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

Doc Type: **Working Group Document**
Title: **Preliminary proposal for encoding the Nabataean script in the SMP of the UCS**
Source: **Michael Everson**
Status: **Individual Contribution**
Action: **For consideration by JTC1/SC2/WG2 and UTC**
Date: **2010-07-25**

1. Introduction.

The Nabataean Kingdom was at first limited to the territories surrounding the city of Petra, north of the Red Sea, but eventually annexed the Transjordan to Damascus and spread south to Medina. Aramaic was used as the official language of the Kingdom, but developed a cursive Aramaic script which became widely used, from the second century BCE to 106 CE, when the Roman province of Arabia was formed. Late Nabataean script is found in the third century CE, in graffiti drawn by passers-by on the rocks of the Sinai peninsula.

The glyphs of the Nabataean script are more ornate than other Aramaic scripts, with circles, loops, and flourishes found in some inscriptions. As the script developed, a range of ligatures and final forms was introduced, evidently as an aid to faster letter writing, ideal for business transactions.

Nabataean is generally considered to be the immediate parent of the Arabic script. An inscription found at En-Nemar, Syria, dating from the fourth century CE, is believed to be the oldest know Arabic text; its letterforms show that the Arabic script is derived from the Nabataean prototype.

2. Processing. Nabataean is written from right to left horizontally. Nabataean language inscriptions usually have no space between words; modern editors tend to insert U+0020 SPACE.

3. Glyphs. Although Nabataean inscriptions show many ligatures, in fact the standardization of such ligatures is very erratic, and so it is not possible to treat Nabataean as though it were Arabic, with cursive classes and regular joinings, final forms, etc. In fact even comprehensive treatises on the Nabataean script like Yardeni 2000 do not give such charts. Accordingly only the base characters are proposed for encoding, and no particular ligation or shaping behaviour is defined, as it cannot be mandated.

4. Character names. The names used for the characters here are based on those used for Imperial Aramaic, Other West Semitic names may have some currency, but the UCS Imperial Aramic names have been preferred here since Nabataean is an Aramaic language.

5. Numerals. Nabataean numerals are built up out of 1, 2, 3, 4, 5, 10, 20, and 100. The numbers 2 ᚢ and 3 ᚣ and 4 ᚤ are composed of multiples of 1 ǀ, but because in practice the numbers are clumped together as units separate from one another they are encoded as individual characters; divided forms ǁ, ǁǀ, ǁǁ are also found. Nabataean also has numbers for 5 ʃ, 10 ∧, 20 3, and 100 ٩. The numbers have right-to-left directionality. In the chart below, the third and sixth columns are displayed in visual order.

1	|	1 ←	11	|∧ (also ⋔)	1 + 10 ←	
2	Ա	2 ←	12	Ա∧	2 + 10 ←	
3	Ш	3 ←	13	Ш∧	3 + 10 ←	
4	Ш	4 ←	14	Ш∧	4 + 10 ←	
5	ς	5 ←	15	ς∧	5 + 10 ←	
6	|ς (also ᛃ)	1 + 5 ←	16	|ς∧	1 + 5 + 10 ←	
7	Աς	2 + 5 ←	17	Աς∧	2 + 5 + 10 ←	
8	Шς	3 + 5 ←	18	Шς∧	3 + 5 + 10 ←	
9	Шς	4 + 5 ←	19	Шς∧	4 + 5 + 10 ←	
10	∧	10 ←	100	٩|	100 + 1 ←	
20	Ʒ (21 |Ʒ or ʐ)	20 ←	200	٩Ա	100 + 2 ←	
30	∧Ʒ	10 + 20 ←	300	٩Ш	100 + 3 ←	
40	ƷƷ	20 + 20 ←	400	٩Ш	100 + 4 ←	
50	∧ƷƷ	10 + 20 + 20 ←	500	٩ς	100 + 5 ←	
60	ƷƷƷ	20 + 20 + 20 ←	600	٩|ς	100 + 1 + 5 ←	
70	∧ƷƷƷ	10 + 20 + 20 + 20 ←	700	٩Աς	100 + 2 + 5 ←	
80	ƷƷƷƷ	20 + 20 + 20 + 20 ←	800	٩Шς	100 + 3 + 5 ←	
90	∧ƷƷƷƷ	10 + 20 + 20 + 20 + 20 ←	900	٩Шς	100 + 4 + 5 ←	

