Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type: **Working Group Document**
Title: **Proposal for encoding the Sundanese script in the BMP of the UCS**
Source: **Michael Everson**
Status: **Individual Contribution**
Action: **For consideration by JTC1/SC2/WG2 and UTC**
Date: **2006-01-09**

The Sundanese script, or *aksara Sunda*, is used for writing the Sundanese language, one of the languages of the island of Java in Indonesia. It is a descendent of the ancient Brahmi script of India, and so has many similarities with modern scripts of South Asia and Southeast Asia which are also members of that family. The script has official support and is taught in schools and is used on road signs. The SIL Ethnologue indicates that there are 27,000,000 Sundanese.

Sundanese has been written in a number of scripts. *Pallawa* or *Pra-Nagari* was first used in West Java to write Sanskrit from the fifth to eighth centuries CE, and from Pallawa was derived *Sunda Kuna* or *Old Sundanese* which was used in the Sunda Kingdom from the 14th to 18th centuries. Both Javanese and Arabic script were used from the 17th to 19th centuries and the 17th to the mid-20th centuries respectively. Latin script has had currency since the 20th century. The modern Sundanese script, called *Sunda Baku* or *Official Sundanese*, was made official in 1996. The modern script itself was derived from Old Sundanese, the earliest example of which is the Prasasti Kawali stone (see Figure 1).

Structure
Consonants bear the inherent vowel, which vowel matras may modify. The explicit PAMAAEH or *virama* is used where there is no inherent vowel; this does not cause Brahmic consonant clustering. Initial consonants can be followed by the medial consonants ◌ᮡ -YA, ◌ᮢ -RA, and ◌ᮣ -LA: ᮊᮡ *kya*, ᮊᮢ *kra*, ᮊᮣ *kla*.

Dependent vowel signs are used in a manner similar to that employed by other Brahmic scripts:

ka	ᮊ	ki	ᮊᮤ	ku	ᮊᮥ	kè	ᮊᮦ
ko	ᮊᮧ	ke	ᮊᮨ	keu	ᮊᮩ	k	ᮊ᮪

Independent vowels are used as in other Brahmic scripts as well:

a	ᮃ	i	ᮄ	u	ᮅ	è	ᮆ
o	ᮇ	e	ᮈ	eu	ᮉ		

Three final consonants also found:

kang	ᮊᮀ	kar	ᮊᮁ	kah	ᮊᮂ

Two supplemental letters have been added recently, ᮮ KHA and ᮯ SYA. These are used to represent sounds denoted by the Arabic letters خ KHA and ش SHEEN. The word خاشع *xušu* 'state of grace' was written ᮠᮥᮞᮥ *husu* before these letters were created; now it can be written ᮮᮥᮯᮥ *khusyu*.

Rendering

The vowel signs, consonant signs, and finals in Sundanese follow the consonant to which they are applied. One of them is rendered before the consonant: *kè*; three of them are drawn below the consonant: *ku*, *kra*, *kla*; four of them are rendered after the consonant: *ko*, *kya*, *kah*, *k*; five of them are drawn above the consonant: *ki*; *ke*; *keu*, *kang*, *kar*. When more than one sign appears above or below a consonant, the two are drawn side-by-side: *king*, *kru*, *koh*, *klu*. The backing store for *kyur* is phonetic: KA + -YA + -U + -R.

Digits and punctuation

Digits have distinctive forms. For punctuation, Sundanese uses European marks like COMMA, FULL STOP, and QUESTION MARK.

