

**ccNSO Study Group on the use of Emoji in
Second Level Domains -
Report and Findings**

Table of Contents

Background	3
2. Evolution of domain names which include emoji	5
Definition of emoji	5
History of domain names which include emoji	6
Registering a domain name which include an emoji	8
Registries offering domain names which include emoji	8
Registrars offering domain names which include emoji	9
3. Emoji domain names – Issues and Consideration	10
Recommendations of the SAC 095 Advisory:	10
Observations:	13
Conclusions: TBD	13
Annex A – Request for Information	14
Annex B – History of Emoji	15
Annex C – Implementation of Emoji	19
Annex D – Details regarding registries which accept the registration of domain names which include emoji	21
Composition of second level domain names accepted by registries which register domain names which include emoji	21
WHOIS and Registration of domain names which include emoji	23
Registration Policies/Terms of Use and similar information for domains which include emoji	25
Annex E - Glossary	26

Executive Summary

(TBC at the end)

1. Background

Since 2001 there has been community interest in the use of emoji in domain names and some country code Top Level Domains (ccTLDs) currently allow domain names with emoji to be registered at the second level.

The SSAC analyzed the use of emoji for domain names and published the findings in the SAC 095 advisory¹ on 25 May 2017. The SAC 095 advisory recommends not allowing the use of emoji in TLDs, and discourages their use in a domain name in any of its labels. The SSAC also advises registrants of domain names with emoji that such domains may not function consistently or may not be universally accessible.

On 2 November 2017² the ICANN Board requested:

“ The Board recognizes that mandating labels beyond the top level is out of the policy remit of the ccNSO. However, in this case the ccNSO could play a role in promoting the use of standards developed by the technical community at the Internet Engineering Task Force (IETF) for the secure, stable and interoperable use of Internet identifiers, similar to what it did in the case of discouraging the use of wildcards. Thus, the Board requests that the ccNSO and GNSO engage with the SSAC to more fully understand the risks and consequences of using a domain name that includes emoji in any of its labels, and inform their respective communities about these risks. Furthermore, the Board requests that the ccNSO and GNSO integrate conformance with IDNA2008 and its success or into their relevant policies so as to safeguard security, stability, resiliency and interoperability of domain names.”

Per the Board request the study group was established on 26 February 2018 by the ccNSO Council³ to provide it and the ccTLD community a comprehensive overview of the issues associated with the use of emoji in second level domains, and the need for and current practice by ccTLD managers to allow emojis as second level domains. If considered appropriate by the Study group, the Study group could advise on a course of further action.

¹ <https://www.icann.org/en/system/files/files/sac-095-en.pdf>

² <https://www.icann.org/resources/board-material/resolutions-2017-11-02-en#1.e>

³ <https://ccnso.icann.org/en/announcements/announcement-26feb18-en.htm>

The Study group will⁴:

- Summarise the issues associated with the use of emoji as second level domains as identified by SSAC in its reports (see references);
- Liaise with SSAC to seek further clarification and input if considered needed and appropriate by the group, for example to better understand the threats that the registration and use of emoji as second level domains create for Internet users and related issues;
- Liaise with relevant departments of ICANN to seek further clarification and input if considered needed and appropriate by the group, for example to better understand the threats that the use of emoji create for Internet users and related issues; (Paul Hoffman, Sarmad GDD, Soave GDD)
- Provide an overview of the references in the Fast Track Implementation Plan and Draft overall IDN ccTLD policy to IDNA 2008 and its successor, which disallow emoji. If not included suggest a course of action to include it in the overall draft IDN ccTLD policy.
- Liaise with the ccTLDs who are currently allowing registration of emoji to solicit their views and perspective on possible threats and security issues and provide an overview from their perspective;
- If deemed appropriate and necessary prepare a session at a ccNSO meeting, to solicit views of the community and to present and discuss the results of the study to the ccTLD community;
- Provide a final report of its findings to the ccNSO Council.

The study group may also undertake other activities the members deem to be appropriate for the purpose of the study group.

⁴ <https://ccnso.icann.org/sites/default/files/field-attached/emoji-sld-purpose-scope-activities-26feb18-en.pdf>

2. Evolution of domain names which include emoji

Definition of emoji⁵

Unicode defines emoji as:

Emoji are pictographs (pictorial symbols) that are typically presented in a colorful cartoon form and used inline in text. They represent things such as faces, weather, vehicles and buildings, food and drink, animals and plants, or icons that represent emotions, feelings, or activities.

.....

Emoji may be represented internally as graphics or they may be represented by normal glyphs encoded in fonts like other characters. These latter are called emoji characters for clarity. Some Unicode characters are normally displayed as emoji; some are normally displayed as ordinary text, and some can be displayed both ways.

This definition is unsatisfactory to a number of people for a variety of reasons which include that it fails to provide a complete and definitive explanation of what should be considered an emoji.

