TO: UTC L2/15-204
FROM: Deborah Anderson, Ken Whistler, Rick McGowan, Roozbeh Pournader, and Laurentiu Iancu
SUBJECT: Recommendations to UTC #144 July 2015 on Script Proposals
DATE: 25 July 2015

The recommendations below are based on documents available to the members of this group at the time they met, and do not include documents submitted later to the document registry.

SOUTH ASIA
Indic
1. Vedic
a) VEDIC SIGN ATIKRAMA
Document: L2/15-160 Proposal to encode 1CF7 VEDIC SIGN ATIKRAMA – Sharma

Discussion: We reviewed this document, which requests one character in the Vedic Extensions block. The character is shown contrastively with 1CE1 VEDIC TONE ATHARVAVEDIC INDEPENDENT SVARITA on page 2 of the proposal. The code point seems reasonable.

Recommendations: We recommend the UTC approve this character, after discussion.

b) Updates to Devanagari Sama Vedic annotations and informative aliases
Documents: L2/15-164 Updates to Devanagari Sama Vedic informative aliases – Sharma
L2/15-203 Feedback on "Updates to Devanagari Sama Vedic informative aliases" – Scharf

Discussion: We reviewed Sharma's document, which requests edits to the informative aliases and annotations in the Vedic Extensions block, modification to sub-headers in the Vedic Extensions block, and changes to the informative aliases in the Devanagari Extended block. We note that in general, annotations are intended to identify characters, not act as an alternative glossary with detailed information for Sama Vedic mark-up. However, existing outright errors should be corrected.

Based on the comments from Peter Scharf (L2/15-203), who worked on the original Vedic proposals, we recommend that scholars working on these materials collaborate and submit a document when agreement has been reached.

Recommendations: We recommend the UTC review these documents and send feedback to the authors.

2. Devanagari
a) COMBINING DEVANAGARI SIGN AVAGRAHA
Document: L2/15-162 Script property of A8F1 COMBINING DEVANAGARI SIGN AVAGRAHA – Sharma

Discussion: We reviewed this document, which requested changing the script property of U+A8F1 COMBINING DEVANAGARI SIGN AVAGRAHA from "Devanagari" to "Inherited". The document also recommends changes to the names list indicating that the character is used across multiple scripts.

According to UAX 24 §2.1, "if it becomes established that a character is regularly used with more than one script, it will be assigned the Common or Inherited Script property value". However, it is difficult to determine whether U+A8FI "regularly" is used in scripts outside Devanagari.

Based on the evidence provided in L2/15-162, Devanagari appears to be the predominant script, and provides identity and stability for the character.

At this point, we recommend retaining “Devanagari” as the script property for this character, but adding Bengali to the set of Script Extensions values for it.".

Recommendations: We recommend the UTC add Bengali to the set of Script Extensions values for U+A8F1 COMBINING DEVANAGARI SIGN AVAGRAHA.

b) COMBINING DEVANAGARI DIGIT SIX
Document: L2/15-163 Concerning A8E6 COMBINING DEVANAGARI DIGIT SIX – Sharma

Discussion: We reviewed this document, which requested no action by the UTC, but calls into question the original description of the character in the proposal for Vedic characters (L2/17-343), which described it as marking an atisvārya tone (the sixth) in Sāma Gāna.

Recommendations: We recommend the UTC note this feedback.

c) DEVANAGARI SIGN SPACING ANUSVARA
Document: L2/15-178 Proposal to encode A8FE DEVANAGARI SIGN SPACING ANUSVARA – Sharma

Discussion: We reviewed this document which requested the encoding of one character, U+A8FE DEVANAGARI SIGN SPACING ANUSVARA.

Because we find no evidence of the character participating in the script’s syllabic structure, our preference is to represent this mark with the existing character U+0966 DEVANAGARI DIGIT ZERO, which is the second option listed in §5.

Recommendations: We recommend the UTC discuss this proposal. If use of U+0966 DEVANAGARI DIGIT ZERO is recommended, we suggest the character be annotated accordingly (i.e., “used for an anusvara after digits indicating secondary svara-s in Samavedic”).