Ordering

The arrangement of characters in the code table here follows the Brahmic ordering, for transparency to implementors of ISO/IEC 10646 and Unicode. The *ha-na-ca-ra-ka* order found in Javanese and Balinese does not seem to have currency in Sundenese. Two Brahmic orderings do have some currency, however. One of these inserts modern additions to the character set into the order according to place of articulation and lettershape. The principle followed here is the same principle that was applied in the ordering of Balinese; it also follows the way the characters are taught (as in the children's primer *Ngalagena*, 2002).

a > *i* > *u* > *ae* > *o* > *e* > *eu* >
ka > *kha* > *qa* > *ga* > *nga* > *ca* > *ja* > *za* > *nya* >
ta > *da* > *na* > *pa* > *fa* > *va* > *ba* > *ma* >
ya > *ra* > *la* > *wa* > *sa* > *xa* > *sya* > *ha*

The other order is based on a separation of the Old Sundanese letters from the additional letters. The former are given in standard Brahmic order, and the latter are then given in the order of their Latin transliteration (though SYA and KHA are given out of order). This is a recommendation of H. R. Hidayat Suryalaga, given at a script symposium held 1997-10-27 at Padjadjaran University, Bandung:

a > *i* > *u* > *ae* > *o* > *e* > *eu* >
ka > *ga* > *nga* > *ca* > *ja* > *nya* >
ta > *da* > *na* > *pa* > *ba* > *ma* >
ya > *ra* > *la* > *wa* > *sa* > *ha* >
sya > *kha* > *fa* > *qa* > *va* > *xa* > *za*

In searching operations, PANGLAYAR should be equivalent to RA, so that word-forms like *batur* 'another person' and *batureun* 'a sad person who needs company' will be found together. The same applies to two other characters in Sundanese: PANYECEK should be equivalent to NGA, so that word forms like *monong* 'ape' and *monongeun* 'act like an ape' will be found together; and PANGWISAD should be equivalent to HA, so that word forms like *panah* 'arrow' and *panaheun* 'target for an arrow (such as a deer)' will be found together.

Linebreaking

Opportunities for hyphenation occur after any full orthographic syllable. A syllable is structured (and represented in the backing store) as C(Y)(R)(L)(V)(F); the initial consonant "C" may be followed by -YA (Y) or -RA (R) or -LA (L), by an optional vowel matra (V), and optionally by one of the finals PANYECEK -*ṅ*, PANGLAYAR -*r*, or PANGWISAD -*h* (F).

Implementations

Dian Tresna Nugraha implemented Sundanese in an eight-bit font in 2005, together with a keyboard layout design (see Figure 6).

