

2013-11-22 Siddham Script (梵字) Meeting @ Tokyo, JAPAN, Earth

(Last Update: 2013-12-04 13:30 PST)

Attendees: Deborah Anderson (via Skype), Lee Collins, Bill Eidson, Andrew Glass (via Skype), Shoken HARADA (原田聖賢), Taichi KAWABATA (川幡太一), Ken Lunde, Koju MOTOYAMA (元山公寿; morning only), Kiyonori NAGASAKI (永崎研宣), Anshuman Pandey, Michel Suignard (morning only), Toshiya SUZUKI (鈴木俊哉), Taro YAMAMOTO (山本太郎)

Meeting Time: 10:00–17:00

Meeting Report (Lunde)

As a neutral party, it was very clear that the right people attended this meeting, and if this meeting didn't happen, the progress that we made could have taken months or years. And, as usual, having the meeting face-to-face made all the difference in the world. I am thankful that everyone took the time out of their busy schedules to participate.

While the agenda was helpful in guiding the meeting, some items were skipped, mainly because everyone realized what the important points were, such as the criteria for encoding character variants and the status of the six character variants in ISO/IEC 10646 Fourth Edition, and the meeting focused on them.

At the very beginning of the meeting, in order to eliminate any possible confusion, Michel Suignard provided details about the current status of the Siddham script in ISO/IEC 10646. The standard Siddham characters (U+11580 through U+115B5 and U+115B8 through U+115C9) are in Amendment 2 of Third Edition (equivalent to Unicode Version 7.0), and are considered a done deal (frozen). The section marks and character variants are in Fourth Edition, which is undergoing its last technical ballot, and is expected to be finalized during the February 2014 WG2 meeting.

As background, WG2 N4294, which is the original Siddham script proposal (Pandey) states that Siddham is an Indian script that is no longer used in India. We learned during the meeting that the user community is approximately 12 million people, 10 million of which are in Japan. The rest are primarily in China and Korea. Japan also coined digits for Siddham, and Korea has unique Siddham forms.

WG2 N4407R (Japan) proposed six Siddham character variants (U+115E0 through U+115E5), which are reflected in ISO/IEC 10646 Fourth Edition. Professor Motoyama stated that there are no more than 10 character variants (when the agreed-upon criteria, which is effectively the same criteria that Japan used to select these first six character variants, is applied), which means that the current Siddham block is of sufficient size to accommodate them in the future. It was also noted (by Japan) that standardized variation sequences cannot be used for the character variants that are combining forms (U+115E4 and U+115E5).

Siddham ligatures were discussed, including the possibility of encoding the high-frequency ones. There was mutual agreement not to do this, and to instead use font features, such as 'liga' (GSUB). WG2 N4490 (Pandey), which proposed a separate block for Siddham logographic forms, was discussed, but there also was mutual agreement not to do this.

To quote Anshuman, it became clear that Siddham needs to be handled from a Pan-Buddhist perspective. This means that each user community will have their own needs, based on their particular usage of the script. Bill's needs are met by the standard Siddham characters, mainly because character variants are handled via separate font resources. Japan's needs will be met by encoding the six character variants that are in ISO/IEC 10646 Fourth Edition.

Part of the difficulty in handling or interpreting Siddham character variants is that their usage is often based on user will, which amounts to a semantic distinction. Also, Bill pointed out that there are known errors in the historical documents that may produce unique forms, and when passing them down, one must decide whether to propagate, correct, or annotate such errors in the process.

One issue that came up, which also comes up in similar meetings, is how to define plain text. In my experience, one way to think about plain text is to open a PDF that includes Siddham content that is likely to be stylized, copy the text (which is done as plain text), then paste it into a text editor, word processor, or comparable application. If any meaning or information is lost in this process, then the plain text representation is insufficient.

Three very important things came out of this meeting:

- Mutual agreement by all attendees was reached on the following three sets of objective criteria for determining whether a Siddham character variant is suitable for encoding:
 1. Both forms are semantically distinct in a logographic context.
 2. Both forms cannot be algorithmically derived by context.
 3. Both forms co-occur in a single source.

There was also mutual agreement that only general documents should be considered valid sources, and those documents that intentionally use character variants for pedagogical purposes should be excluded.