Part of the problem in defining what is an emoji is that the Unicode list of emojis is growing rapidly. Unicode version 9.0 added 72 emoji (June 2016), version 10.0 added 56 emoji (June 2017), version 11.0 added 145 emoji (June 2018) while version 12.0 added 72 emoji (February 2019) for a current total of 1,719 emoji⁶ excluding skin tone modifiers.

An additional complexity which is not reflected in the above definition is that not each emoji has a one to one correspondence with a unique Unicode code point. In certain cases several emoji can be amalgamated into a single new emoji by using a Zero Width Joiner (ZWJ). An example of this is joining 🙋 (U+1F469) and 🚀 (U+2708) using a ZWJ (U+200D) which produces 🙋🚀 (U+1F469 U+200D U+2708 U+FE0F). Although the result of this joining is a single emoji, 🙋🚀, which has a unique entry in the Unicode emoji table⁷ it cannot be represented by a single Unicode code point like 🙋 (U+1F469).

⁵ <http://www.unicode.org/reports/tr51/index.html>

⁶ <https://unicode.org/emoji/charts/full-emoji-list.html>

⁷ https://unicode.org/emoji/charts/full-emoji-list.html#1f471_200d_2640_fe0f

Similarly some emoji can also have Skin Tone Modifiers⁸ applied to them which create new emoji. An example of this is 🙄 (U+1F469) which when used in conjunction with the Dark Skin Tone Modifier (U+1F3FF) becomes 🙄🏿 (U+1F469 U+1F3FF). Skin Tone Modifiers also apply to emoji created using ZWJ, an example of this is 🙄👤 (U+1F469 U+200D U+2708 U+FE0F) which when used in conjunction with the Dark Skin Tone Modifier becomes 🙄🏿👤 (U+1F469 U+1F3FF U+200D U+2708 U+FE0F). The use of the 5 Skin Tone Modifiers in Unicode version 12.0 of emoji creates an additional 1,295 emoji⁹.

Another concern is associated with the Symbols class of emoji. This class of emoji includes, as examples, the question mark 🤔 ((U+2753) and the exclamation mark 🗨 (U+2757) but these marks are also part of the Unicode Basic Latin block as ? (U+003F) and ! (U+0021)¹⁰.

Additional concerns arise from the Country-Flags class of emoji. These are based on the ISO3166-1¹¹ list and portions of its Exceptionally Reserved list without any clear explanation as to why the EZ and FX codes which are on the Exceptionally Reserved list are not included. Emojipedia¹² also notes the following regarding emoji country flags:

If the ISO 3166-1 standard was updated to add a new country tomorrow, that would almost certainly end up on the emoji flag list.

Which fails to clearly spell out the procedure for doing this but more importantly does not mention any procedure for removal.

In trying to arrive at an authoritative and all encompassing definition of what is an emoji one may, at least for the moment, have to settle on referring to the latest version of emoji as documented by the Unicode Consortium¹³.

History of domain names which include emoji

The first emoji domain name was registered on April 19, 2001 . The “Definitive Guide to emoji Domains” notes that three domains were registered that day: 🇺🇸.com, 🇸🇬.com,

⁸ http://unicode.org/reports/tr51/#Emoji_Characters

⁹ <https://unicode.org/emoji/charts/full-emoji-modifiers.html>

¹⁰ <https://unicode.org/charts/PDF/U0000.pdf>

¹¹ <https://www.iso.org/iso-3166-country-codes.html>

¹² <https://emojipedia.org>

¹³ <https://unicode.org/emoji/charts/index.html>

🌐.net all of which still resolve. (for a general history on emoji and DNS please see Annex B).

The next major step in the evolution of the use of emoji in domain names came with the adoption of IDNA 2003 to allow for the use of Internationalized Domain Names (IDN) which also had an impact on the use of emoji in domain names. IDNA 2003 adopted the notion of "stored" strings and "non-stored" strings (strings used for querying) where unassigned code points were not to be allowed as "stored" strings. Because of this definition competing interpretations evolved where in one the registration of code points which were unregistered according to IDNA2003 was not valid while in the competing interpretation IDNA2003 did allow encoding and use of unregistered strings (which by definition did not have any mapping). Strings which included such code points were valid for use in the DNS.

IDNA 2008 replaced IDNA 2003 and clearly addressed the use of emoji in domain names by disallowing their use in any labels for the DNS. All registries which had contracts with ICANN, gTLDs, were contractually obliged to respect IDNA 2008 and could therefore not allow the registration of domain names which contained emoji.

Post IDNA 2008 the first domain name registered consisting of a single emoji was 🇱🇦.la (U+1F4A9) in 2011. .LA no longer registers domain names which include emoji but 🇱🇦.la still resolves.

The first reported commercial use of an emoji domain name was in 2015 with Coke launching a major advertising campaign¹⁴ which used 😊.ws (xn--e28h.ws, U+263A) on a standard Coke background (the 😊.ws domain is registered and delegated but is now owned by an Australian photographer).