3. Bengali
a) BENGALI LETTER VEDIC ANUSVARA
Document: L2/15-161 Proposal to encode 09CF BENGALI LETTER VEDIC ANUSVARA – Sharma

Discussion: We reviewed this document which proposes the character U+09CF BENGALI LETTER VEDIC ANUSVARA. The proposal provides clear examples of usage, and notes that parallel characters are already encoded in Devanagari, Grantha, and Kannada.

The proposed location, U+09CF, fills a hole at the bottom of a middle column in the Bengali block. We recommend instead U+09FC, at the end of the Bengali block but close to the visually similar ISSHAR character (U+09FA), so those creating fonts will be be able to harmonize these two similar-looking glyphs.

Recommendations: We recommend the UTC review this proposal, accept the character, and discuss the location. Our recommendation is to place BENGALI LETTER VEDIC ANUSVARA at U+09FC.

b) BENGALI ABBREVIATION SIGN
Document: L2/15-172 Proposal to Encode an Abbreviation Sign for Bengali – Srinidhi and Sridatta

Discussion: We reviewed this document, which requested an abbreviation sign for Bengali. The authors noted that the sign was already proposed in 2010 (L2/10-029), but not as a separate character, only as part of a "Doctor" sign. Anshuman Pandey recommended a separate encoding of a Bengali abbreviation sign (L2/10-050), comparable to the abbreviation signs for other scripts (Devanagari, Gujarati, Kaithi, etc.).

The examples are clear and support the request.

Rather than put this sign in the middle of the block, however, we suggest it be placed in the last column at U+09FD (after BENGALI LETTER VEDIC ANUSVARA, above). (Cf. Devanagari and Gujarati, where the abbreviation sign is located in the last column.)

Recommendations: We recommend the UTC approve this character, but discuss its location. Our suggestion for BENGALI ABBREVIATION SIGN is U+09FD.

4. Malayalam
MALAYALAM VOWEL SIGN VOCALIC RR
Document: L2/15-174 Request to change the representative glyph of MALAYALAM VOWEL SIGN VOCALIC RR – Srinidhi

Discussion: We reviewed this request to change the glyph for U+0D44 MALAYALAM VOWEL SIGN VOCALIC RR. The evidence provided in support of the change appears to be solid.

It was noted that the glyph in the charts is in current fonts (Lohit, Kartika, and Nirmala, for example).
Current glyph: Requested glyph change is to:

The original proposal (L2/05-309r2) for this character gave the shape currently in the code charts as its representative glyph. The proposal summary form identified Rachana Akshara Vedi as the glyph source (for the proposal).[1]

Recommendations: We recommend the UTC make the glyph change, based on the evidence provided.

5. Kannada

[1] The exact source for the chart glyph was not able to be determined during the script ad hoc meeting. However, at the 2011 WG2 meeting in Busan, South Korea, Ireland had requested use of the Rachana font (see page 17 of http://www.unicode.org/L2/L2011/11100-n3903-wg2-minutes.pdf), and this was accepted to in principle, so it could be the source of the current chart glyphs.

KANNADA SIGN SPACING CANDRABINDU
Document: L2/15-158 Attestations for Sama Vedic usage of 0C80 KANNADA SIGN SPACING CANDRABINDU – Sharma

Discussion: We reviewed this document, which provides evidence from academic published sources for U+0C80 KANNADA SIGN SPACING CANDRABINDU in Samavedic texts, improving on the examples in L2/14-166. Sharma's examples demonstrate that the character has two variants, one with a stroke underneath, and one without a stroke. Sharma queried scholars and discovered they prefer the glyph without the stroke, which is the glyph shape in the current DAM2 ballot.

Recommendations: We recommend the UTC note this feedback, and consider adding an annotation for Unicode 9.0 on the use of U+0C80 (i.e., it appears in the Badaga orthography and Samavedic texts).

