

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 the addition of ten Latin characters to the UCS
Source: Michael Everson, Denis Jacquerye, Chris Lilley
Status: Individual Contribution
Date: 2012-07-26

This proposal requests the encoding of ten Latin characters including three casing pairs, and four capital letters which provide casing support for existing characters. If this proposal is accepted, the following characters will exist:

Ɑ	A7AE	LATIN CAPITAL LETTER TURNED K x 029E Ɱ latin small letter turned k • used in Americanist orthographies
Ɱ	A7AF	LATIN CAPITAL LETTER TURNED T x 0287 Ɐ latin small letter turned t • used in Unifon and Americanist orthographies
Ɔ	A7B0	LATIN CAPITAL LETTER BETA
β	A7B1	LATIN SMALL LETTER BETA • used in Gabonese orthographies
Ɔ	A7B2	LATIN CAPITAL LETTER OMEGA
ω	A7B3	LATIN SMALL LETTER OMEGA • used in Gabonese orthographies
Ɔ	A7B4	LATIN CAPITAL LETTER TAU GALLICUM
ɖ	A7B5	LATIN SMALL LETTER TAU GALLICUM • Gaulish
Ɑ	A7BE	LATIN CAPITAL LETTER CHI x 026A Ɱ latin small letter stretched x [latin small letter chi]
Ɱ	A7BF	LATIN CAPITAL LETTER SMALL CAPITAL I x 026A Ɱ latin letter small capital i • used in Unifon and Gabonese orthographies

1. LATIN LETTER TURNED K and **LATIN LETTER TURNED T** were used in transcriptions of the Dakota language in publications of the American Board of Ethnology in the late 19th century. Most of the base letters used here have capital forms encoded in the UCS: Cc, Ɔo (used as a consonant), Çç, Ćć, Ee, Чч, Jj, Qq—but not ꝥ or ꝧ (used for [k̥]~[k̚] and [t̥]~[t̚] respectively): Ꝩ and ꝩ are missing.

ă	as a in <i>what</i> or as o in <i>not</i> .
c	sh, given as ś by the author and Matthews.
o	a medial sound, between sh (ś) and zh (ż).
ç	as th in <i>thin</i> , the surd of φ.
dϕ	a d sound followed by a dh sound which is scarcely audible. This combination is peculiar to the Biloxi, Hidatsa, and Kwapa languages. Given as d by Matthews.
ϕ	dh, or as th in <i>the</i> , the sonant of ç.
e	a short e as in <i>get</i> .
q	a sound heard at the end of certain syllables, but slightly audible, nearer h than kh. Given by Matthews as an apostrophe after the modified vowel.
i	as in <i>it</i> .
j	zh, or as z in <i>azure</i> . Given as ż by the author and as z by Matthews.
ꝥ	a medial k, between g and k, heard in Teton, Çegiha, etc.
Ꝧ	an exploded k. Given as Ꝧ by the author.
n	a vanishing n, scarcely audible, as the French n in <i>bon</i> , <i>vin</i> , etc., occurring after certain vowels. Given as ŋ by the author.
ñ	as ng in <i>sing</i> , <i>singer</i> , but not as ng in <i>finger</i> ; heard sometimes before a k-mute, at others just before a vowel, as in Ꝭiwere (i-çũñ-e, i-yũñ-e, wañ-e, etc.). Given as ŋ by the author.
q	kh or as ch in German <i>ach</i> . Given as h by the author and Matthews.
ꝧ	a medial sound, between d and t.
ũ	as oo in <i>foot</i> .
û	as u in <i>but</i> , given by Matthews as “a” with a dot subscript.
tc	as ch in <i>church</i> . Given as ć by the author.
tç	a t sound followed by a ç (th) sound, as th in <i>thin</i> , but scarcely audible. It is the surd of dϕ, and is peculiar to the Biloxi, Hidatsa, and Kwapa languages. Given as t by Matthews.
ɽo	a medial sound, between dj (j as in judge) and tc.
ɽs	a medial sound, between dz and ts.

Figure 1. In this text, **LATIN LETTER TURNED T** is shown in both capital and lower-case forms, and **LATIN LETTER TURNED K** is shown in lower-case. Typographically the Bureau simply turned the T sort upside down, which accounts for its hanging below the descender. It is unlikely that this should be imitated today, and the name of the Ꝭiwere dialect of Lakota (now spelt Chiwere) should be written with a Ꝭ on the baseline.

