ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
Please fill all the sections A, B and C below.
(Please read Principles and Procedures Document for guidelines and details before filling this form.)
See http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html **for latest *Form*.**
See http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html **for latest *Principles and Procedures* document.**
See http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest roadmaps.**

A. Administrative

1. **Title:** **_ Proposal to WG2 for Encoding Syloti Nagri Script in the BMP _____**

2. Requester's names: INCITS/L2; UTC; Peter Constable (SIL International); James Lloyd-Williams and Sue Lloyd-Williams (Sylheti Translation And Research, London); Advocate Shamsul Islam Chowdhury (Chairman, Sylot Academy, Sylhet, Bangladesh); Professor Asaddar Ali (Vice Chairman, Sylot Academy, Sylhet, Bangladesh); Mohammed Sadique; Matiar Rahman Chowdhury (Chairman, Sylot Academy (UK and Europe), London)

3. Requester type (Member body/Liaison/Individual contribution): Member (L2), Liaison (UTC), expert (others)

4. Submission date: 2003-06-12

5. Requester's reference (if applicable): (This proposal is a revision of L2/03-151R.)

6. (Choose one of the following:)
This is a complete proposal: Yes
or, More information will be provided later: See also the accompanying document, "Documentation in support of proposal for encoding Syloti Nagri in the BMP" (= L2/02-388).

B. Technical - General

1. (Choose one of the following:)
   a. This proposal is for a new script (set of characters): Yes
      Proposed name of script: Syloti Nagri
   b. The proposal is for addition of character(s) to an existing block: Yes (1 character)
      Name of the existing block: General Punctuation

2. Number of characters in proposal: 45

3. Proposed category (see section II, Character Categories): Category A

4. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000): Level 2 or 3
   Is a rationale provided for the choice? Yes
      If Yes, reference: Syloti Nagri script contains combining diacritics.

5. Is a repertoire including character names provided? Yes
   a. If YES, are the names in accordance with the 'character naming guidelines in Annex L of ISO/IEC 10646-1: 2000? Yes
   b. Are the character shapes attached in a legible form suitable for review? Yes

6. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard? Sue Lloyd-Williams (Sylheti Translation And Research)
   If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used: Sue Lloyd-Williams (Sylheti Translation And Research), < james@jslw.fsworld.co.uk>
   _____
   _____
   _____

7. References:
   a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? Yes (see accompanying document, "Documentation in support of proposal for encoding Syloti Nagri in the BMP")
   b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? Yes (see accompanying document, "Documentation in support of proposal for encoding Syloti Nagri in the BMP")

8. Special encoding issues:
   Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?
   Yes: it discusses conjunct formation and control of display of a virama-like character. See also the

[1] Form number: N2352-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09)

accompanying document, “Documentation in support of proposal for encoding Syloti Nagri in the BMP”

9. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see http://www.unicode.org/Public/UNIDATA/UnicodeCharacterDatabase.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard. See section D below.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? No (not to WG2; it has been submitted to and approved by UTC)

If YES explain _____

2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? Yes

If YES, with whom? Submitters (other than L2 and UTC) are members of the user community in England and Bangladesh.

If YES, available relevant documents: See accompanying document, “Documentation in support of proposal for encoding Syloti Nagri in the BMP”

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? The user community consists of sizeable communities in Sylhet region of Bangladesh, in Calcutta, in England and elsewhere.

Reference: See accompanying document, “Documentation in support of proposal for encoding Syloti Nagri in the BMP”

4. The context of use for the proposed characters (type of use; common or rare) Limited but increasing usage with renewed interest. Publishing in the script was known to have been done by several presses in Bangladesh and Calcutta as recently as the 1970s. More recently, no metal-type facilities have been available, but development of digital type has resulted in renewed use and growing interest.

Reference: See accompanying document, “Documentation in support of proposal for encoding Syloti Nagri in the BMP”

5. Are the proposed characters in current use by the user community? Yes

If YES, where? Reference: In London and Birmingham, and in Sylhet, possibly elsewhere. See accompanying document, “Documentation in support of proposal for encoding Syloti Nagri in the BMP.”