6. Punctuation. Script-specific punctuation for Nabataean is not known.

7. Unicode Character Properties

```
10860;NABATAEAN LETTER ALEPH;Lo;0;R;;;;;N;;;;;
10861;NABATAEAN LETTER BETH;Lo;0;R;;;;;N;;;;;
10862;NABATAEAN LETTER GIMEL;Lo;0;R;;;;;N;;;;;
10863;NABATAEAN LETTER DALETH;Lo;0;R;;;;;N;;;;;
10864;NABATAEAN LETTER HE;Lo;0;R;;;;;N;;;;;
10865;NABATAEAN LETTER WAW;Lo;0;R;;;;;N;;;;;
10866;NABATAEAN LETTER ZAYIN;Lo;0;R;;;;;N;;;;;
10867;NABATAEAN LETTER HETH;Lo;0;R;;;;;N;;;;;
10868;NABATAEAN LETTER TETH;Lo;0;R;;;;;N;;;;;
10869;NABATAEAN LETTER YODH;Lo;0;R;;;;;N;;;;;
1086A;NABATAEAN LETTER KAPH;Lo;0;R;;;;;N;;;;;
1086B;NABATAEAN LETTER LAMEDH;Lo;0;R;;;;;N;;;;;
1086C;NABATAEAN LETTER MEM;Lo;0;R;;;;;N;;;;;
1086D;NABATAEAN LETTER NUN;Lo;0;R;;;;;N;;;;;
1086E;NABATAEAN LETTER SAMEKH;Lo;0;R;;;;;N;;;;;
1086F;NABATAEAN LETTER AYIN;Lo;0;R;;;;;N;;;;;
10870;NABATAEAN LETTER PE;Lo;0;R;;;;;N;;;;;
10871;NABATAEAN LETTER SADHE;Lo;0;R;;;;;N;;;;;
10872;NABATAEAN LETTER QOPH;Lo;0;R;;;;;N;;;;;
10873;NABATAEAN LETTER RESH;Lo;0;R;;;;;N;;;;;
10874;NABATAEAN LETTER SHIN;Lo;0;R;;;;;N;;;;;
10875;NABATAEAN LETTER TAW;Lo;0;R;;;;;N;;;;;
10878;NABATAEAN NUMBER ONE;No;0;R;;;;1;N;;;;;
10879;NABATAEAN NUMBER TWO;No;0;R;;;;2;N;;;;;
1087A;NABATAEAN NUMBER THREE;No;0;R;;;;3;N;;;;;
1087B;NABATAEAN NUMBER FOUR;No;0;R;;;;4;N;;;;;
1087C;NABATAEAN NUMBER FIVE;No;0;R;;;;5;N;;;;;
1087D;NABATAEAN NUMBER TEN;No;0;R;;;;10;N;;;;;
1087E;NABATAEAN NUMBER TWENTY;No;0;R;;;;20;N;;;;;
1087F;NABATAEAN NUMBER ONE HUNDRED;No;0;R;;;;100;N;;;;;
```

8. Bibliography

Ballhorn, Friedrich. 1864. *Alphabete orientalischer und occidentalischer Sprachen*. Leipzig: Brockhaus.

Christian, Arthur. 1905. *Débuts de l'imprimerie en France: L'Imprimerie Nationale; L'Hôtel de Rohan*. Paris: Imprimerie Nationale.

Faulmann, Carl. 1990 (1880). *Das Buch der Schrift*. Frankfurt am Main: Eichborn. ISBN 3-8218-1720-8

Fossey, Charles. 1948. *Notices sur les caractères étrangers anciens et modernes rédigées par un groupe e savants*. Nouvelle édition míse à jour à l'occasion du 21e Congrès des Orientalistes. Paris: Imprimerie Nationale de France.

