

Emoji Flags for Dependent Regions

- **Author:** Christoph Päper (christoph.paeper@crissov.de)
- **Date submitted:** 2017-06-18
- **Latest revision:** <https://crissov.github.io/unicode-proposals/dependent-regions>

Unicode emoji flags rely on ISO 3166. Contemporary sovereign countries (and some other geographical entities) are encoded in ISO 3166-1 and, for some of them, their administrative subdivisions in ISO 3166-2. Unicode Emoji (UTS#51) describes two separate methods to encode flags *Emoji Flag Sequences* and *Flag Emoji Tag Sequences*: The first uses Regional Indicator Symbol Letters A through Z (U+1F1E6..1F1FF) for code elements from Part 1 of ISO 3166, the second so far uses Tag Latin Small Letters A through Z (U+E0061..E007A) and U+E007F with the Waving Black Flag U+1F3F4 as a common base. Only the former are of interest here.

For each region code in ISO 3166-1, a parameter named `independent` is specified. Either CLDR or UTS#51 should explicitly make support optional for those listed as `independent=no`, just like is already done for "deprecated" codes. This is especially relevant for emoji input methods that are used to select a single flag out of hundreds. I am proposing to keep the independent states as `idStatus="regular"` and split off the rest as territories with the new status "dependent".

Currently, UTS#51 just warns that some region sequences for territories may not have an official flag of their own. It might be wise to add territories and subdivisions in sub-menus like skin colors or genders in some implementations, but this UX detail is probably out of scope of this specification.

Current text in UTS#51 Annex B: Flags

- The valid region sequences are specified by Unicode region subtags as defined in [CLDR], with `idStatus="regular"`, `"deprecated"`, or the `"macroregion"`. However, for macroregions, only **UN** and **EU** are valid.
- Deprecated region sequences should not be generated, but may be supported for backward compatibility.
- Macroregion region sequences generally do not have official flags, with the exception of the **UN** and **EU**.

Some region sequences represent countries (as recognized by the United Nations, for example); others represent territories that are associated with a country. Such territories may have flags of their own, or may use the flag of the country with which they are associated. Depictions of images for flags may be subject to constraints by the administration of that region.

Caveats:

- Although a pair of REGIONAL INDICATOR symbols is referred to as an *emoji_flag_sequence*, it really represents a specific region, not a specific flag for that region. The actual flag displayed for the pair may be different on different platforms, for example for territories which do not have an official flag. The displayed flag may change over time as regions change their flags and platforms update their software.
- For some territories (especially those without separate official flags), the displayed flag may be the same as the flag for the country with which they are associated. For more about cases where characters have the same appearance, see UTR #36: Unicode Security Considerations [UTR36].

For additional information see the sub-section on Regional Indicator Symbols in Section 22.10 Enclosed and Square of [Unicode].

Current `region.xml` in CLDR

```
<supplementalData>
  <version number="$Revision$"/>
  <idValidity>
    <id type="region" idStatus="regular">
      <!-- 256 items -->
      AC~G AI AL~M AO AQ~U AW~X AZ BA~B BD~J BL~O BQ~T BV~W BY~Z CA CC~D CF~I
      CK~P CR CU~Z DE DG DJ~K DM DO DZ EA EC EE EG~H ER~T FI~K FM FO FR GA~B GD~I
      GL~N GP~U GW GY HK HM~N HR HT~U IC~E IL~O IQ~T JE JM JO~P KE KG~I KM~N KP KR
      KW KY~Z LA~C LI LK LR~V LY MA MC~H MK~Z NA NC NE~G NI NL NO~P NR NU NZ OM PA
      PE~H PK~N PR~T PW PY QA RE RO RS RU RW SA~E SG~O SR~T SV SX~Z TA TC~D TF~H
      TJ~O TR TT TV~W TZ UA UG UM US UY~Z VA VC VE VG VI VN VU WF WS XK YE YT ZA ZM
      ZW
    </id>
    <id type="region" idStatus="macroregion">
      <!-- 34 items -->
      001~3 005 009 011 013~5 017~9 021 029 030 034~5 039 053~4 057 061 142~3
      145 150~1 154~5 419 EU EZ QO UN
    </id>
    <id type="region" idStatus="deprecated">
      <!-- 12 items -->
      AN BU CS DD FX NT QU SU TP YD YU ZR
    </id>
    <id type="region" idStatus="private_use">
      <!-- 38 items -->
      AA QM~N QP~T QV~Z XA~J XL~Z
    </id>
    <id type="region" idStatus="unknown">
      <!-- 1 item -->
      ZZ
    </id>
  </idValidity>
</supplementalData>
```