Currently a number of companies use domain names which contain emoji at the second level such as Budweiser which uses ❤️🍺.ws (xn--qei8618m.ws, U+2764 and U+1F37A) and 🍺🍺🍺.ws (xn--xj8haa.ws, U+1F37A) and Sony Pictures using 😊🎬.ws (xn--dl8h11b.ws, U+1F60A and U+1F3AC).

¹⁴ <https://www.adweek.com/adfreak/coca-cola-spreads-happiness-online-first-emoji-web-addresses-163044/>

Registering a domain name which include an emoji

- **Registries offering domain names which include emoji**

As of the writing of this report the only ccTLDs which could be confirmed or strongly suspected¹⁵ to currently accept the registration of domain names which include emoji were:

1. .AZ (Azerbaijan) - Official registrar offered to register test strings for the price of registration.
2. .CF (Central African Republic) – test registration of a domain name which include emoji completed.
3. .FM (Micronesia) - Registry clearly advertises it supports the registration of domain names which include emoji.
4. .JE (Jersey) - Official registrar offered to register test strings for the price of registration
5. .GA (Gabon) – Test registration of a domain name which include emoji completed.
6. .GE (Georgia) - Registrar Marcaria.com has confirmed the registry accepts domain names which contain emoji.
7. .GG (Guernsey) - Official registrar offered to register test strings for the price of registration.
8. .GQ (Equatorial Guinea) – Test registration of a domain name which include emoji completed.
9. .ML (Mali) – Test registration of a domain name which include emoji completed.
10. .ST (São Tomé and Príncipe) - Registrar Marcaria.com has confirmed the registry accepts domain names which contain emoji.
11. .SU (former Soviet Union cc) - Suspected of supporting domain names which include emoji. Registry contacted and awaiting confirmation.
12. .TO (Tonga) - Registry clearly advertises it supports the registration of domain names which include emoji.

¹⁵ The ESG did not dispose of funds to complete the registration process by paying for the registration of test strings. As such the testing process consisted of either registering domain names in registries which support free registrations (.CF, .GA, .GQ, .ML, .TK) or finding advertising which promoted the registration of domain names which include emoji in a registry (.FM, .TO and .WS) or checking if a registry WHOIS supported the test string which include an emoji and then initiating the registration of that test string in that registry via a registrar. If the registry and registrar accepted the test string then the registry was considered to support the registration of domain names which include emoji.

13. .TK (Tokelau) – Test registration of a domain name which include emoji completed.
14. .UZ (Uzbekistan) - Registrar Marcaria.com has confirmed the registry accepts domain names which contain emoji.
15. .VU (Vanuatu) - Registry has confirmed it supports the registration of domain names which contain emoji.
16. .WS (Western Samoa) - Registry clearly advertises it supports the registration of domain names which include emoji.

Further information on these registries can be found in Annex D – Details regarding registries which accept the registration of domain names which include emoji

- **Registrars offering domain names which include emoji**

Not all registrars accept the registration of domain names which contain emoji (Google domains does not seem to accept registration services for any ccTLD which accepts domain names which include emoji). Additionally there is little standardization amongst those registrars which do accept the registration of domains which contain emoji as to how to search for and present results of such searches.

It is probably important to point out that even if a registrar offers the registration of domain names which include emoji their search processes do not, in most cases, verify which registries they represent support the registration of domains which contain emoji. As an example searching for 📎 (xn--xy8h, U+1F587 - linked paperclips) in GODADDY.COM results in offers for registration in 28 registries including .NET, .INFO, and .ORG which cannot accept the registration of domain names which include emoji as they are gTLDs. Similarly on HOVER.COM a search for the same punycode string, xn--xy8h, returns availability in 62 registries (mostly gTLDs) including .IN and .ORG neither of which support the registration of domain names with emoji. When Hover customer service was contacted about this they advised that their application will allow potential customers to complete the registration application process but will cancel the request for the registration after it is submitted – a number of registrars supporting the registration of domain names which contain emoji seem to follow a similar approach.

3. Emoji domain names – Issues and Consideration

Recommendations of the SAC 095 Advisory:

- **Recommendation 1:** Because the risks identified in this Advisory cannot be adequately mitigated without significant changes to Unicode or IDNA (or both), the SSAC recommends that the ICANN Board reject any TLD (root zone label) that includes emoji.
- **Recommendation 2:** Because the risks identified in this Advisory cannot be adequately mitigated without significant changes to Unicode or IDNA (or both), the SSAC strongly discourages the registration of any domain name that includes emoji in any of its labels. The SSAC also advises registrants of domain names with emoji that such domains may not function consistently or may not be universally accessible as expected.

With respect to Recommendation 1 the ccNSO Fast Track Process for IDN ccTLDs adheres to IDNA 2008 which does not allow for the use of emoji in any labels and is therefore aligned with this recommendation. There is no expectation that this should be changed.