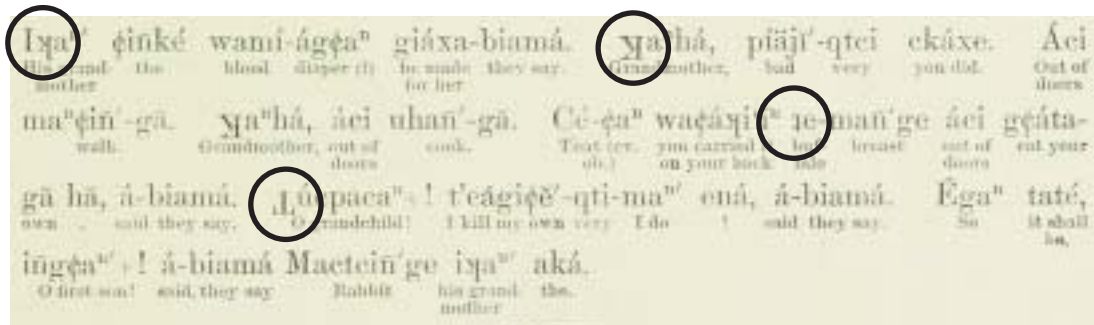


Figure 2. Here, the LATIN CAPITAL LETTER TURNED K is shown, in the word *Ya'ha'* 'Grandmother', and LATIN CAPITAL LETTER TURNED T is also shown, in the word *Lu'paca'* 'Grandchild'. Again, typographically the Bureau simply turned the K sort upside down, which accounts for its hanging below the descender.