6. Tamil
TAMIL FRACTIONS
Documents:
L2/15-176 Encoding of Tamil fractions and special symbols in Tamil block and a new Tamil Supplement block – Gov't of India / Manoj Jain
L2/15-185 Recommendations on Tamil Fractions and special symbols in Unicode - Postponement of decision – Ramachandran

Discussion: We reviewed these two documents on Tamil fractions and symbols. The characters referred to in the documents are contained in the current DAM2 ballot. The two documents do not contain any technical comments.

Recommendations: We recommend the UTC discuss these documents, and decide whether to make DAM2 ballot comments.

EUROPE
7. Cyrillic
TWO LETTERS FOR MONTENEGRIN LANGUAGE
Document: L2/15-169 Addition of two letters from Montenegrin language, CYRILLIC script – Gov't of Montenegro

Discussion: We reviewed this document, which asks for four precomposed Cyrillic characters–two letters, each in uppercase and lowercase–for the Montenegrin language.

The language can be written in Latin or Cyrillic. Montenegrin Cyrillic replaced Serbian Cyrillic in 2009, according to Wikipedia. The two proposed precomposed characters are not found in Serbian, but were represented as digraphs in both Latin and Cyrillic.

Although the precomposed characters are not eligible for atomic encoding in Unicode, this document provides useful input for CLDR on the changing orthography of Montenegrin.

Recommendations: We recommend the UTC review the document, and the information be funneled to CLDR.

MIDDLE EAST

8. Syriac

SYRIAC LETTERS FOR GARSHUNI MALAYALAM

Documents: L2/15-166 Feedback on the Alphabetization of the Syriac letters for Garshuni (Suryani) Malayalam – Perczel

Discussion: We reviewed this document.

Recommendations: We recommend the UTC note this input for Unicode 9.0.

INDONESIA

9. Old Makassarese Script

Documents: L2/15-179 Proposal to Encode the Old Makassarese Script – Pandey
(replaces L2/15-100 Preliminary Proposal to Encode the Makassarese Bird Script)

Discussion: We reviewed this document, which is a proposal for the Old Makassarese script, an historical script that was used in Indonesia for the Makassar language.

Specific comments:

- The name seems reasonable.
- We agree the model should be the same as Buginese; both scripts are used to represent the same language, Makassar.
- We do not recommend including the last character, OLD MAKASSARESE END OF TEXT, which is a decorative version of Arabic text and can be used next to Arabic text.
- The proposal could potentially add *hīr* (top of page 11), the Arabic word for the Gregorian era, as a character
- We suggest the range allocation for the script be moved to where 3 columns are available, such as after Khotanese, from 11ED0..11EFF.
- What is the origin of the numbers? The shapes appear to be unique enough to justify encoding, so an additional column appears warranted. Details on the use of numbers with Arabic should be investigated further.

Recommendations: We recommend the UTC review this proposal and send comments to the author.

AFRICA

10. Medefaidrin

Document*:* L2/15-117R Preliminary proposal for encoding the Medefaidrin (Oberi Okaime) script in the SMP of the UCS (Revised) – Rovenchak

Discussion: We reviewed this revised proposal, which addresses comments made in the May 2015 script ad hoc recommendations (L2/15-149). We understand that members of the user community will be reviewing the proposal during the week of the UTC, and we await their feedback.

Recommendations: We recommend the UTC review this proposal, and take into consideration any new input from the user community, if available.

11. Mandombe

Document: L2/15-118 R Preliminary proposal for encoding the Mandombe script in the SMP of the UCS (Revised) – Rovenchak

Discussion: We reviewed this revised preliminary proposal, which has made changes based on comments from the May 2015 script ad hoc report (L2/15-149). The proposal now provides a better explanation of how the script works.

After studying the proposal, we suggest a different encoding model for this script: instead of encoding graphical script elements, we recommend syllable units be encoded, as has been done for many of the other African scripts (Mende Kikakui, Bamum, Vai, etc.). This approach will make rendering easy. Once a count on the number of characters has been established, it can be given a slot on the Roadmap alongside the other big African syllabaries.