Imprimerie Nationale. 1990. *Les caractères de l'Imprimerie Nationale*. Paris: Imprimerie Nationale Éditions. ISBN 2-11-081087-8

Yardeni, Ada. 2000. *Textbook of Aramaic, Hebrew, and Nabataean documentary texts from the Judaean desert and related material*. Jerusalem: The Ben-Zion Dinur Center for Research in Jewish History, The Hebrew University. ISBN 965-350-083-X

	1088	1089
0	٥ 10880	ﬞ 10890
1	ﬞ 10881	ﬞ 10891
2	ﬞ 10882	ﬞ 10892
3	ﬞ 10883	ﬞ 10893
4	ﬞ 10884	ﬞ 10894
5	ﬞ 10885	ﬞ 10895
6	ﬞ 10886	
7	H 10887	
8	ﬞ 10888	ﬞ 10898
9	ﬞ 10889	ﬞ 10899
A	ﬞ 1088A	ﬞ 1089A
B	ﬞ 1088B	ﬞ 1089B
C	ﬞ 1088C	ﬞ 1089C
D	ﬞ 1088D	ﬞ 1089D
E	ﬞ 1088E	ﬞ 1089E
F	ﬞ 1088F	ﬞ 1089F

Letters

10880	٥	NABATAEAN LETTER ALEPH
10881		NABATAEAN LETTER BETH
10882		NABATAEAN LETTER GIMEL
10883		NABATAEAN LETTER DALETH
10884		NABATAEAN LETTER HE
10885		NABATAEAN LETTER WAW
10886		NABATAEAN LETTER ZAYIN
10887	H	NABATAEAN LETTER HETH
10888		NABATAEAN LETTER TETH
10889		NABATAEAN LETTER YODH
1088A		NABATAEAN LETTER KAPH
1088B		NABATAEAN LETTER LAMEDH
1088C		NABATAEAN LETTER MEM
1088D		NABATAEAN LETTER NUN
1088E		NABATAEAN LETTER SAMEKH
1088F		NABATAEAN LETTER AYIN
10890		NABATAEAN LETTER PE
10891		NABATAEAN LETTER SADHE
10892		NABATAEAN LETTER QOPH
10893		NABATAEAN LETTER RESH
10894		NABATAEAN LETTER SHIN
10895		NABATAEAN LETTER TAW

Numbers

10898		NABATAEAN NUMBER ONE
10899		NABATAEAN NUMBER TWO
1089A		NABATAEAN NUMBER THREE
1089B		NABATAEAN NUMBER FOUR
1089C		NABATAEAN NUMBER FIVE
1089D		NABATAEAN NUMBER TEN
1089E		NABATAEAN NUMBER TWENTY
1089F		NABATAEAN NUMBER ONE HUNDRED

10. Figures.

Hauranitisch		Nabathäisch	Palmyrenisch	Wert
nach Halévy	H. Müller			
				a
				b
				g
				d
				h
				w
				z
				χ
				ṭ
				y
				k
				l
				m
				n
				s
				a
				p
				ṣ
				q
				r
				š
				t

In der Wüste Hauran, welche sich zwischen Palästina und Arabien erstreckt, befinden sich Ruinen mit merkwürdigen Inschriften, welche ein Mittelglied zwischen himyarischen und aramäischen Formen bilden. An sie schliessen sich die palmyrenische und die Schrift der Nabathäer, deren Hauptstadt Petra war; die hier gegebenen Zeichen sind theils den Münzen aus dem 2. Jahrhundert v. Chr., theils den sinaitischen Inschriften entnommen, deren Entzifferung besonders den Studien Levy's zu verdanken ist.

Figure 1. Chart of "scripts in the east of Palestine" from Faulmann 1880, showing Nabataean.