Dependent regions in ISO 3166-1

- [Full list](#) (as [Markdown](#) or as [tab-separated values](#))

English	Alpha-2	Alpha-3	Numeric
Anguilla	AI	AIA	660
Antarctica	AQ	ATA	010
American Samoa	AS	ASM	016
Aruba	AW	ABW	533
Åland Islands	AX	ALA	248
Saint Barthélemy	BL	BLM	652
Bermuda	BM	BMU	060
Bonaire, Sint Eustatius and Saba	BQ	BES	535
Bouvet Island	BV	BVT	074
Cocos (Keeling) Islands	CC	CCK	166
Cook Islands	CK	COK	184
Curaçao	CW	CUW	531
Christmas Island	CX	CXR	162
Western Sahara	EH	ESH	732
Falkland Islands (Malvinas)	FK	FLK	238
Faroe Islands	FO	FRO	234
French Guiana	GF	GUF	254
Guernsey	GG	GGY	831
Greenland	GL	GRL	304
Guadeloupe	GP	GLP	312

English	Alpha- 2	Alpha- 3	Numeric
South Georgia and the South Sandwich Islands	GS	SGS	239
Guam	GU	GUM	316
Hong Kong	HK	HKG	344
Heard Island and McDonald Islands	HM	HMD	334
Isle of Man	IM	IMN	833
British Indian Ocean Territory	IO	IOT	086
Jersey	JE	JEY	832
Saint Martin	MF	MAF	663
Macao	MO	MAC	446
Northern Mariana Islands	MP	MNP	580
Martinique	MQ	MTQ	474
Montserrat	MS	MSR	500
New Caledonia	NC	NCL	540
Norfolk Island	NF	NFK	574
Niue	NU	NIU	570
French Polynesia	PF	PYF	258
Saint Pierre and Miquelon	PM	SPM	666
Pitcairn	PN	PCN	612
Puerto Rico	PR	PRI	630
Palestine	PS	PSE	275
Réunion	RE	REU	638

English	Alpha-2	Alpha-3	Numeric
Saint Helena, Ascension and Tristan da Cunha	SH	SHN	654
Svalbard and Jan Mayen	SJ	SJM	744
Sint Maarten	SX	SXM	534
Turks and Caicos Islands	TC	TCA	796
French Southern Territories	TF	ATF	260
Tokelau	TK	TKL	772
Taiwan	TW	TWN	158
United States Minor Outlying Islands	UM	UMI	581
British Virgin Islands	VG	VGB	092
US Virgin Islands	VI	VIR	850
Wallis and Futuna	WF	WLF	876
Mayotte	YT	MYT	175

Differences between ISO and CLDR

CLDR currently has 256 `regular` codes, but ISO has only 249 “officially assigned codes”. That’s because Unicode adds 6 codes that are exceptionally reserved by ISO 3166/MA and 1 private-use code (`xx`). The former group can safely be considered `dependent` as well.