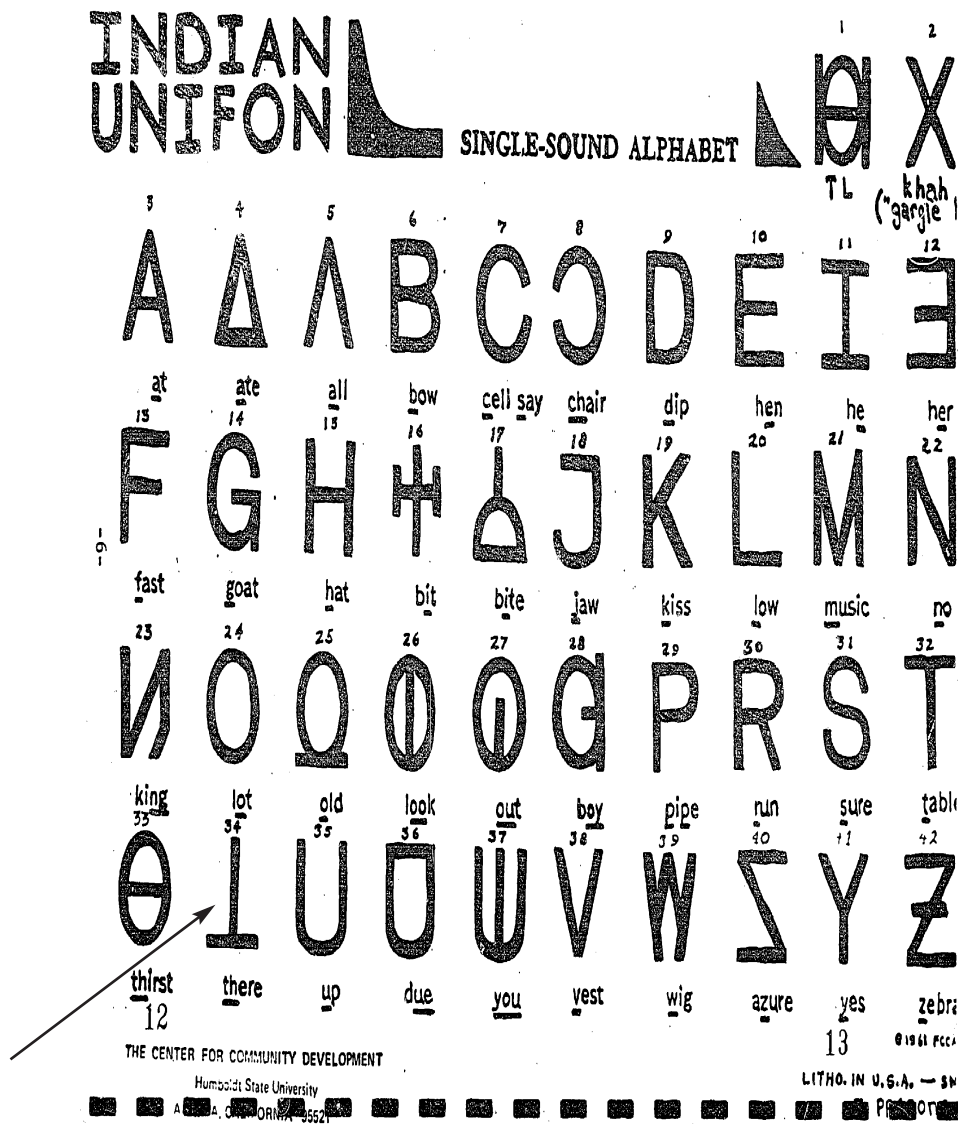


Figure 3. In the Unifon phonetic alphabet, the TURNED CAPITAL T is used for the sound [ð] in earlier orthographies for English, and in orthographies for various Native American languages.

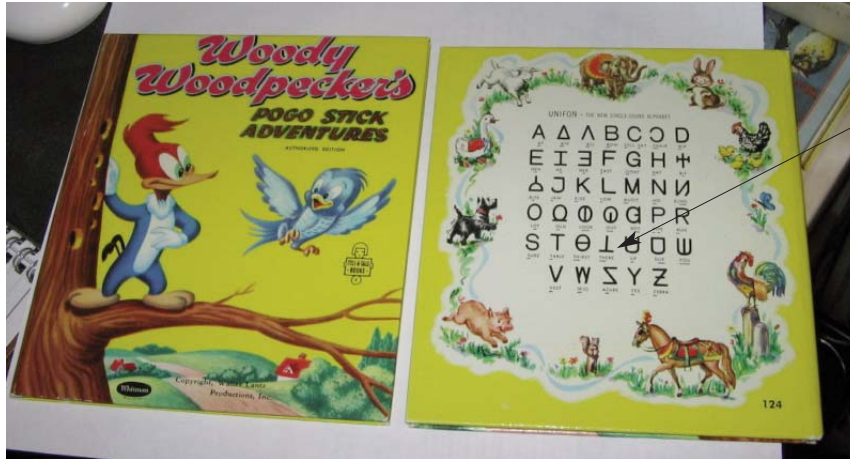


Figure 4. Example of a children's book published in Unifon in 1954. The alphabet given is:
 AΔΛBCODEIEIÆFGH†ΔJKLNMNŃOOQOΘIPRSTΘLUSWVWSYZ

2. LATIN LETTER CAPITAL CHI was devised by Carl Richard Lepsius and used in *Das allgemeine linguistische Alphabet* (1855) and in the second edition *Standard Alphabet for Reducing Unwritten Languages and Foreign Graphic Systems to a Uniform Orthography in European Letters* (1863)

wird das *m* von *χnum*,
 in *Χνοῦβι θεῶ* geweiht; *ix*

Figure 5. Here is an example of Lepsius distinguishing GREEK CAPITAL CHI and LATIN CAPITAL CHI in roman and italic type. The shape of the italic LATIN CAPITAL CHI is not used in Greek.

32	χ ^{ar}	Reten, and Šasu by Daniel Hy Haigh.	[Januar u. Februar
Θμῶσις	9. 8.	1449. Jan. 1444. Jan.	Thothmes IV. „ and Amunhotep III.
Ἀμένωφις	30. 10.	1440. Sept.	Amunhotep III.
Ἦρος	36. 5.	1409. July.	Amunmes Amuntutanχ Ai (?) Horemheb.
Ἀκεγχρήσις	12. 1.	1373. Dec.	(no monuments).

I believe that these reigns are included in the 36. 5. assigned to Horemheb.