With respect to Recommendation 2 the situation is more complex. As noted in the Board resolution:

“ The Board recognizes that mandating labels beyond the top level is out of the policy remit of the ccNSO ”

The previous sections of this report confirms that there are a number of ccTLDs which currently accept the registration of domain names which include emoji.

The Emoji Study Group identified a number of registries which could be accepting the registration of domain names which include emoji and wrote to these to confirm they were in fact accepting domain names with emoji and to obtain more information on their practices concerning the registration of such domains. The Study Group only received a few responses (See Annex A).

In addition to the issues raised by the SSAC in its SAC 095 the following issues should also be considered:

- Implementation of emoji – There are significant variations in the implementation of emoji by the various vendors¹⁶ (see Annex C of this report for examples).
- Registrar support for emoji – as discussed in a previous section of this report results for registrar searches of domain names which contain emoji provide a significant number of False Positives, and in many cases the registrars allow the customer to complete the application process only to have that application

¹⁶ Paper on the rendering of Emoji vs user understanding of this - http://www.brenthecht.com/publications/csew2018_emojiimpact.pdf

rejected because the selected registry does not support domain names which contain emoji (all gTLDs).

- Direction of writing - Everything emoji seems to be based on a left to right writing system and the impacts of using emoji with those languages which use right to left systems seems to remain an open question.
- Cultural, linguistic, generational or religious significance of emoji - Emojis may be standardized via Unicode but the meaning of emojis can vary greatly depending on culture, language, generation and religion. An example of this is the thumbs-up symbol which is a sign of approval in Western culture, however traditionally in Greece and the Middle East it has been interpreted as vulgar and even offensive¹⁷.

Obviously the main concern when considering using emoji in domain names remains confusability which touches on the foundation of domain names. Domain names should be unique identifiers for resources on the Internet.

The issue of confusability is not new to domain names and can be traced back to the registration of popular domains with common spelling errors or “typos” and continues with IDN domain names, in part, because of the similarity of certain Greek and Cyrillic letters to certain Latin ones.

Although it can be argued that confusability of domain names, at many levels, has always been part of the DNS and that the use of IDN (based on IDNA 2008) in the DNS has not broken the Internet why would the inclusion of emoji?

This is partially correct as a general argument but fails in the specific as the potential for confusability is at least several orders or magnitude greater when considering the following factors:

- Glyph similarity – Emoji include 🌱(growing heart, U+1F497) and ❤️(red heart, U+2764) which are two examples of very similar code points. It is true that some Latin letters suffer from some glyph similarity (il) and there are a few other limited cases when considering the Greek and Cyrillic alphabets vs the Latin alphabet but the number of these letter similarities is significantly dwarfed when compared to the glyph similarities in emoji.
- Country-Flags class of emoji - These are based on the ISO3166-1¹⁸ list and portions of its Exceptionally Reserved list. Emojipedia¹⁹ also notes the following regarding emoji country flags:
 - *Vendors aren't required to support all of these flags (Microsoft doesn't support any country flags on Windows, instead showing the two-letter country codes), but generally do support everything in the list for compatibility.*

¹⁷ <http://www.bbc.com/future/story/20181211-why-emoji-mean-different-things-in-different-cultures>







¹⁸ <https://www.iso.org/iso-3166-country-codes.html>

¹⁹ <https://emojipedia.org>

Additionally the Unicode documentation regarding emoji flags states the following under caveats²⁰:

- *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].*

An example of this for Google, Apple, Facebook and other implementers is that the emoji flag for the United States of America (U+1F1FA, U+1F1F8) and the flag for the U.S. Outlying Islands (U+1F1FA U+1F1F2) both use the same image for the flag  which creates multiple code points for the same glyph and becomes an additional factor to consider with respect to confusability.

- Multiple Unicode code points for the same symbol - An example of this is the question mark emoji  ((U+2753, xn--8di), which can be registered as a domain name in certain ccTLD's, and the question mark ? (U+003F) in the Unicode Basic Latin block which cannot. Potentially even more confusing is the minus sign emoji  (U+2796) and the minus sign in the Unicode Basic Latin block - (U+002D) where the minus sign can be used in ASCII only domain names such as Hello-All.CA and is used in the Punycode representation of IDN and emoji characters as with the emoji question mark  ((U+2753) which has a Punycode representation of xn--8di (the minus sign emoji  (U+2796, xn--5fi) is also available for registration as a domain name in certain ccTLD's and .WS is active and currently redirects to an active website).
- Lack of a standard representation of glyphs between implementers (see Annex C). A search for the number of fonts for alphabets generates results from a hundred thousand to incalculable which could be considered significantly more problematic than the current set of emoji producers. Although the variation between font implementations for a given letter is important, the similarity of specific characters between these is very high (probably, in part, through conditioning from reading). This is not necessarily the case for all emoji..
- Evolution of the representation of a given emoji over time for the same implementer (see the "cow face" example in Annex C). (see previous point)
- Variation in Skin Tones - some emoji can also have Skin Tone Modifiers²¹ applied to them which create new emoji. An example of this is  (U+1F469)

²⁰ <http://www.unicode.org/reports/tr51/#Flags>

²¹ http://unicode.org/reports/tr51/#Emoji_Characters

which when used in conjunction with the Dark Skin Tone Modifier becomes 🧑🏿 (U+1F469 U+1F3FF). Fonts can also have colour but **regardless** of the colour of a **specific** character it does not change **meaning**.