Other comments:

- Remove MANDOMBE SPACE SEPARATOR, since it is a SPACE (U+0020)
- Remove MANDOMBE WORD SEPARATOR, and recommend use of U+002E FULL STOP
- The character 1xxx0 MANDOMBE SYMBOL MANDOMBE (page 11) is not a proper character, and should be removed.
- In place of the ELLIPSIS character, recommend use of three MANDOMBE DOT characters
- On copyright: Additional information is still required. Copyright claims on particular documents, books, pamphlets, etc. is not the concern here, and owners of such documents are not being asked to drop their claims of copyright ownership. The only issue would be if a copyright claim has been made on the Mandombe script.
- Provide more details on the numbers of actual users, i.e., how many users of the script are in Africa and how many in other areas of the world?

Recommendations: We recommend the UTC review this proposal, and send the authors feedback.

SYMBOLS
12. Currency Symbols
LARI SIGN
Document: L2/15-168 The Lari Symbol: Implementation Principles and Supplementary Manual -- National Bank of Georgia / Giorgi Shermazanashvili

Discussion: We reviewed this document, which provides guidance on the glyph for the LARI SIGN. The current glyph in the charts is:

20BE

The implementation principles and manual (L2/15-168) gives the following shape as reflecting the symbol's basic structure and principles (but provides other acceptable forms on page 5):

Recommendations: We recommend the UTC discuss this document.

++

Previous recommendations (carried over for script and character proposals not yet discussed in the UTC)

MIDDLE EAST
13. Arabic
Document: L2/15-087 Proposal to encode one Arabic honorifics - Lateef Sagar Shaikh

Discussion: We reviewed this request for one Arabic honorific, similar to the proposal for seventeen honorifics by Roozbeh Pournader (L2/14-147).

The proposal requests one character, ARABIC LIGATURE KARAM ALLAHU WAJHULKARIM. However, the evidence presented does not show the character as a combined ligature. Also, no evidence is provided of the character in a font.

Recommendations: We recommend the UTC respond to the author, and request he provide more attestation of the ligature in plain text and evidence of the character in a font.

14. Hebrew Nomina Sacra
Document: L2/15-092 Typographic Concerns and the Hebrew Nomina Sacra - Shoulson

Discussion: We reviewed this preliminary document, which gives a selection of examples showing how the Tetragrammaton appears in Hebrew, but whose representations cannot be handled by current Unicode characters.

Based on the document, it appears that the encoding of two characters would make it possible to represent many of the examples. The characters are (with the following approximate glyphs):

(1)  (2) 

Recommendations: We recommend the UTC discuss this proposal and respond to the author, based on the comments above and UTC discussion.

NUMBER SYSTEMS
15. Siyaq Number Mark
Document: L2/15-074 Siyaq Number Mark - Pandey

Discussion: We reviewed this request for one character, U+ 061D ARABIC SIYAQ NUMBER MARK, which appears over numbers and indicates that the numbers belong together as a set.

The script ad hoc recommended the proposal include wording describing what characters it appears on top of, such as: “This character appears over any sequence of Siyaq numbers, from any Siyaq block, and full stop”. Additionally, can the author specify the maximum number of characters over which it can extend? Does it ever go across a word break? (If so, provide evidence.)

Recommendations: We recommend the UTC discuss this proposal and approve U+ 061D ARABIC SIYAQ NUMBER MARK, after modifications based on the comments above and discussion in the UTC.

--- CARRIED OVER FROM OLDER MEETINGS ---

EAST AND CENTRAL ASIA

16. Small Seal Script

L2/14-242 Proposal to encode Small Seal Script – TCA and China

Discussion: We reviewed this proposal, which proposes 799 characters out of a projected 10,516. In our opinion, the proposal is still far from mature, and would benefit from coordinating work with experts in the U.S. and Japan in order to formalize mapping data, which is needed to evaluate a final proposal. The proposal should also provide demonstrated need for including the script in the international standard.

Recommendations: We recommend the UTC members review this proposal and consider sending the authors the comments above.