Figure 6. The journals *Zeitschrift für ägyptische Sprache und Altertumskunde* and *Abhandlungen der Königlichen Akademie der Wissenschaften* used glyphs made for Lepsius' Standard alphabet, in both upper-case and lower-case. Note how his GREEK CHI and LATIN CHI differ.

welcher tödtet, man pfllegt ihn zu tö
 ?). *χ^{ar}-t^o-i p^e-t^obe p^e-m^asi-^aχ^en*
 (τωθε) zu bezeichnen die Nachforschun

Figure 7. Example of Lepsius' LATIN CAPITAL LETTER CHI and LATIN SMALL LETTER CHI in transcription of ancient Egyptian.

tundirhum; lā yu-minūna. 6. *Ḥatīma* ʾllāhu ʾsalā qulābihim, wa ʾsalā samʾihim, wa ʾsalā absārūhim yisāwatun, wa lahum ʾadābun ʾadīmūn. 7. Wa minā ʾl-nāsi man yaqūlu: āmannā bi ʾllāhi wa bi ʾl yaumi ʾl āḫiri, wa mā lum bimūminīna, 8. *yuḫādīsūna* ʾllāha wa ʾladīna āmanū, wa mā yaḫdaʾūna illā *anfusahum*, wa mā yašʾurūna.

Figure 8. Example from Lepsius 1863:245 showing LATIN CAPITAL LETTER CHI and LATIN SMALL LETTER CHI in his Standard Orthography transcription of Arabic.

Juse ama eme-i saiin ʼeʼe babe kemuni soikolome yabumbi.
Saxaku bitʼxe-be xulači saiin guči be baɣa gese, xulaxa bitʼxe-
*be sabučī fe guču-be ačaxa gese. *Ḥančikiŋge*-be urgujebu *Kūnte*-i*
ʼxenduʼxenge. Sinagan-de dakitara aŋgala gosiʼolo. Manju
*bitʼxe-be urunaku urebu akuči Nikan bitʼxe-be *χaj* kiyame ge-*
dukeleme mutembio.

Figure 9. Example from Lepsius 1863:245 showing LATIN CAPITAL LETTER CHI and LATIN SMALL LETTER CHI in his Standard Orthography transcription of Manchu.

3. LATIN LETTER SMALL CAPITAL I, LATIN LETTER BETA, and LATIN LETTER OMEGA. As has often happened with IPA characters used in practical orthographies, a capital form has been innovated on the basis of the lower-case.

LATIN LETTER SMALL CAPITAL I is also used with upper-case and lower-case forms in several Kulango language publications from Bondoukou. This example (from Psalm 118) comes from *Ἦ de bī dali bitese*, p 10:

Alleluya!
 → Ἦ kpre Yego bɔ gyasole, gyɪgaleɪ hō kyere,
 gyɪgaleɪ bɔ korigyo tuben haa ti-ɪ.
 Izrael bɔ yɔgɔɪywoɔ, ɪ dw-ke:
 bɔ korigyo tuben haa ti-ɪ.

Figure 10. Note here how the first words “Ἦ kpre” has a CAPITAL SMALL CAPITAL I, contrasting with the ordinary CAPITAL I in “Izrael”. At the end of the third line, the words “ti-ɪ” show contrasting use of the two lower-case letters. In all caps, these would be written “Ἦ KPRE, IZRAEL, TI-ɪ”.

LATIN LETTER BETA was arguably disunified from Greek by the devisers of the IPA. David Abercrombie describes this in his 1967 *Elements of General Phonetics*:

A good source from which letters can be borrowed is the Greek alphabet, and β γ ε θ φ χ, for example, have been made use of for centuries in roman-based phonetic notations. Borrowed Greek letters are sometimes redesigned so as to fit in with the general appearance of roman letters. The preceding six characters, for example, have for this reason been modified as follows: β γ ε θ φ χ. The Cyrillic alphabet can also offer possible new characters, such as Ъ ѣ, and the script form *u*.

Figure 11. Here we see the already-disunified γ LATIN LETTER GAMMA, ε LATIN LETTER OPEN E, and φ LATIN LETTER PHI, which are quite distinct from the usual Greek γ, ε, and φ (shown here also in Times). The normal shape of β GREEK LETTER BETA has no serif on its lower descender, and as can be seen here, the shape of the Latinized IPA β BETA is based more on the β SHARP S than the original Greek letter.

