Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation

Doc Type: **Working Group Document**
Title: **Response to N3313**
Source: **Martin Hosken**
Status: **Individual contribution**
Action: **For consideration by JTC1/SC2/WG2**
Date: **2007-09-10**

Introduction

N3313 presents what it considers to be serious problems with the Lanna encoding in FPDAM4. This paper addresses the issues raised in N3313 showing that the current encoding is not deficient.

Script Name

This issue was addressed at the Germany WG2 meeting at which a resolution was passed on the name.

Character Names

The character names are based on the script name and follow them.

Alignment Issue

There is indeed a naming alignment problem in FPDAM4. With the removal of the two characters (ᩢ)LANNA VOWEL SIGN AM and (ᩣ)LANNA VOWEL SIGN TALL AM, the names have been adjusted to account but not the glyphs. Thus the two glyphs for AM should be removed and all the other glyphs in the range 1A65..1A7D shifted backwards two codes. This will realign the glyphs and the names.

AE name change

In keeping with the name of (ᩮ)1A71 LANNA VOWEL SIGN AE, (ᩖ)1A54 should be renamed to LANNA LETTER AE.

Character Ordering

The relative order of character groups within a script block is somewhat arbitrary. In this case the block order follows the Northern Thai sorting order. But this makes no difference to actual sorting. Chinese based Tai Lue sorting cannot be achieved without text transformation because finals take precedence over vowels. So a default order is given that at least is close to one language's sorting convention (if it can be called a convention).

The relative order within the consonant group is clear across all the languages using the script and the only disagreement between them is that some languages do not include certain letters or

spell them differently. Therefore the the listed characters are in the right order and the corrected names as per FPDAM4 are correct.

Missing Characters

1A28

1A28 and 1A4B are both in the alphabet list for Northern Thai. It is only in Tai Lue where the (ᦨ)HIGH CHA (U+1A28) is spelled differently.

1A72

The issue here is whether U+1A72 LANNA VOWEL SIGN THAM AI is necessary. It is used in Khuen and in the archaic old Tham Lao orthography where it contrasts with (ᩥ)U+1A71 LANNA VOWEL SIGN AI.

1A55

Visually, 1A55 is correctly analysed as being representable by (ᩣ)1A4E LANNA LETTER A followed by (᩠)1A58 LANNA CONSONANT SIGN MEDIAL RA. But while this sequence would never occur as a consonant medial sequence, it would require special handling for it to be interpreted as a vowel. Instead a separate code is used for the letter vowel in keeping with other letter vowels in the group.

1A5D

The claim is made that 1A50, which could be analysed as being visually representable as a stylistic variant of (ᩃ)1A3E LANNA LETTER LOW PA ()1A60 LANNA SIGN SAKOT (ᩆ)1A26 LANNA LETTER NGA. But for the same reasons as for 1A55, a separate code is used for the vowel letter. N3313 claims that it could be encoded using U+1A5D, but this is not the case as will be made more evident in the discussion of additional letters. U+1A5D is used solely for the LOW PA preceding in a a ligature.

Additional Characters

1A7C, 1A7D

These are resolved by tidying up the FPDAM4 code chart and glyphs.

The characters with codes listed in parentheses are codes from the proposed new character table. All other codes are considered to be taken from FPDAM4.

(1A5E)

This ligature from the proposed chart is represented in the encoding in FPDAM4 as the ligature sequence (ᩈ)1A34 LANNA LETTER RANA ()1A60 LANNA SIGN SAKOT (ᩈ)1A34 LANNA LETTER RANA which is linguistically what it is, a special rendering of these two characters in subjoined relationship.

(1A60)

A first visual examination of this character seems to imply that it is the same kind of ligature involved (ᩈ)1A34 LANNA LETTER RANA, but this sequence doesn't occur. So instead the

analysis implies that this would be stored as (ᨪ)1A40 LANNA LETTER LOW PHA 1A5D LANNA CONSONANT SIGN LOW PA. Linguistically the order is actually reversed and the storage order wants to be 1A5D 1A40. But in order to conform to the Unicode convention of a diacritic always following its base character the order is reversed to 1A40 1A5D. Since this sequence only occurs in one word and that very rarely (the author has only seen one example of the use of the character) it was felt that the compromise was acceptable. Further discussion of this character can be found in N3207 section 4.

(1A61)

This syllable is encoded in the FPDAM4 encoding as (ᨬ)1A4C LANNA LETTER HA ()1A60 LANNA SIGN SAKOT (ᩊ)1A42 LANNA LETTER LOW YA 1A6F LANNA VOWEL SIGN OY.