Unicode Character Properties

```
1B80;SUNDANESE SIGN PANYECEK;Mn;0;NSM;;;;;N;;anusvara;;;
1B81;SUNDANESE SIGN PANGLAYAR;Mn;0;L;;;;;N;;repha;;;
1B82;SUNDANESE SIGN PANGWISAD;Mc;0;L;;;;;N;;visarga;;;
1B83;SUNDANESE LETTER A;Lo;0;L;;;;;N;;;;;
1B84;SUNDANESE LETTER I;Lo;0;L;;;;;N;;;;;
1B85;SUNDANESE LETTER U;Lo;0;L;;;;;N;;;;;
1B86;SUNDANESE LETTER AE;Lo;0;L;;;;;N;;;;;
1B87;SUNDANESE LETTER O;Lo;0;L;;;;;N;;;;;
1B88;SUNDANESE LETTER E;Lo;0;L;;;;;N;;;;;
1B89;SUNDANESE LETTER EU;Lo;0;L;;;;;N;;;;;
1B8A;SUNDANESE LETTER KA;Lo;0;L;;;;;N;;;;;
1B8B;SUNDANESE LETTER QA;Lo;0;L;;;;;N;;;;;
1B8C;SUNDANESE LETTER GA;Lo;0;L;;;;;N;;;;;
1B8D;SUNDANESE LETTER NGA;Lo;0;L;;;;;N;;;;;
1B8E;SUNDANESE LETTER CA;Lo;0;L;;;;;N;;;;;
1B8F;SUNDANESE LETTER JA;Lo;0;L;;;;;N;;;;;
1B90;SUNDANESE LETTER ZA;Lo;0;L;;;;;N;;;;;
1B91;SUNDANESE LETTER NYA;Lo;0;L;;;;;N;;;;;
1B92;SUNDANESE LETTER TA;Lo;0;L;;;;;N;;;;;
1B93;SUNDANESE LETTER DA;Lo;0;L;;;;;N;;;;;
1B94;SUNDANESE LETTER NA;Lo;0;L;;;;;N;;;;;
1B95;SUNDANESE LETTER PA;Lo;0;L;;;;;N;;;;;
1B96;SUNDANESE LETTER FA;Lo;0;L;;;;;N;;;;;
1B97;SUNDANESE LETTER VA;Lo;0;L;;;;;N;;;;;
1B98;SUNDANESE LETTER BA;Lo;0;L;;;;;N;;;;;
1B99;SUNDANESE LETTER MA;Lo;0;L;;;;;N;;;;;
1B9A;SUNDANESE LETTER YA;Lo;0;L;;;;;N;;;;;
1B9B;SUNDANESE LETTER RA;Lo;0;L;;;;;N;;;;;
1B9C;SUNDANESE LETTER LA;Lo;0;L;;;;;N;;;;;
1B9D;SUNDANESE LETTER WA;Lo;0;L;;;;;N;;;;;
1B9E;SUNDANESE LETTER SA;Lo;0;L;;;;;N;;;;;
1B9F;SUNDANESE LETTER XA;Lo;0;L;;;;;N;;;;;
1BA0;SUNDANESE LETTER HA;Lo;0;L;;;;;N;;;;;
1BA1;SUNDANESE CONSONANT SIGN PAMINGKAL;Mc;0;L;;;;;N;;subjoined ya;;;
1BA2;SUNDANESE CONSONANT SIGN PANYAKRA;Mn;0;NSM;;;;;N;;subjoined ra;;;
1BA3;SUNDANESE CONSONANT SIGN PANYIKU;Mn;0;NSM;;;;;N;;subjoined la;;;
1BA4;SUNDANESE VOWEL SIGN PANGHULU;Mn;0;L;;;;;N;;i;;;
1BA5;SUNDANESE VOWEL SIGN PANYUKU;Mn;0;NSM;;;;;N;;u;;;
1BA6;SUNDANESE VOWEL SIGN PANAELAENG;Mc;0;L;;;;;N;;ae;;;
1BA7;SUNDANESE VOWEL SIGN PANOLONG;Mc;0;NSM;;;;;N;;o;;;
1BA8;SUNDANESE VOWEL SIGN PAMEPET;Mn;0;NSM;;;;;N;;e;;;
1BA9;SUNDANESE VOWEL SIGN PANEULEUNG;Mn;0;L;;;;;N;;eu;;;
1BAA;SUNDANESE SIGN PAMAAEH;Mc;9;L;;;;;N;;virama;;;
1BAE;SUNDANESE LETTER KHA;Lo;0;L;;;;;N;;;;;
1BAF;SUNDANESE LETTER SYA;Lo;0;L;;;;;N;;;;;
1BB0;SUNDANESE DIGIT ZERO;Nd;0;L;;0;0;0;N;;;;;
1BB1;SUNDANESE DIGIT ONE;Nd;0;L;;1;1;1;N;;;;;
1BB2;SUNDANESE DIGIT TWO;Nd;0;L;;2;2;2;N;;;;;
1BB3;SUNDANESE DIGIT THREE;Nd;0;L;;3;3;3;N;;;;;
1BB4;SUNDANESE DIGIT FOUR;Nd;0;L;;4;4;4;N;;;;;
1BB5;SUNDANESE DIGIT FIVE;Nd;0;L;;5;5;5;N;;;;;
1BB6;SUNDANESE DIGIT SIX;Nd;0;L;;6;6;6;N;;;;;
1BB7;SUNDANESE DIGIT SEVEN;Nd;0;L;;7;7;7;N;;;;;
1BB8;SUNDANESE DIGIT EIGHT;Nd;0;L;;8;8;8;N;;;;;
1BB9;SUNDANESE DIGIT NINE;Nd;0;L;;9;9;9;N;;;;;
```

Bibliography

Baidillah, Idin, Cucu Komara, & Deuis Fitni. [2002]. *Ngalagena: Panglengkep Pangajaran Aksara Sunda pikeun Murid Sakola Dasar/Dikdas 9 Taun*. [Bandung]: CV Walatra.

