

DOC TYPE: Working Group Document

TITLE: Comments on Tangut proposal N3797

SOURCE: Deborah Anderson and Richard Cook, Script Encoding Initiative, UC Berkeley

STATUS: Liaison member contribution (from SEI/UCB)

ACTION: For consideration by WG2

DISTRIBUTION: ISO/IEC JTC1/SC2/WG 2

SEI welcomes further progress on Tangut, but we request the following:

(1) All LFW 1997 (and 1986) characters should be included in the proposed repertoire. As noted on page 4 of the proposal document N3797, "Li Fànwén's dictionary (LFW1997) was used as the basis for the Mojikyo Tangut character set, which has been widely used for electronic publishing." As such, these characters should be included in the repertoire; they were included in the earlier Tangut proposal (N3521).

(2) The iteration mark should be removed for future study. This character is said to be modern (and is not attested in LFW2008). In comparison, consider also that e.g. U+303B VERTICAL IDEOGRAPHIC ITERATION MARK is not a CJK Unified Ideograph.

(3) The final font containing *only* the final glyphs in their proposed repertoire should be provided for review.

(4) Mappings (from item #3) to N3521. (See <http://linguistics.berkeley.edu/~rscook/UTC/Tangut/> for associated charts and data files.)

(5) Collation data determining the proposed block order should be available as plain text tab-delimited.

(6) The Tangut character names need to be changed. They should be algorithmically derivable from code points. Tangut encoding is supported by a wide variety of source mappings, not only LFW2008. Any non-algorithmically derived names would be "single-source". Such a names list would have to be maintained explicitly, and this would place an undue burden on the editors of the names files, collation input files, and other property data files, particularly for a repertoire of over 6,000 characters.

Given those items, we will be able to prepare multi-column chart for final rounds of proofing, and will not object to proposed block order.