(184) *Das Verhältnis der mandäischen und nabatäischen Schrift zum aramäischen*
Konsonantenalphabet

Lautwert	Aram.	Nabatäisch	Sinaitisch	Mandäisch	Lautwert	Aram.	Nabatäisch	Sinaitisch	Mandäisch
'					l				
b					m				
g, ǧ (arab.)					n				
d					s				
h					ʿ				
w					p; f (arab.)				
z					ṣ				
ḥ					q				
ṭ					r				
j					š				
k					t				

Figure 2. Chart of Semitic scripts from Haarmann 1990, showing Nabataean.

6

TABLE 5.5: *Scripts Derived from Aramaic Script (Garbini 1979, fig. 7)*[a]

	XVII	XVIII	XIX	XX
ʾ				
b				
g				
d				
h				
w				
z				
ḥ				
ṭ				
y				
k				
l				
m				
n				
s				
ʿ				
p				
ṣ				
q				
r				
š				
t				

a. Col. XVII, Hebrew square script; col. XVIII, Palmyrene script; col. XIX, Nabatean script; col. XX, Ancient Arabic script.

Figure 3. Chart of Semitic scripts from Haarmann 1990, showing Nabataean.

ALPHABETS ARAMÉENS COMPARÉS.

(ARAMÉEN ANCIEN, PALMYRÉNIEN, NABATÉEN.)

VALEUR.	ARAMÉEN ANCIEN.			PALMYRÉNIEN.	NABATÉEN.
	VIIIᵉ SIÈCLE.	VIᵉ SIÈCLE.	IVᵉ SIÈCLE.		
’					
b					
g					
d					
h					
w					
z					
ḥ					
ṭ					
y					
k					
l					
m					
n					
s					
‘					
f					
ṣ					
ḳ					
r					
š					
t					

Figure 4. Chart of "Aramaic alphabets" from Ballhorn 1864, showing Nabataean.

TEXTE NABATÉEN.

———

ÉPITAPHE DE L'AN 26 DE NOTRE ÈRE,
À MEDAÏN-SALEH (HEDJRA).

ꙏꙮ	ꙧꙛ꙯	꙱	ꙮꙛꙮꙣ	꙯ꙧꙣ	꙱꙰	ꙧ꙯	ꙮꙣꙛ꙱	꙱ꙛ
h š p nl	*nl a w*	*r b*	*a y s a*	*nl h k*	*db ʿ*	*yd*	*a r b ḳ*	*hnd*
pour lui-même	,de Waalan	fils	,le médecin	Kahlan	a fait	que	le tombeau	Ceci est

ꙮꙣꙛ꙱ꙧ	꙱꙯	ꙣ꙯	꙯ꙛꙣꙛ	꙯ꙛꙣꙛ	꙱ꙮꙣꙛ꙯	꙱ꙣꙛꙧ꙯
y t y a w	*ml ʿ*	*d ʿ*	*ḳ d ṣ ab*	*ḳ d ṣ a*	*h r ḥ a w*	*h d l y w*
et est	;toujours	jusque	en héritier	d'héritier	et sa postérité	et pour ses enfants

ꙮꙧꙣꙛ꙯	ꙮꙛꙣꙧ	ꙣ꙯	ꙮꙣꙛꙣ	ꙣꙮꙧꙣ꙯	ꙮꙛꙣ	ꙣꙣ꙯	꙱ꙛꙣ꙯
a r š w dl	*mr ḥ m*	*yd*	*a m r ḥ*	*t ḳ yl ḥ k*	*mr ḥ*	*hnd*	*a r b ḳ*
à Dusarès	consacré	qui est	du sanctuaire	à la façon	un sanctuaire	ce tombeau	

ꙛꙣ	ꙣ꙯	ꙣꙧ꙯꙯	꙯ꙛꙣꙛ	ꙛꙣꙛꙮ	ꙛꙣ	ꙛ꙯	꙯ꙧꙣ꙱ꙮ꙯	꙯ꙛꙣꙛꙛ
a l	*yd*	*t r yw*	*ḳ d ṣ a*	*š wn a*	*l k*	*l ʿ*	*wml š w*	*w ṭ bnb*
ne	que	et héritier	parent	homme	tout	Pour	.et chez les Salamiens	en Nabatène