English	Alpha-2	Alpha-3	Numeric
Ascension Island	AC	—	—
Clipperton Island	CP	—	—
Diego Garcia	DG	—	—
Ceuta, Melilla	EA	—	—

English	Alpha-2	Alpha-3	Numeric
Canary Islands	IC	—	—
Tristan da Cunha	TA	—	—
Kosovo	XK	—	—

Proposed changes to `region.xml`

Replace `<id type="region" idStatus="regular">...</id>` by an equivalent of the following XML code:

```
<id type="region" idStatus="dependent">
  <!-- 53 items -->
  <!-- (AC) AI AQ AS AW AX BL BM BQ BV CC CK (CP) CW CX (DG) (EA) EH FK FO GF
GG GL GP GS GU HK HM (IC) IM IO JE MF MO MP MQ MS NC NF NU PF PM PN PR PS RE
SH SJ SX (TA) TC TF TK TW UM VG VI WF (XK) YT -->
  AC AI AQ AS AW AX BL~M BQ BV CC CK CP CW CX DG EA EH FK FO GF~G GL GP GS GU
HK HM IC IM IO JE MF MO~Q MS NC NF NU PF PM~N PR~S RE SH SJ SX TA TC TF TK TW
UM VG VI WF XK YT
</id>
<id type="region" idStatus="regular">
  <!-- 196 items -->
  <!-- AD AE AF AG AL AM AO AR AT AU AZ BA BB BD BE BF BG BH BI BJ BN BO BR
BS BT BW BY BZ CA CD CF CG CH CI CL CM CN CO CR CU CV CY CZ DE DJ DK DM DO DZ
EC EE EG ER ES ET FI FJ FM FR GA GB GD GE GH GI GM GN GQ GR GT GW GY HN HR HT
HU ID IE IL IN IQ IR IS IT JM JO JP KE KG KH KI KM KN KP KR KW KY KZ LA LB LC
LI LK LR LS LT LU LV LY MA MC MD ME MG MH MK ML MM MN MR MT MU MV MW MX MY MZ
NA NE NG NI NL NO NP NR NZ OM PA PE PG PH PK PL PT PW PY QA RO RS RU RW SA SB
SC SD SE SG SI SK SL SM SN SO SR SS ST SV SY SZ TD TG TH TJ TL TM TN TO TR TT
TV TZ UA UG US UY UZ VA VC VE VN VU WS YE ZA ZM ZW -->
  AD~G AL~M AO AR AT~U AZ BA~B BD~J BN~O BR~T BW BY~Z CA CD CF~I CL~O CR CU~V
CY~Z DE DJ~K DM DO DZ EC EE EG ER~T FI FJ FM FR GA~B GD~E GH~I GM~N GQ~R GT
GW GY HN HR HT~U ID IE IL IN IQ~T JM JO JP KE KG~I KM~N KP KR KW KY~Z LA~C LI
LK LR~V LY MA MC~E MG~H MK~N MR MT~Z NA NE NG NI NL NO~P NR NZ OM PA PE PG~H
PK~L PT PW PY QA RO RS RU RW SA~E SG SI SK~O SR~T SV SY~Z TD TG~H TJ TL~O TR
TT TV TZ UA UG US UY~Z VA VC VE VN VU WS YE ZA ZM ZW
</id>
```

Proposed text for UTS#51 Annex B: Flags

- The valid `region+emoji flag` sequences are specified by Unicode region subtags as defined in [CLDR], with `idStatus="regular"`, `"deprecated"`, `"dependent"`,

"unknown", or "macroregion". However, for macroregions, only **UN** and **EU** are valid.

- ~~Deprecated region and dependent emoji flag sequences~~ should not be generated, but may be supported for backward compatibility.
- ~~Macroregion region emoji flag~~ sequences generally do not have official flags, with the exception of the **UN** and **EU**.
- The unknown emoji flag sequence should not be displayed with the same glyph as an invalid or unsupported emoji flag sequence or flag emoji tag sequence.

~~Some region~~Most regular emoji flag sequences represent countries (as recognized by the United Nations, for example); ~~others~~most dependent emoji flag sequences represent territories that are associated with a country. Such territories may have flags of their own, or may use the flag of the country with which they are associated. (...)

References

- [Unicode Emoji \(UTS#51\)](#)
- [ISO 3166-1 code elements database](#)
- [UN M.49](#)
- [CLDR Regions XML data](#)