17. Naxi Dongba

L2/14-241 Supplement on Proposal for Encoding Naxi Dongba Pictograph Script (L2/11-178) - China
L2/14-245 Feedback on Naxi Dongba Supplement document - Anderson

Discussion: We reviewed the "Supplement" document, which answered questions posed at the June 2011 WG2 meeting in Helsinki, Finland (see Naxi Dongba Ad Hoc report, L2/11-244). Specifically, the authors in the "Supplement" confirmed that the encoding is for modern use, not traditional use of the characters, and that alphabetical ordering is preferred.

The "Feedback" document posed additional questions and made suggestions. During WG2 discussion, the Naxi Dongba proposal authors stated the script is both a logography and syllabary, and the variation shown in some glyphs is due to regional differences, but only one glyph per character is warranted in the encoding. They agreed to revise the proposal and provide information on the proposed characters, with glyphs, Romanized transcription, Chinese gloss (and English translation) and references.

Recommendations: We recommend the UTC members review this proposal and send comments to the authors.

18. Shuishu

L2/14-243 Proposal for encoding Shuishu – China

Discussion: We reviewed this proposal, which is still at an early stage. In our view, it is not yet clear that Shuishu is an encodable writing system. In order to move forward, we recommend the authors prepare and publish a standard sign list for Shuishu, which can then be circulated for review by other scholars and gain scholarly support. The next version of the proposal should also provide a rationale for the digital representation of their sign list, answering the question why these shapes should be put into an international character encoding standard.

Recommendations: We recommend the UTC members review this proposal and send comments to the

authors. The UTC may want to relay the suggestions to the authors above, regarding recommended next steps.

19. Khitan Large Script
L2/14-234 Proposal on Encoding Khitan Large Script – China
L2/14-233 Preliminary Review of Proposal on Encoding Khitan Large Script – West
L2/14-246 Ad hoc reports for Tangut and Khitan Large Script – Anderson

Discussion: We reviewed these documents. As noted in L2/14-233, the Khitan Large Script is largely undeciphered without any character list or recent dictionaries, vocabulary lists, or secondary linguistic materials, so the current proposal should be viewed as preliminary.

Also as mentioned in L2/14-233, the script appears to have a significant percentage of characters (18%) that are either Han borrowings or identical in shape to already encoded CJK ideographs. A revised proposal should discuss the pros and cons of unifying those Khitan Large Script characters with CJK characters already encoded: what are the costs/benefits to unification? Because Khitan Large Script is an historical script, the security risk would not arise if Khitan Large Script used CJK characters, only if it encoded a large set of identical CJK characters.

Additionally, we suggest the proposal also create a "Uni-Khitan" database (or spreadsheet) to document sources.

Recommendations: We recommend the UTC members discuss these documents.

20. Ranjana
L2/09-192 Preliminary proposal for encoding the Rañjana script in the SMP (WG2 N3649)
L2/14-221 Comparison between Ranjana Proposals - Anderson
L2/13-243 Proposal to Encode Ranjana Script - Manandhar
L2/14-253 Recommendations to UTC from Script Meeting in Nepal - Anderson

Discussion: We discussed these documents. Since decisions on the repertoire and encoding model for Ranjana depend upon those for "Nepaalalipi", discussion on Ranjana was limited. It was noted that a future Ranjana proposal should also discuss the unification with Wartu and Lanydza, and should provide details on any specific characters and behaviors of the script in Tibet and other locations outside Nepal.

Recommendations: We recommend the UTC review the document, but postpone discussion until after the "Nepaalalipi" encoding is resolved.

21. Bhujinmola
L2/14-253 Recommendations to UTC from Script Meeting in Nepal
L2/14-283 Introducing the Bhujinmol Script - Pandey

Discussion: We briefly discussed the section in the "Recommendations" on Bhujinmola. Bhujinmola has a characteristic wavy headline (see examples in "Roadmapping the Scripts of Nepal" L2/09-325). The question on whether Bhujinmola represents a stylistic variation of "Nepaalalipi" or should be separately encoded needs to be discussed in a separate document, with examples of how vowels and consonants join differently from "Nepaalalipi" and other rendering issues.

Recommendations: We recommend the UTC review the document, but wait for further research to support separately encoding Bhujinmola. (Note: The script ad hoc did not yet review L2/14-283 Introducing the Bhujinmol Script by Pandey.)