- Zero Width Joiner issues - The insertion of an “invisible” character between emoji can compress several emoji into a single new emoji. An example of this is using a ZWJ (U+200D) with 🧑 (U+1F469) and ✈️ (U+2708) produces 🧑✈️ (U+1F469 U+200D U+2708 U+FE0F) .
- Voice recognition applications - As noted by SSAC in its report there is no standard way to verbally pronounce an emoji in any language. This could be a significant issue as voice to text applications, especially in the mobile market, have become a standard feature on most platforms. Failure to have an agreed to standard way of pronouncing or verbally describing what an emoji is could lead to significant confusability issues. This issue also has impacts on applications which assist the visually impaired.
- Direction²² – There is a possibility that emoji will also include the option to have some emoji facing right or left. If this were to be included in emoji modifiers it would significantly increase the potential for confusability.

Observations:

- Potential for confusability with emoji is significant but is currently contained given the small number of registries which accept the registration of domain names which include emoji.
- Some in the emoji domain name industry have proposed Whitelisting as a potential solution to address confusability²³
- In addition to the confusability issues, SSAC has pointed out there is no agreed to way of pronouncing emoji or otherwise representing them in a way that they can be managed by (for example) various aids people with disabilities uses.

Conclusions: TBD

²² <https://www.superlinguo.com/post/130501329351/emoji-deixis-when-emoji-dont-face-the-way-you>

²³ https://medium.com/@Emoji_Domains/ssac-response-d8d2ad6e800c

Annex A – Request for Information

The following registries were sent the letter below to inquire about the registration of domain names which include emoji:

- .CF (Central African Republic)
- .FM (Federated States of Micronesia)
- .GA (Gabon)
- .GG (Guernsey)
- GQ (Equatorial Guinea)
- .JE (Jersey)
- .LA (Laos)
- .ML (Mali)
- .MP (Northern Mariana Islands)
- .ST (São Tomé and Príncipe)
- .TK (Tokelau)
- .TO (Tonga)
- .TV (Tuvalu)
- .UZ (Uzbekistan)
- .VG (British Virgin Islands)
- .VU (Vanuatu)
- .WS (Samoa)

(ccTLD Manager/operator)

Subject: CCNSO Study on the use of emoji's at the second level in ccTLDs.

On 26 February 2018 the ccNSO Council constituted the Emoji Study Group (ESG) to provide it with a comprehensive overview of the issues associated with the use of Emoji in second level domains as well as any current practices by ccTLDs which accept such registrations. For more details please refer to <https://ccnso.icann.org/sites/default/files/field-attached/emoji-sld-purpose-scope-activities-26feb18-en.pdf>

You are receiving this communication as the result of some initial ad-hoc work by the ESG which potentially identified your registry as one which accepts the registration of Emojis as or in second level domains. If you do not accept such registrations we apologize for any inconvenience and would appreciate you advising us of this.

If you do accept such registrations we are seeking your assistance with our study. I would greatly appreciate you forwarding to us via Kimberly.carlson@ICANN.org pointers to relevant public information related to this practice in your registry. This might be a list of accepted Emoji's, any technical specifications or requirements.

The ESG would also welcome any presentations, memoranda or any other material you may deem relevant, and you would wish to share on this subject. The intent is to understand the different approaches and incorporate as many perspectives as possible into the Study Group report.

The ESG's current plan is to complete the first draft of its report for the ccNSO Council by ICANN 64 to be held in Kobe Japan and as such if you wish to provide any information, I kindly request you to do so as soon as possible.

For further information on the study group, please refer to <https://ccnso.icann.org/en/workinggroups/emoji-sld.htm>

Thanking you for your collaboration

Peter Koch

Chair ccNSO-ESG

Annex B – History of Emoji²⁴

²⁴ <https://www.unicode.org/reports/tr51/#Introduction>

Emoji are pictographs (pictorial symbols) that are typically presented in a colorful cartoon form and used inline in text. They represent things such as faces, weather, vehicles and buildings, food and drink, animals and plants, or icons that represent emotions, feelings, or activities.

Emoji on smartphones and in chat and email applications have become extremely popular worldwide. As of March 2015, for example, Instagram reported that “nearly half of text [on Instagram] contained emoji.” Individual emoji also vary greatly in popularity (and even by country), as described in the SwiftKey Emoji Report²⁵. See emoji press page for details about these reports and others.