806. The sound **w** causes difficulty to many foreigners, especially to Germans. They generally replace it by a different kind of bi-labial fricative, namely one in which the lips are kept flat instead of being rounded and pushed forward, and in which the tongue is in a neutral position instead of being raised at the back. The phonetic symbol for this consonant is **β**. Its lip-position is the same as that of **ϕ** (Fig. 89). It is a sound intermediate in acoustic effect between **w** and **v**; it is very frequently heard in German words like *Quelle* 'kβelə or 'kvelə, *zwei* tsβai or tsvai. Sometimes foreigners replace **w** by **v**.

Figure 12. In Daniel Jones' 1932 *An outline of English phonetics*, an even less "Greek-like" beta can be seen (Jones was Assistant Secretary of the International Phonetic Association from 1907 to 1927, Secretary from 1927 to 1949, and President from 1950 to 1967):

t	Ṭ	tētētē	"tuméfié"	omyene-mpongwe
u	Ṭ	tsúlù	"source "	lekaniji
ü	Ṭ	ékùri	"piège"	fan-ntumu
v	V	úvèyà	"donner"	yisangu
β	β	íβéβè	"voler "	beŋga
w	W	wàyéndà	"étrangers "	yefia
w	W	njwé	"chef"	fan-ntumu

Figure 13. The LATIN BETA in lower-case and upper-case form has been found in *Revue Gabonaises des Sciences de l'Homme*, No. 2, 1990, p.113. The usage is based on the "Alphabet scientifique des langues du Gabon" (ASG) first published 1989, which was followed by the "Orthographe des langues du Gabon" intended for the educational system in 1999. The unique Latin capital form **β** is unknown for Greek B:

Y E B O B E

Figure 14. From the same journal, page 193, The word *yeβoβe* in all caps (**YEB OBE**):

18. Dìbá:là yâ :— "Mî ngà nà dùnyùrù dùnámòlè Yèbòtùyé tì nìyèβù:ndà".	18. L'homme dit : "Moi, je suis fatigué. Va-t-en que je me repose."
19. Dìtèngú tî : — "Mî nzìmá:ngà βù:ndà ywàndì. Mî sâ:yùbwí:là dùnyùrù." Dìbá:là tsìβù:ndè.	19. Le revenant dit : "Non, je ne veux pas. Repose-toi donc. Je ne te tiens pas le corps." L'homme se reposa.

Figure 15. Text in Lumbu, a Bantu language spoken around the city of Mayumba in Gabon. Here the upper-case glyph has a rounded top like the lower-case—yet again a shape unknown for Greek B:

LATIN LETTER OMEGA is used in several Kulango language publications from Bondoukou or its area, including an alphabet primer ("syllabaire"), four liturgy books (a missal), a funeral book, and two catechism books. All were published in the 1990s, most probably by the diocese of Bondoukou.

(Y) BI tɛsɛ Jezɛ, ywɔ bɛrɛ pɪɪ pɛɛ, waa kprɛ-bɛ lɛ bɔ nyĩ
 ũguogo vɛɛ. Dɛ lɛ tɪɪ ω swmɛ ... hŵ daa gbigo hɔɔ.
 ω kyɛrɛ: mɔm mĩɪ lɛ mŵ, kpŵkprɛ'rɛ hŵ hɛ saakɔ hɔɔ-rɪ, zɛɛ-ɛ
 lɛ kyɛ-lɛ, lɛ hŵ nyĩ ãyako ω kyĩŋɔ ariginā-nɪ. Prɛŵ ω yi lɛ
 bɔŋwɪ drunyā, lɛ ω kprɛ-gɛ. Bɪɪ dalɪ-ω lɛ ω kpele bɪ ywɔkɔ-rɪ
 ω sɪrɔ-rɔ

Figure 16. This example comes from *Les funérailles chrétiennes*, p. 17. It can be seen that Latin Ū ω are not the same as Greek Ω ω. The Times glyph for ω was based on the IPA character ω LATIN SMALL LETTER CLOSED OMEGA.

The unique Latin capital forms of LATIN CHI, LATIN BETA, and LATIN OMEGA meet one of the major criteria for disunification. Whether the existence of these characters would affect the recommendations of the International Phonetic Association is a matter for the eventual decision of the Association. Certainly much IPA text currently uses the Greek β, for instance. Much probably uses the Latin sharp β, and certainly much IPA text also currently uses Latin B for the same character in pre-UCS encoded fonts, which are still unfortunately more widespread than one might wish. Support for a disunification in principle has been given by John Wells, currently a member of the Council of the Association, though he agreed that the preferred mapping would be a matter for the Association to consider and decide in future. In the meantime, the letters proposed here are required for natural language text in Gabon.

