Add Review Note below it**

**Review Note:** If emoji tag sequences are to be allowed within ZWJ sequences, then **emoji\_tag\_sequence** would be added not to ED-17, but to [ED-15b](#). emoji zwj element.

---

**Add new Annex C.**

---

## Annex C. Emoji Tag Sequences

While the syntax of emoji tag sequences is defined in ED15a, only valid emoji tag sequences are to be displayed. The only valid sequences in this version of Unicode Emoji are defined by sections in this annex. Whenever an emoji tag sequence is not well-formed according to ED15a or not valid according to Annex C, it should be shown as the base character with an overlay, as in [Flag Presentation](#).

For a given <tag\_key>, only specified <tag\_base> characters and only specified <tag\_value> sequences are valid.

In examples, underlined ASCII characters represent the corresponding tag characters, while **◆** represents the tag\_term.

### C.1 Flag emoji tag sequences

A valid flag emoji tag sequence must satisfy the following constraints:

- a. The tag\_base, tag\_key, and tag\_value are limited to the following:

|                  |                               |
|------------------|-------------------------------|
| <b>tag_key</b>   | Tag_F                         |
| <b>tag_base</b>  | U+1F3F3 WAVING WHITE FLAG     |
| <b>tag_value</b> | (Tag_0..Tag_9, Tag_a..Tag_z)+ |

- b. The tag\_base must only have one tag\_key consisting of Tag\_F.
  - i. Any duplicate tag\_key renders the entire emoji tag sequence invalid.
- c. Let SD be the result of mapping each character in the TAG value to [0-9a-z] by subtracting 0xE0000.
  - i. SD must then be a specification as per [CLDR] of either a Unicode [subdivision\\_id](#) or a 3-digit [unicode\\_region\\_subtag](#), and
  - ii. SD must have CLDR idStatus equal to "regular" or "deprecated".







**Review Note:** If we allow the tag\_key in general to be empty (as discussed in a review note for **ED-15a. emoji tag sequence**) we could have *this* tag\_key be empty instead of being F, since it would be the primary use case for an emoji tag sequence with WAVING WHITE FLAG as a tag\_base.

**Notes:**

1. The deprecated values are only included for compatibility, and should not be used.
2. Note that there is no hyphen in the tag\_value, unlike ISO subdivisions like "GB-SCT".
3. Like the emoji flag sequences, these are used to 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.
4. 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.

**Review Note:** To promote interoperability, we expect to catalog those subdivision sequences that are commonly supported on platforms, as is done with ZWJ sequences.

**Examples:**

| Name              | Image   | Encoding   |
|-------------------|---|--|
| Flag for England  |  |  <u>Egbeng</u> ♦ |
| Flag for Scotland |  |  <u>Egbsct</u> ♦ |
| Flag for Wales    |  |  <u>Egbwls</u> ♦ |

**Note to committee:** We had had Private Use defined in tr51 as follows. Adding it would allow some level of experimentation, such as in L2/16-105. However, a possible danger is a profusion of forms, with some that people feel obliged to support because of their frequency; to avoid that it might be better to not add private use emoji tag sequences.

A related emoji tag sequence that we discussed earlier would allow for variant images for an emoji: it would have tag\_key=U, also apply to any tag\_base, and require that the tag value must result in a well-formed URL when prefixed by “https://” and postfixed by “.png”, with that URL pointing to a valid PNG.

## C.2 Private use emoji tag sequences

Private Use emoji tag sequences are for closed interchange within a given system. As with private use codes in general, the emoji tag sequence may have no meaning or a different meaning outside that system, so it is not suitable for general interchange.

A valid private use emoji tag sequence must satisfy the following constraints:

d. The tag\_base, tag\_key, and tag\_value are limited to the following:

|                  |   |
|------------------|---|
| <b>tag_key</b>   | tag-Z   |
| <b>tag_base</b>  | Any character with Emoji=Yes  |
| <b>tag_value</b> | [ $\backslash$ {E0020}- $\backslash$ {E0040} $\backslash$ {E005B}- $\backslash$ {E007E}] <sup>+</sup> |

*Example:*

Private use runner with ski boots:  $\square$ Zskiboosts  $\blacklozenge$