Emoji are most often used in quick, short social media messages, where they connect with the reader and add flavor, color, and emotion. Emoji do not have the grammar or vocabulary to substitute for written language. In social media, emoji make up for the lack of gestures, facial expressions, and intonation that are found in speech. They also add useful ambiguity to messages, allowing the writer to convey many different possible concepts at the same time. Many people are also attracted by the challenge of composing messages in emoji, and puzzling out emoji messages.

The word emoji comes from the Japanese:

絵 (e ≅ picture) 文字 (moji ≅ written character).

Emoji may be represented internally as graphics or they may be represented by normal glyphs encoded in fonts like other characters. These latter are called emoji characters for clarity. Some Unicode characters are normally displayed as emoji; some are normally displayed as ordinary text, and some can be displayed both ways.

There’s been considerable media attention to emoji since they appeared in the Unicode Standard, with increased attention starting in late 2013. For example, there were some 6,000 articles on the emoji appearing in Unicode 7.0, according to Google News. See the emoji press page for many samples of such articles, and also the Keynote from the 38th Internationalization & Unicode Conference.

²⁵ <https://www.scribd.com/doc/262594751/SwiftKey-Emoji-Report>
<https://www.scribd.com/doc/267595242/SwiftKey-Emoji-Report-Part-Two>

Emoji became available in 1999 on Japanese mobile phones. There was an early proposal in 2000 to encode DoCoMo emoji in the Unicode standard. At that time, it was unclear whether these characters would come into widespread use—and there was not support from the Japanese mobile phone carriers to add them to Unicode—so no action was taken.

The emoji turned out to be quite popular in Japan, but each mobile phone carrier developed different (but partially overlapping) sets, and each mobile phone vendor used their own text encoding extensions, which were incompatible with one another. The vendors developed cross-mapping tables to allow limited interchange of emoji characters with phones from other vendors, including email. Characters from other platforms that could not be displayed were represented with = (U+3013 GETA MARK), but it was all too easy for the characters to get corrupted or dropped.

When non-Japanese email and mobile phone vendors started to support email exchange with the Japanese carriers, they ran into those problems. Moreover, there was no way to represent these characters in Unicode, which was the basis for text in all modern programs. In 2006, Google started work on converting Japanese emoji to Unicode private-use codes, leading to the development of internal mapping tables for supporting the carrier emoji via Unicode characters in 2007 external link.

There are, however, many problems with a private-use approach, and thus a proposal was made to the Unicode Consortium to expand the scope of symbols to encompass emoji. This proposal was approved in May 2007, leading to the formation of a symbols subcommittee, and in August 2007 the technical committee agreed to support the encoding of emoji in Unicode based on a set of principles developed by the subcommittee.

It is important to note that the Unicode Consortium provides character code charts that show a representative glyph (in a black-and-white text presentation), but is not a designer or purveyor of emoji images, nor is it the owner of any of the color images used in Unicode emoji documents and charts, nor does it negotiate licenses for their use. Inquiries for permission to use vendor images should be directed to those vendors, not to the Unicode Consortium. See [Emoji Images and Rights](#).²⁶

The Sample Colored Glyphs columns use a variety of different styles to illustrate some possible presentations. However, the actual presentations on phones and other devices are up to vendors, subject to the considerations in UTR #51, Unicode Emoji.













²⁶ http://www.unicode.org/faq/emoji_dingbats.html

Annex C – Implementation of Emoji









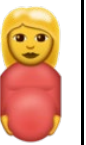

The Unicode Consortium provides character code charts that show a representative glyph (in a black-and-white text presentation), but is not a designer or purveyor of emoji images, nor is it the owner of any of the color images used in Unicode emoji documents and charts, nor does it negotiate licenses for their use. Inquiries for permission to use vendor images should be directed to those vendors, not to the Unicode Consortium. See emoji Images and Rights.²⁷

The Sample Colored Glyphs columns use a variety of different styles to illustrate some possible presentations. However, the actual presentations on phones and other devices are up to vendors, subject to the considerations in UTR #51, Unicode emoji.

As an example of this the well known “smiley face”²⁸ (GRINNING FACE - U+1F600, punycode xn—74h) is currently implemented as follows by various vendors:

Google	Micros oft	Samsun g	WhatsA pp	Twitter	Facebo ok	EmojiO ne	emojide x	Messen ger	LG	HTC	Mozilla
											

For this type of a pictograph, although the variation between the implementations is important, the similarity is fairly high. This is not necessarily the case for all emoji especially those depicting persons. The emoji depicting a pregnant woman²⁹ (PREGNANT WOMAN, U+1F930, punycode xn--pq9h) shows a very high degree of variation in the implementation and has limited similarity between them:









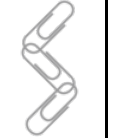

Apple	Google	Microsoft	Samsung	WhatsApp p	Twitter	Facebook	EmojiOne	Emojiped ia	emojindex
									

²⁷ http://www.unicode.org/faq/emoji_dingbats.html




²⁸ *Grinning Face* was approved as part of [Unicode 6.1](#) in 2012 and added to [Emoji 1.0](#) in 2015.