Hardjasaputra, A. Sobana, Tedi Permadi, Undang A. Darsa, & Edi S. Ekadjati. 1998. *Rancangan Pembakuan Aksara Sunda*. Bandung.

Acknowledgements
This project was made possible in part by a grant from the U.S. National Endowment for the Humanities, which funded the Universal Scripts Project (part of the Script Encoding Initiative at UC Berkeley).

Figures

Figure 1. The Prasasti Kawali stone, source of the Sundanese script.
Taken from the Sundanese Wikipedia, **http://su.wikipedia.org/wiki/Prasasti_Kawali**

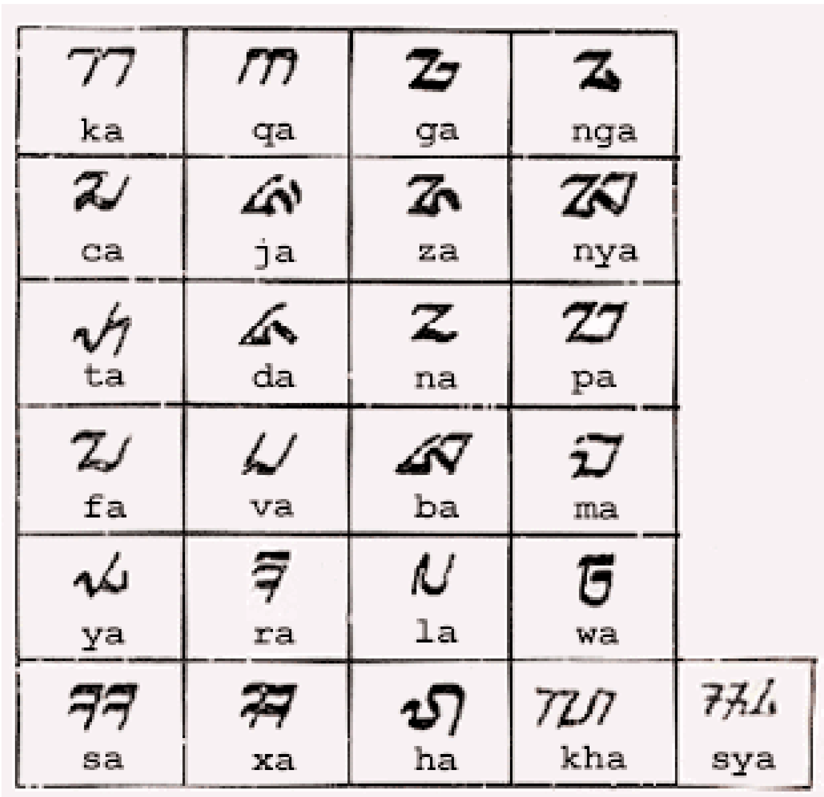


Figure 2. Table of Sundanese consonants, showing the additional letters ᮮ KHA and ᮯ SYA.
The table is given at **http://sundanet.com/artikel.php?id=233**, where the text of H. R. Hidayat Suryalaga's recommendations are given.

Figure 3. Street sign showing the Sundanese script alongside the Latin script.
The text reads ᮏ. ᮜ. ᮜᮤᮀᮌᮏᮚ, strictly transliterated *Ja. La. Liṅgajaya.*

Figure 4. Public signage from a university in Bandung.
The building name *Loka Wastuwidya* is written in Latin and Sundanese scripts.

