

Title: Issues with kSpoofingVariant data
Source: William T. Nelson
Status: Individual contribution
Action: For consideration by UTC
Date: 2019-10-02

L2/19-328

Recommendations for UTC

1. Put action item [160-A15](#) on hold pending revised draft data for `kSpoofingVariant`
2. In the future, UTC should be more careful to ensure proposals are reviewed before taking action

Background

As described in [L2/19-186](#), the Unihan ad hoc group was assigned to:

[G]enerate (and review) a set of data for the `kSpoofingVariant` ... [field] for Unicode 13.0, to be submitted to the UTC in time to be discussed at UTC #160.

Accordingly, [L2/19-282](#) *Proposed Contents for kSpoofingVariants Field* posted to the document register on 2019-07-18. UTC accepted its contents for Unicode version 13.0 at meeting #160, as indicated by action item [160-A15](#).

However, the accepted data is not usable. The issue was raised and acknowledged on the Unihan list in late July.

Commentary

The data issue is an honest mistake, which can be readily corrected using existing comments from Unihan list members.

My greater concern is that proposals presented as work product of Unihan ad hoc may be accepted by UTC without scrutiny. In the case of [L2/19-282](#), the reader is assured the listed variants "can confidently be labeled" as such, but the data is invalid to the degree that I doubt if anyone looked over it prior to acceptance. Unihan list members did not endorse this data set, and the proposal document was submitted to the register too late to ensure public review for UTC #160.

UTC should positively verify that proposals have been reviewed by experts other than their own authors. When verification is not possible, please consider holding proposals for the next meeting, particularly when documents arrive past the submission deadline.

UTC should also encourage that initial proposals for new fields, or major revisions to existing fields, include draft data (when applicable). Proposals are more concrete, and issues more apparent, when changes to draft text and draft data can be evaluated together. When proposals with many data records are submitted as PDF documents, the data should be made available in a text-based format.

Note: [L2/19-281](#) *Proposed Contents for kZVariant Field* has the same issues as [L2/19-282](#). The UTC #160 [meeting minutes](#) do not reference [L2/19-281](#), so I omitted it from the body of this report.