꙯ꙛꙣ	ꙣꙛꙮꙛ꙯	ꙛꙣꙮꙣꙣ꙯	꙯ꙛꙣꙛ꙯ꙣ	ꙛꙣꙛ꙯	ꙧꙛꙣ	ꙣꙮꙣꙛ	ꙣꙮ꙯	꙱ꙛꙣꙧ	ꙛꙣ꙯ꙮ
kl m	*t t r ḥ l*	*š m ḥ w*	*ny t l t*	*tn š*	*ry a*	*ḥ ryb*	*hnd*	*a r b ḳ*	*nbzy*
roi	,d'Harétat	et cinq	trente	de l'an	Iyar	Au mois	.ce	tombeau	il vende

ꙛꙣ꙯꙯
w ṭ bn
.de Nabatène

R. DUSSAUD, 1927.
(J. FÉVRIER, 1948.)

Figure 5. Sample text in Nabataean from Fossey 1948.

(1) ... (2) ...

(3) ...

(4) ...

(5) ...

Inscription de Pétra

corps 14

(1) ... (2) ...

(3) ...

(4) ...

(5) ...

Inscription de Pétra (lecture rectifiée dans la copie du R. P. Vincent)
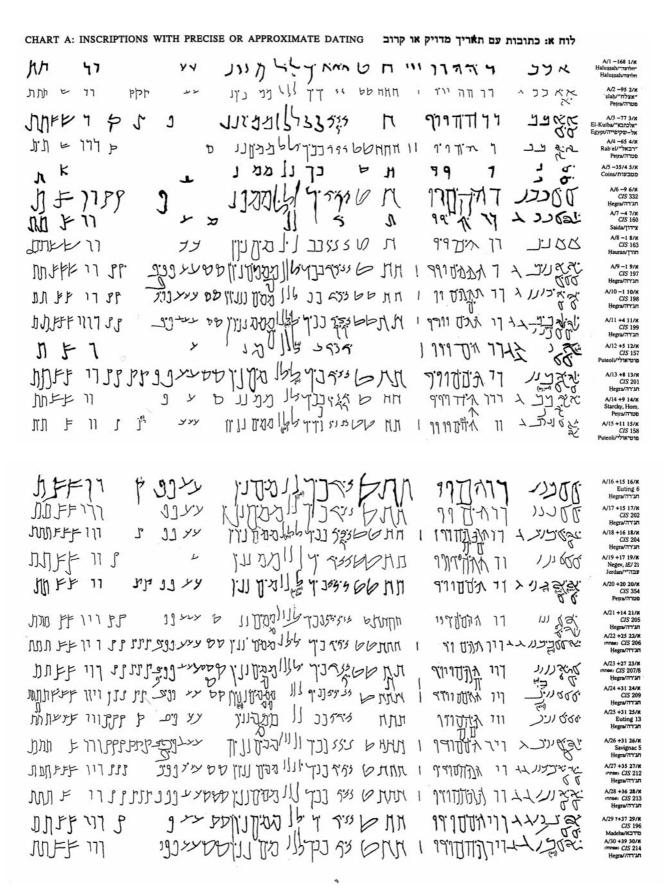
Figure 6. Sample of Nabataean text from Christian 1905.

Figure 7a. Table of dated Nabataean inscriptions from Yardeni 2000.
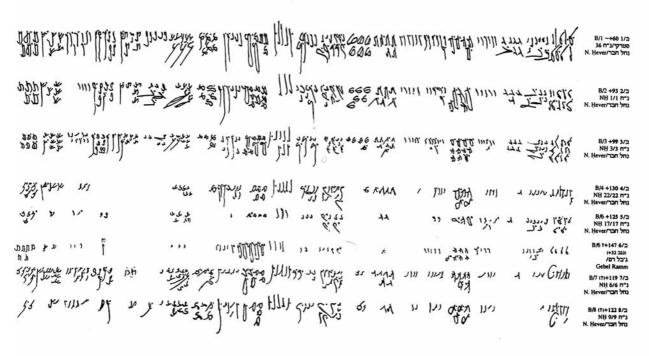
Figure 7b. Table of dated Nabataean inscriptions from Yardeni 2000.