6. After giving due considerations to the principles in *Principles and Procedures document* (a WG 2 standing document) must the proposed characters be entirely in the BMP? Yes

If YES, is a rationale provided? Yes: living script

If YES, reference: See accompanying document, “Documentation in support of proposal for encoding Syloti Nagri in the BMP.”

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? Yes, with the exception of one character, as noted below. (Note: the code positions shown in this proposal have been approved by UTC.)

8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? No

If YES, is a rationale for its inclusion provided? _____

If YES, reference: _____

9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? No

If YES, is a rationale for its inclusion provided? _____

If YES, reference: _____

10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? No. Syloti Nagri has similarities with other North Indic scripts, but is distinct; it is no more like any other Indic scripts already in the UCS than are any of them to one another. Thus, for purposes of this proposal, these characters cannot be considered similar to existing characters.

If YES, is a rationale for its inclusion provided? _____

If YES, reference: _____

11. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)? Yes

If YES, is a rationale for such use provided? Yes

If YES, reference: See accompanying document, “Documentation in support of proposal for encoding Syloti Nagri in the BMP.”

Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? _____

If YES, reference: _____

12. Does the proposal contain characters with any special properties such as control function or similar semantics? Yes

If YES, describe in detail (include attachment if necessary) See discussion in section D below, and also in the accompanying document, “Documentation in support of proposal for encoding Syloti Nagri in the BMP.”

13. Does the proposal contain any Ideographic compatibility character(s)? No

If YES, is the equivalent corresponding unified ideographic character(s) identified? _____

If YES, reference: _____

D. Proposed Characters

As discussed in section II.6 of the accompanying document, “Documentation in support of proposal for encoding Syloti Nagri in the BMP” (= L2/02-388), the basic collation order for Syloti Nagri is well established, but there are attested variations in ordering in relation to a few details. In the following chart, characters are ordered in what we believe to be the most preferable ordering.

As discussed in section II.1.6 of the accompanying document, “Documentation in support of proposal for encoding Syloti Nagri in the BMP” (= L2/02-388), there is very limited attestation in manuscripts of a set of Syloti Nagri digits. The evidence is not sufficient to support a proposal to encode Syloti Nagri digits at this time. It is conceivable that a proposal for Syloti Nagri digits may become justified at some future point if further manuscripts are discovered, in which case an additional column of characters would be required. We have no reason to anticipate this happening in the near future, however.

The code chart is presented on a new page. Note that there are existing characters in the 205 column that are not shown.

Code Chart:

	A80	A81	A82
0	ꠀ	ꠐ	ꠤ
1	ꠁ	ꠑ	ꠥ
2	ꠂ	ꠒ	ꠦ
3	ꠃ	ꠓ	ꠧ
4	ꠄ	ꠔ	꠨
5	ꠅ	ꠕ	꠩
6	꠆	ꠖ	꠪
7	ꠇ	ꠗ	꠫
8	ꠈ	ꠘ	꠬
9	ꠉ	ꠙ	꠭
A	ꠊ	ꠚ	꠮
B	ꠋ	ꠛ	꠯
C	ꠌ	ꠜ	
D	ꠍ	ꠝ	
E	ꠎ	ꠞ	
F	ꠏ	ꠟ	

205
*

Character Names:

U+A800 SYLOTI NAGRI LETTER A
U+A801 SYLOTI NAGRI LETTER I
U+A802 SYLOTI NAGRI SIGN DVISVARA
U+A803 SYLOTI NAGRI LETTER U
U+A804 SYLOTI NAGRI LETTER E
U+A805 SYLOTI NAGRI LETTER O
U+A806 SYLOTI NAGRI SIGN HASANTA
= halant, virama
U+A807 SYLOTI NAGRI LETTER KO
U+A808 SYLOTI NAGRI LETTER KHO
U+A809 SYLOTI NAGRI LETTER GO
U+A80A SYLOTI NAGRI LETTER GHO
U+A80B SYLOTI NAGRI SIGN ANUSVARA
U+A80C SYLOTI NAGRI LETTER CO
U+A80D SYLOTI NAGRI LETTER CHO
U+A80E SYLOTI NAGRI LETTER JO
U+A80F SYLOTI NAGRI LETTERJHO
U+A810 SYLOTI NAGRI LETTER TTO
U+A811 SYLOTI NAGRI LETTER TTHO
U+A812 SYLOTI NAGRI LETTER DDO
U+A813 SYLOTI NAGRI LETTER DDHO
U+A814 SYLOTI NAGRI LETTER TO
U+A815 SYLOTI NAGRI LETTER THO
U+A816 SYLOTI NAGRI LETTER DO
U+A817 SYLOTI NAGRI LETTER DHO
U+A818 SYLOTI NAGRI LETTER NO
U+A819 SYLOTI NAGRI LETTER PO
U+A81A SYLOTI NAGRI LETTER PHO
U+A81B SYLOTI NAGRI LETTER BO
U+A81C SYLOTI NAGRI LETTER BHO
U+A81D SYLOTI NAGRI LETTER MO
U+A81E SYLOTI NAGRI LETTER RO
U+A81F SYLOTI NAGRI LETTER LO
U+A820 SYLOTI NAGRI LETTER RRO
U+A821 SYLOTI NAGRI LETTER SO
U+A822 SYLOTI NAGRI LETTER HO
U+A823 SYLOTI NAGRI VOWEL SIGN A
U+A824 SYLOTI NAGRI VOWEL SIGN I
U+A825 SYLOTI NAGRI VOWEL SIGN U
U+A826 SYLOTI NAGRI VOWEL SIGN E
U+A827 SYLOTI NAGRI VOWEL SIGN OO
U+A828 SYLOTI NAGRI POETRY MARK-1
U+A829 SYLOTI NAGRI POETRY MARK-2
U+A82A SYLOTI NAGRI POETRY MARK-3
U+A82B SYLOTI NAGRI POETRY MARK-4
U+A82C..U+A82F (These positions shall not be used.)

U+2055 FLOWER PUNCTUATION MARK
= phul, puspika
- used with Syloti Nagri, Bengali and other Indic scripts

Unicode character properties

With the exception of the characters listed below, we propose that all characters have a General Category property of “Lo”, a Canonical Combining Class property of 0, a Bidirectional Class property of “L”, a Mirrored property of “N”, an East Asian Width property of “N”, and a Line Breaking property of “AL”. In the information provided below, only those values that differ from the defaults just mentioned are listed.

Characters	Properties
U+A806	General Category = “Mn”, Canonical Combining Class = 9, Bidi Class = NSM, Line Breaking = “CM”
U+A802, U+A80B, U+A826	General Category = “Mn”, Canonical Combining Class = 230, Bidi Class = NSM, Line Breaking = “CM”
U+A823, U+A824, U+A827	General Category = “Mc”, Canonical Combining Class = 0, Line Breaking = “CM”
U+A825	General Category = “Mn”, Canonical Combining Class = 220, Bidi Class = NSM, Line Breaking = “CM”
U+A828..U+A82B	General Category = “Po”, Line Breaking = “AL”
U+2055	General Category = “Po”, Line Breaking = “QU”

The canonical combining classes assigned to characters of general category Mc (other than hasanta / virama) do not match typical practice for Indic scripts (which would be 0). We propose these classes as they better deal with the behaviours of the script and are less likely to lead to problems in implementation and use. For instance, the DVISVARA and VOWEL SIGN U can co-occur, but these combining marks do not typographically interact, so alternate orderings of the characters cannot result in visible distinction and hence cannot have any distinct information value. By putting these two combining marks in distinct, non-zero combining classes, these two orderings become canonically equivalent.

Additional notes

The accompanying document “Documentation in support of proposal for encoding Syloti Nagri in the BMP” (= L2/02-388) proposes a particular model for control of conjunct formation and of the display overt hasanta (virama). This particular aspect of that document is revised in this proposal. The current proposal is to use an encoding model like that used for Myanmar script: U+A806 SYLOTI NAGRI HASANTA is used to determine formation of a conjunct of the preceding and following characters, in which case the HASANTA has no direct visual display. If, however, the HASANTA is followed by ZWNJ, no conjunct is formed, and the HASANTA becomes visible.