- These criteria were applied to the six Siddham character variants, and there was mutual agreement by all parties that the following four are encodable:

U+115E0 through U+115E3

Japan feels that they can provide sufficient evidence that the combining character variants, U+115E4 and U+115E5, should be encoded.

- Mutual agreement that a joint document be submitted to WG2 to clarify how Siddham character variants should be handled in terms of encodability, including clarification of the six that are in ISO/IEC 10646 Fourth Edition.

Lee Collins raised the following open technical questions with regard to the handling of the character variants that represent vowels:

- Do we need to define a default or canonical representation of the U vowel sign?
- Which character variant is correct for use with RA, and how do we render the RA followed by non-canonical forms?
- What guidelines will we offer for transliteration software between Romanized Sanskrit and Siddham if we introduce unresolvable ambiguities by separate encoding of U?

That is all.

Meeting Agenda

I. Introductions

II. Current status of Siddham block (U+11580–U+115FF)

III. Presentations on perspectives of Siddham encoding

1. Japanese members
2. SBII's approach to translation of Siddham Texts, Mandalas & Bijas (Shuji)
3. A. Pandey

IV. Letter Variants

1. Criteria for selection of variants
 - a. Which variants to include [letter A? Short I Jogon?]
 - b. VOWEL SIGN U and VOWEL SIGN UU [only used with base HA?]
2. How to handle letter variants
 - a. Separately-encoded characters
 - b. Via fonts
3. Location of letter variants
 - a. In a separate block?
 - b. Location in relation to current Siddham block
4. Names of letter variants

V. Demo: Siddham TENCHIJI Font & Transliteration Tools

VI. Other characters worthy of encoding

1. Word ligatures
2. Stroke primitives
3. Pedagogical letters
4. Digits
5. Other symbols
6. Korean and Chinese versions of Siddham

VII. Next steps

1. Confirm property values
2. What to include in ISO/IEC 10646 Fourth Edition

Relevant Documents

- [Kawabata-20131121](#) (*Request to Encode Variant Characters of Siddham Script*, Kawabata et al.)
- [Eidson-20131114](#) (*Comments on N4407R Proposal to Encode Variants for Siddham Script*, Eidson)
- [Eidson-20131114-Appendix](#) (*Siddham Variations Template*, Eidson)
- [WG2 N4507](#) (*Comments on the name of the "Siddham" script (L2/12-011R = N4185)*, Baums & Glass)

- [WG2 N4506](#) (*Comments on naming the "Siddham" encoding, Rajan & Sharma*)
- [WG2 N4490](#) (*A Practical Approach to Encoding Siddham Variants, Pandey*)
- [WG2 N4486](#) (*Comments on N4407R Proposal to Encode Variants for Siddham Script, Glass*)
- [WG2 N4468](#) (*Additional Siddham Variants, Pandey*)
- [WG2 N4467](#) (*Proposal to Encode a Set of Digits for Siddham, Pandey*)
- [WG2 N4460](#) (*Siddham ad hoc report*)
- [WG2 N4459](#) (*Draft additional repertoire for ISO/IEC 10646:2014 (4th edition) [see 11580-115FF on pp 23 & 24], Suignard*)
- [WG2 N4457](#) (*Name changes for Siddham Section marks, Anderson et al.*)
- [WG2 N4407](#) (*Proposal to Encode Variant Forms for Siddham Script, Kawabata et al.*)
- [WG2 N4391](#) (*Additional expert feedback on Siddham Section marks, Eidson*)
- [WG2 N4378](#) (*Additional Information on Siddham Section Marks (N4336), Anderson*)
- [WG2 N4369](#) (*Feedback on Siddham proposal (WG2 N4294), Eidson*)
- [WG2 N4361](#) (*Feedback on Siddham proposal (WG2 N4294), Suzuki*)
- [WG2 N4336](#) (*Proposal to Encode Section Marks for Siddham in ISO/IEC 10646, Pandey*)
- [WG2 N4294](#) (*Proposal to Encode the Siddham Script in ISO/IEC 10646, Pandey*)
- [WG2 N4185](#) (*Preliminary Proposal to Encode Siddham in ISO/IEC 10646, Pandey*)