²⁹ *Pregnant Woman* was approved as part of [Unicode 9.0](#) in 2016 and added to [Emoji 3.0](#) in 2016.

It is important to note that limited similarity is not only an issue for emoji representing persons as shown below for linked paperclips (U+1F587, punycode xn--xy8h):

Apple	Google	Microsoft	Samsung	WhatsApp	Twitter	Facebook	EmojiOne	emojidex	LG
									

An additional complication is that vendors do modify emoji over time, sometimes significantly, as shown below for the Facebook emoji representing the face of a cow (U+1F42E, punycode xn--2o8h):

3.0	2.0	1.0
		

Annex D – Details regarding registries which accept the registration of domain names which include emoji

- **Composition of second level domain names accepted by registries which register domain names which include emoji**

The Study Group felt it was important to understand what the registries which register domain names which contain emoji actually accept as domain names. Specifically can the domain names include:






- Multiple emoji
- Emoji which use ZWJ
- Emoji with Skin Tone Modifiers
- Emoji which are Symbols
- Emoji mixed with ASCII characters.
- Emoji mixed with IDN characters.

Given this was not meant to be an exhaustive study of the possibilities the same limited set of test strings were used for each registry tested:

- Multiple emoji - 📎🍺
 - 📎 Linked Paperclips (U+1F587, xn--xy8h)
 - 🍺 Beer Mug (U+1F47A)
 - Punycode = xn--xj8hz6a
- Emoji which use ZWJ - 📎👮
 - 📎 Linked Paperclips (U+1F587)
 - 👮 (U+1F469 U+200D U+2708 U+FE0F)
 - Punycode = xn--3bi9068mrya
- Emoji which use Skin Tone Modifiers - 👮
 - 👮 (U+1F469 U+1F3FF U+200D U+2708 U+FE0F)
 - Punycode = xn--3bi7648mcja
- Emoji which are Symbols - ©, ●
 - © copyright symbol (U+00A9, xn--gba)
 - Geometric ● Black Circle (U+26AB, xn--g8h)
- Emoji mixed with ASCII characters - AbCd📎🍺
 - AbCd📎🍺 (xn--abcd-on53c8ze)
- Emoji mixed with IDN characters - Aéeè📎🍺
 - Aéeè📎🍺 (xn--ae-8iac18294d8tc)

If a registrar offered to register the test string or noted it as already registered it was considered as accepting the test string and this was recorded as a Y in Table 1. All other responses were recorded as a N in the same table.

Annex D - Table 1

Registry	Registrar				©	●	AbCd 	Aéeè 
.AZ	smarthost.az	Y	Y	N	Y	Y	Y	Y
.CF	Freenom.com	Y	Y	N	Y	Y	Y	Y
.FM	nic.fm	Y	Y	Y	Y	Y	Y	Y
.GA	Freenom.com	Y	Y	N	Y	Y	Y	Y
.GE	fastcloud.ge	Y	Y	Y	Y	Y	Y	Y
.GG	eurodns.com	Y	Y	Y	Y	Y	Y	Y
.GQ	Freenom.com	Y	Y	N	Y	Y	Y	Y
.JE	eurodns.com	Y	Y	Y	Y	Y	Y	Y
.ML	Freenom.com	Y	Y	N	Y	Y	Y	Y
.ST	nic.st	Y	Y	Y	Y	Y	Y	Y
.SU	who.su	Y	Y	Y	Y	Y	Y	Y
.TK	Freenom.com	Y	Y	N	Y	Y	Y	Y
.TO	Register.to	Y	Y	Y	Y	Y	Y	Y
.UZ	cp.billur.com	Y	Y	Y	Y	Y	Y	Y

.VU	www.vunic.vu	Y	Y	Y	Y	Y	Y	Y
.WS	Website.ws	Y	Y	Y	Y	Y	Y	Y

- **WHOIS and Registration of domain names which include emoji**

Given identifying the relevant registries and the registrars associated with them required some research this has been collated in the following Table 2 to facilitate referencing these. The column Registry is the ccTLD, Registration Via Registry notes if the registry can perform the registration without having to go through a registrar, Whois/Search is the website used to initially identify if the registry accepts domain names which contain emoji and Registrars lists the registrars listed on the registry website.