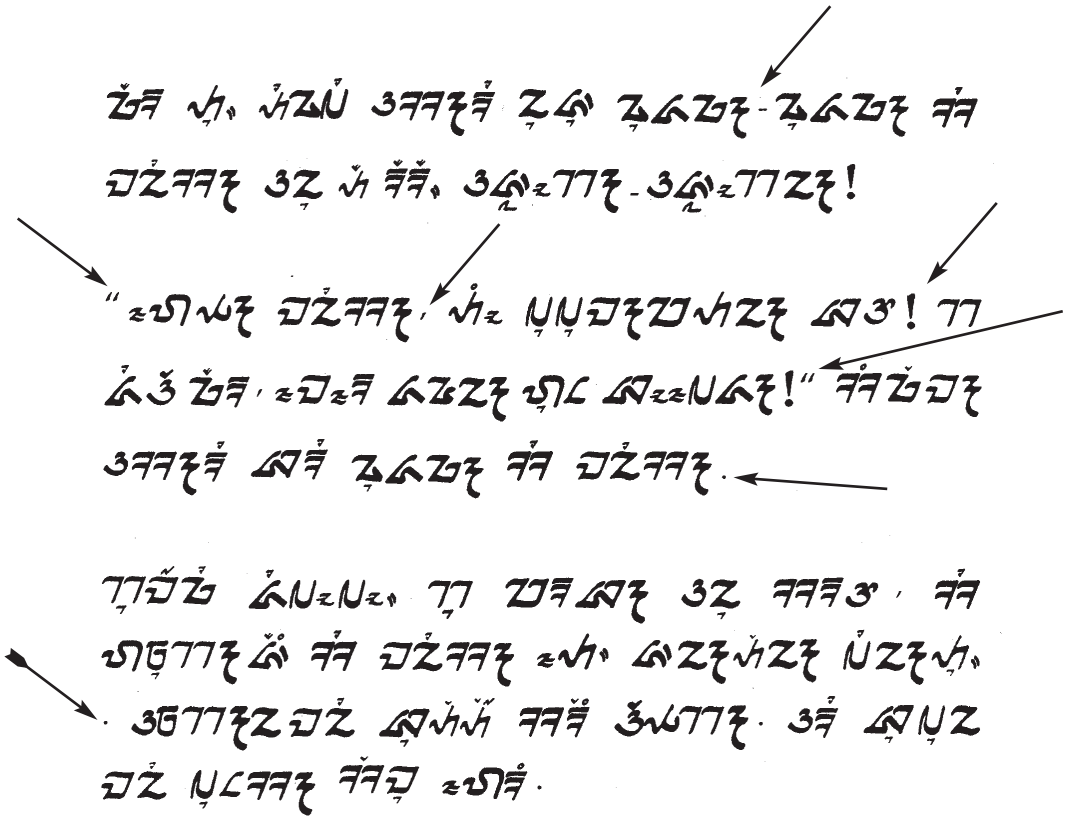


Figure 5. Text in Sundanese from Baidillah et al. 2002, showing European punctuation in use. There is a linebreaking error in the third line of the last paragraph.

