

Re: UCD Emoji Properties?  
 From: Mark Davis  
 Date: 2015-010-31  
 Draft: <https://goo.gl/ixOS9V>

On the unicode.org list, there was a question as to why the emoji data was not in Unicode properties. That is a valid question, because we have given implementers a structure for relating UCD data files to properties, and how to use them (eg in regex).

Ken answered, then I added:

As Ken said, there's been some preliminary discussion, but we wanted to get initial information out in connection with UTR #51 first, and take more time to consider what UCD properties would look like, and which are necessary.

The basic information that people want to access for implementations are:

1. Is a character emoji or not?
2. Which emoji have default text presentation? (others having emoji presentation)
3. Which emoji are modifiers, and which are modifier bases? (others being neither)
4. Which sequences of emoji are recommended (zwj and/or combining marks) for those who support them?
5. flags and modifier sequences are specified algorithmically, and don't need to be listed.

The levels, the distinction between primary and secondary, and the carrier sources were useful in development of the emoji data and tr51 but aren't really necessary for implementations.

At some point we should consider adding such properties. One cut at how it would look based on the above would be:

Emoji\_Presentation (EP)

- Non\_Emoji (NE)
- Default\_Text (DT)
- Default\_Emoji (DE)
- NA

Emoji\_Modifier\_Status (EMS)

- Emoji\_Modifier (EM)
- Emoji\_Modifier\_Base (EMB)
- NA

(For the sequences, it is probably too early to think beyond having them listed in a file, such as <http://unicode.org/draft/Public/emoji/2.0/emoji-sequences.txt> )

Alternatively, we might consider having these not be UCD properties, documented in <http://unicode.org/reports/tr44/>, but instead have a decentralized approach, and treat these as (say) XUCD properties:

1. Use the same file formats, so that implementers can handle them easily, and make sure that the namespaces don't collide, but document them separately.
2. This would be applicable to not only Public/emoji, but also Public/security, Public/idna, etc.
3. To make this work, we'd have separate XPropertyAliases.txt and XPropertyValueAliases.txt files in each directory, plus the property data files for those, plus some extra documentation that would otherwise go into [tr44](http://unicode.org/reports/tr44/).