Annex D - Table 2

Registry	Registration via registry	WHOIS/search	Registrars
.AZ	N	WHOIS.AZ	NT.AZ (non functional) or SMARHOST.AZ
.CF	N	DOT.CF	FREENOM.COM exclusively
.FM	N	NIC.FM/WHOIS	Multiple such as GoDaddy, Hover and NameCheap
.JE	Y?	channelisles.net	Multiple but not all support the registration of domains which contain emoji.
.GA	N	MY.GA	FREENOM.COM exclusively
.GE	N	REGISTRATION.GE	Multiple (all local?) as REGISTRATOR.GE and GRENA.GE
.GG	Y?	channelisles.net	Multiple but not all support the registration of domains which contain emoji.
.GQ	N	DOMINO.GQ	FREENOM.COM exclusively
.ML	N	POINT.ML	FREENOM.COM exclusively
.ST	Y	NIC.ST	Multiple? None listed on site - 101DOMAIN.COM

.SU	N	WHO.SU	REGISTER.SU Exclusive?
.TO	Y	TONIC.TO	REGISTER.TO exclusively
.TK	N	DOT.TK	FREENOM.COM exclusively
.UZ	N	CCTLD.UZ	Multiple including MACARIA.COM and TLOUD.UZ
.VU	Y	VUNIC.VU/WHOIS	None listed
.WS	Y	WEBSITE.WS	Multiple such as GoDaddy, Hover and NameCheap

- **Registration Policies/Terms of Use and similar information for domains which include emoji**

Table 3 is a summary of the information regarding registration policies and terms of use gathered as part of the testing of registries which support the registration of domain names which include emoji. It is important to note that the .WS registry service provider submitted detailed material related to the registration of emoji which is included in Annex (TBD).

Annex D - Table 3

Registry	Registration policy/TOU Found?	Mention of emoji in policy?	Additional info on emoji	Link
.AZ	N	N	N	WHOIS.AZ
.CF	Y	Y	Y	freenom.com/en/termsandconditions.html
.FM	Y	N	N	get.fm/legal/terms-use/
.JE	Y	N	N	https://www.channelisles.net/legal/terms-and-conditions-1
.GA	Y	N	N	freenom.com/en/termsandconditions.html
.GE	?	?	?	REGISTRATION.GE is only in Georgian
.GG	Y	N	N	https://www.channelisles.net/legal/terms-and-conditions-1
.GQ	Y	N	N	freenom.com/en/termsandconditions.html
.ML	Y	N	N	freenom.com/en/termsandconditions.html
.ST	Y	N	N	nic.st/terms-of-service
.SU	Y	?	?	fid.su/su-eng/docks - Documents in Cyrillic
.TO	Y (FAQ)	N	N	tonic.to/faq.htm
.TK	Y	N	N	freenom.com/en/termsandconditions.html
.UZ	?	?	?	cctld.uz/info/ - Documents in Cyrillic
.VU	Y	N	N	vunic.vu/terms.php
.WS	Y	Y	N	http://website.ws/newdesign/documents/Domain%20Name%

Annex E - Glossary

(Derived and adapted from ICANN's IDN Glossary:

<https://www.icann.org/resources/pages/glossary-2014-02-04-en>)

Term	Definition
ASCII	<p>ASCII (American Standard Code for Information Interchange)</p> <p>ASCII is a common numerical code for computers and other devices that work with text. Computers can only understand numbers, so an ASCII code is the numerical representation of a character such as 'a' or '@'. When mentioned in relation to domain names or strings, ASCII refers to the fact that before internationalization only the letters a-z, digits 0-9, and the hyphen "-", were allowed in domain names.</p>
Domain Name	<p>A unique identifier with a set of properties attached to it so that computers can perform conversions. A typical domain name is "an.example.org". Most commonly the property attached is an IP address, like "192.0.2.42" or "2001:DB8::BA5E:53, so that computers can map the domain name onto an IP address. However the DNS is used for many other purposes. The domain name may also be a delegation, which transfers responsibility of all sub-domains within that domain to another entity.</p>
Emoji	<p>Emoji are pictographs (pictorial symbols) that are typically presented in a colorful cartoon form and used inline in text. They represent things such as faces, weather, vehicles and buildings, food and drink, animals and plants, or icons that represent emotions, feelings, or activities</p> <p>More specifically, an emoji is a Unicode Character representing such a pictorial, represented by a single Unicode Code Point.</p>
Punycode	<p>Punycode is the encoding algorithm described in RFC 3492, was used in IDNA2003 and is in use in IDNA2008. This is the method that together with addition of the ace prefix "xn--" is used to encode IDNs into sequences of LDH ASCII characters in order for applications using the Domain Name System (DNS) to understand and manage the names. The intention is that domain name registrants and users will never see this encoded form of a domain name. The sole purpose is for the internationalized domain names to be backwards compatible with ascii based domain names and that to be able to handle them in existing software not being IDN aware.. For examples see A-label under "IDN".</p> <p>A Punycode encoded label always starts with "xn--". Hence this prefix is</p>

	recommended to be reserved by top-level domain Registries in order to avoid confusion when/if registrations of IDNs are introduced under the respective top level domain.