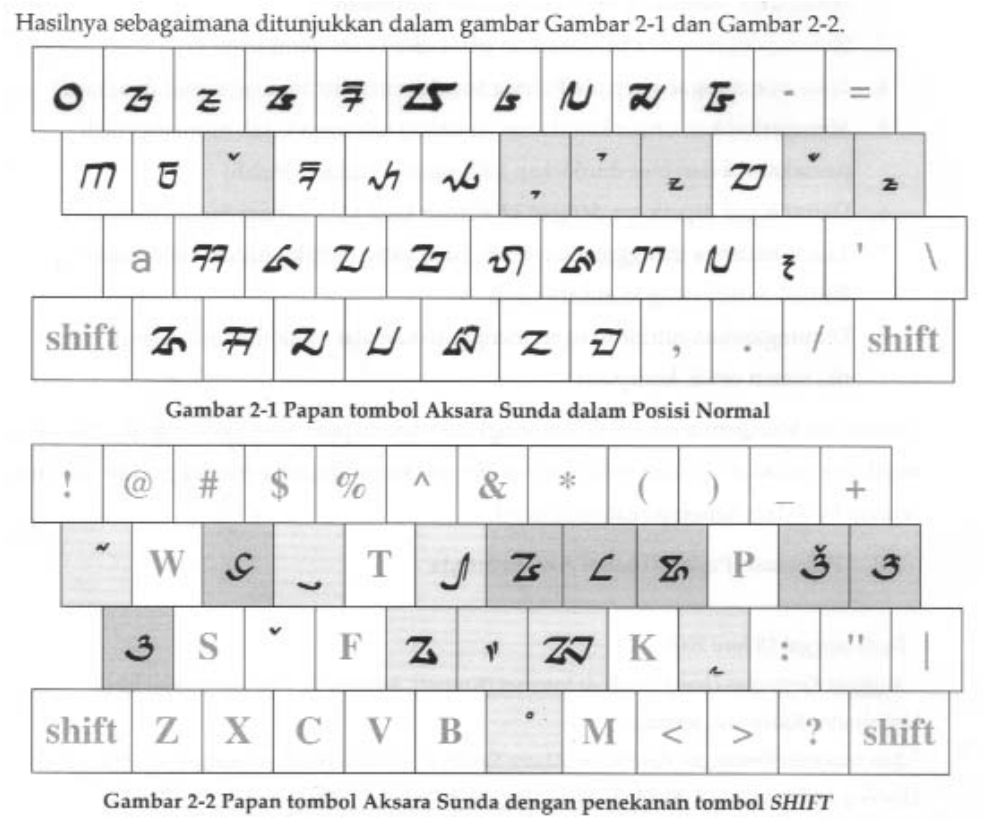
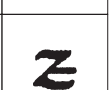
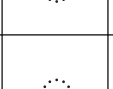
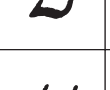


Figure 6. QWERTY-based keyboard layout for Sundanese designed by Dian Tresna Nugraha.

TABLE XX - Row 1B: SUNDANESE

	1B8	1B9	1BA	1BB
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
A				
B				
C				
D				
E				
F				