Figure 8. Table of Nabataean documents in cursive script from Yardeni 2000.

CHART C: INSCRIPTIONS WITH CONJECTURAL DATING　　לוח ג: כתובות עם תאריך משוער

					C/1 ?—100 1/1 Incant./… Kh. Raqiq/קיק
					C/2 –1—25 2/1 CIS 162 Hauran/חורן
					C/3 +25—+1 3/1 CIS 164 Hauran/חורן
					C/4 +40—9 4/1 Negev, IEJ 11 'Avdat/עבדת
					C/5 +40—9 5/1 Negev, IEJ 11 'Avdat/עבדת
					C/6 +70—40 6/1 CIS 219 Hegra/חגרה
					C/7 +50—25 7/1 CIS 216 Hegra/חגרה
					C/8 +50—25 8/1 CIS 228 Hegra/חגרה
					C/9 +139—16 9/1 CIS 217 Hegra/חגרה
					C/10 +8 10/1 CIS 200 Hegra/חגרה
					C/11 +70—40 11/1 St.&St.RB 73 Petra/פטרה
					C/12 +64 12/1 Savignac 38 Hegra/חגרה
					C/13 +75—50 13/1 Star.& Strag.RB 73 Petra/פטרה
					C/14 ?75—50 14/1 CIS 350 Petra/פטרה
					C/15 +100—75 15/1 Savignac 2 Hegra/חגרה
					C/16 +150—100 16/1 Rotenberg/רוטנברג W. Firān/ואדי פיראן
					C/17 +125—106 17/1 CIS 173 Hauran/חורן
					C/18 ?+150 18/1 CIS 340 G.Mishma/גבל משמה
					Niṣṣanah/ניצנה ?+300—100

Figure 9. Table of Nabataean inscriptions with conjectural dating from Yardeni 2000.

Figure 10a. Table of Nabataean graffiti from Yardeni 2000.

Figure 10b. Table of Nabataean graffiti from Yardeni 2000.

Figure 11. Table of Nabataean numbers. Source unidentified; from an online forum.

A. Administrative

1. Title
Preliminary proposal for encoding the Nabataean script in the SMP of the UCS
2. Requester's name
Michael Everson
3. Requester type (Member body/Liaison/Individual contribution)
Individual contribution.
4. Submission date
2010-07-25
5. Requester's reference (if applicable)
6. Choose one of the following:
6a. This is a complete proposal
No.
6b. More information will be provided later
Yes.

B. Technical – General

1. Choose one of the following:
1a. This proposal is for a new script (set of characters)
Yes.
1b. Proposed name of script
Nabataean.
1c. The proposal is for addition of character(s) to an existing block
No.
1d. Name of the existing block
2. Number of characters in proposal
30.
3. Proposed category (A-Contemporary; B.1-Specialized (small collection); B.2-Specialized (large collection); C-Major extinct; D-Attested extinct; E-Minor extinct; F-Archaic Hieroglyphic or Ideographic; G-Obscure or questionable usage symbols)
Category E.
4a. Is a repertoire including character names provided?
Yes.
4b. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?
Yes.
4c. Are the character shapes attached in a legible form suitable for review?
Yes.
5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?
Michael Everson.
5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:
Michael Everson, FontLab.
6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?
Yes.
6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?
Yes.
7. 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.
8. 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?
Laïla Nehmé.
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?
See above.

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

To write the Aramaic language.

4b. Reference

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

Yes.

5b. If YES, where?

In scholarly publications.

6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?

No.

6b. If YES, is a rationale provided?

6c. If YES, reference

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)?

No.

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

11c. If YES, reference

11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

11e. If YES, reference

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

No.

12b. If YES, describe in detail (include attachment if necessary)

13a. Does the proposal contain any Ideographic compatibility character(s)?

No.

13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?