G = 00
P = 00

TABLE XX - Row 1B: SUNDANESE

hex	Name
80	SUNDANESE SIGN PANYECEK (anusvara)
81	SUNDANESE SIGN PANGLAYAR (repha)
82	SUNDANESE SIGN PANGWISAD (visarga)
83	SUNDANESE LETTER A
84	SUNDANESE LETTER I
85	SUNDANESE LETTER U
86	SUNDANESE LETTER AE
87	SUNDANESE LETTER O
88	SUNDANESE LETTER E
89	SUNDANESE LETTER EU
8A	SUNDANESE LETTER KA
8B	SUNDANESE LETTER QA
8C	SUNDANESE LETTER GA
8D	SUNDANESE LETTER NGA
8E	SUNDANESE LETTER CA
8F	SUNDANESE LETTER JA
90	SUNDANESE LETTER ZA
91	SUNDANESE LETTER NYA
92	SUNDANESE LETTER TA
93	SUNDANESE LETTER DA
94	SUNDANESE LETTER NA
95	SUNDANESE LETTER PA
96	SUNDANESE LETTER FA
97	SUNDANESE LETTER VA
98	SUNDANESE LETTER BA
99	SUNDANESE LETTER MA
9A	SUNDANESE LETTER YA
9B	SUNDANESE LETTER RA
9C	SUNDANESE LETTER LA
9D	SUNDANESE LETTER WA
9E	SUNDANESE LETTER SA
9F	SUNDANESE LETTER XA
A0	SUNDANESE LETTER HA
A1	SUNDANESE CONSONANT SIGN PAMINGKAL (subjoined ya)
A2	SUNDANESE CONSONANT SIGN PANYAKRA (subjoined ra)
A3	SUNDANESE CONSONANT SIGN PANYIKU (subjoined la)
A4	SUNDANESE VOWEL SIGN PANGHULU (i)
A5	SUNDANESE VOWEL SIGN PANYUKU (u)
A6	SUNDANESE VOWEL SIGN PANAELAENG (ae)
A7	SUNDANESE VOWEL SIGN PANOLONG (o)
A8	SUNDANESE VOWEL SIGN PAMEPET (e)
A9	SUNDANESE VOWEL SIGN PANEULEUNG (eu)
AA	SUNDANESE SIGN PAMAAEH (virama)
AB	(This position shall not be used)
AC	(This position shall not be used)
AD	(This position shall not be used)
AE	SUNDANESE LETTER KHA
AF	SUNDANESE LETTER SYA
B0	SUNDANESE DIGIT ZERO
B1	SUNDANESE DIGIT ONE
B2	SUNDANESE DIGIT TWO
B3	SUNDANESE DIGIT THREE
B4	SUNDANESE DIGIT FOUR
B5	SUNDANESE DIGIT FIVE
B6	SUNDANESE DIGIT SIX
B7	SUNDANESE DIGIT SEVEN
B8	SUNDANESE DIGIT EIGHT
B9	SUNDANESE DIGIT NINE
BA	(This position shall not be used)
BB	(This position shall not be used)
BC	(This position shall not be used)
BD	(This position shall not be used)
BE	(This position shall not be used)
BF	(This position shall not be used)

hex	Name

A. Administrative

1. Title

Proposal for encoding the Sundanese script in the BMP of the UCS.

2. Requester's name

Michael Everson

3. Requester type (Member body/Liaison/Individual contribution)

Individual contribution.

4. Submission date

2006-01-09

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

Yes.

6b. More information will be provided later

No.

B. Technical – General

1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

Yes.

Proposed name of script

Sundanese.

1b. The proposal is for addition of character(s) to an existing block

No.

1c. Name of the existing block

2. Number of characters in proposal

55

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

Category B.1.

4a. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000)

Level 2

4b. Is a rationale provided for the choice?

Yes.

4c. If YES, reference

Sundanese requires Level 2 implementation as other Brahmic scripts do.

5a. Is a repertoire including character names provided?

Yes.

5b. If YES, are the names in accordance with the character naming guidelines in Annex L of ISO/IEC 10646-1: 2000?

Yes.

5c. Are the character shapes attached in a legible form suitable for review?

Yes.

6a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Michael Everson.

6b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Michael Everson, Fontographer.

7a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes.

7b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

8. 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.

9. 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 Unicode Character Database 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 above.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

No.

2a. 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.

2b. If YES, with whom?

Dian Tresna Nugraha, script user and font designer.

2c. If YES, available relevant documents

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?

Sundanese is used on the island of Java in Indonesia.

4a. The context of use for the proposed characters (type of use; common or rare)

Used to write the Sundanese language.

4b. Reference

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

Yes.

5b. If YES, where?

In Java.

6a. 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. Positions 1B80-1BBF are proposed.

6b. If YES, is a rationale provided?

Yes.

6c. If YES, reference

Contemporary use and accordance with the Roadmap.

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

Yes.

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

No.

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

No.

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

No.

10b. If YES, is a rationale for its inclusion provided?

10c. If YES, reference

11a. 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.

11b. If YES, is a rationale for such use provided?

Yes.

11c. If YES, reference

Brahmic vowels.

12a. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

12b. If YES, reference

13a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

13b. If YES, describe in detail (include attachment if necessary)

14a. Does the proposal contain any Ideographic compatibility character(s)?

No.

14b. If YES, is the equivalent corresponding unified ideographic character(s) identified?