4. LATIN LETTER TAU GALLICUM is a letter which we devised as an extension to the Latin alphabet by native writers of the Gaulish language who lived in Roman Gaul. (Other Gaulish inscriptions use the North Italic or the Greek Alphabets) The TAU GALLICUM represents a dental affricate; most scholars believe it was [ts], though some have wondered if it might be [ð] or [θ], and is itself of uncertain origin, as it has features both of Latin D and of Greek Θ. In environments where the tau gallicum is not available, the Icelandic eth (Ðð) or Croatian barred d (Đđ) is sometimes used as a substitute, but this is a workaround, and the three characters are unrelated.



Figure 17. A Gaulish tomb inscription with the name “ARADDOVNA”, illustrating the use of the TAU GALLICUM. When cased, the name would be written “Araddovna” according to modern practice..



Figure 18. A Gaulish coin of the Trinovantes/Catuvellauni tribe. Although the inscription is worn, the inscription has been read as “ADDEDOMAROS” (“Addedomaros”).

Unicode Character Properties. Character properties are proposed here.

026A;LATIN LETTER SMALL CAPITAL I;Ll;0;L;;;;N;;;A7BF;;A7BF
0287;LATIN SMALL LETTER TURNED T;Ll;0;L;;;;N;;;A7AF;;A7AF
029E;LATIN SMALL LETTER TURNED K;Ll;0;L;;;;N;;;A7AE;;A7AE
A7AE;LATIN CAPITAL LETTER TURNED K;Lu;0;L;;;;N;;;029E;
A7AE;LATIN CAPITAL LETTER TURNED T;Lu;0;L;;;;N;;;0287;
A7B0;LATIN CAPITAL LETTER BETA;Lu;0;L;;;;N;;;A7B1;
A7B1;LATIN SMALL LETTER BETA;Lu;0;L;;;;N;;;A7B0;;A7B0
A7B2;LATIN CAPITAL LETTER OMEGA;Lu;0;L;;;;N;;;A7B3;
A7B3;LATIN SMALL LETTER OMEGA;Lu;0;L;;;;N;;;A7B2;;A7B2
A7B4;LATIN CAPITAL LETTER TAU GALLICUM;Lu;0;L;;;;N;;;A7B4;
A7B5;LATIN SMALL LETTER TAU GALLICUM;Lu;0;L;;;;N;;;A7B5;;A7B5
A7BE;LATIN CAPITAL LETTER CHI;Lu;0;L;;;;N;;;AB53;
A7BF;LATIN CAPITAL LETTER SMALL CAPITAL I;Lu;0;L;;;;N;;;026A;
AB53;LATIN SMALL LETTER STRETCHED X;Ll;0;L;;;;N;;;A7BE;;A7BE

Bibliography

- Abercrombie, David. 1967. *Elements of general phonetics*. Edinburgh: Edinburgh University Press.
- Jones, Daniel. 1932. *An outline of English phonetics*. Cambridge: W. Heffer & Sons.
- Lepsius, Carl Richard. 1855. *Das allgemeine linguistische Alphabet. Nachdruck der Ausgabe Berlin 1855*. 1971 reprint: Niederwallruff bei Wiesbaden.
- Lepsius, Carl Richard. 1863. *Standard Alphabet for Reducing Unwritten Languages and Foreign Graphic Systems to a Uniform Orthography in European Letters*. Second Edition, London/Berlin.
- Riggs, Stephen Return. 1893. *Dakota grammar, texts, and ethnography*. (Department of the Interior: U. S. Geographical and Geological Survey of the Rocky Mountain Region.) Washington: Government Printing Office.

A. Administrative

1. Title

Proposal for the addition of ten Latin characters in the UCS

2. Requester's name

Michael Everson, Denis Jacquerye, Chris Lilley

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

Individual contribution.

4. Submission date

2012-07-26

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)

No.

1b. Proposed name of script

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

Yes

1d. Name of the existing block

Latin Extended-D

2. Number of characters in proposal

10.

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 A.

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, Fontographer.

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?

Michael Everson, Denis Jacquerye, Chris Lilley.

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?

Linguists, phoneticians, Africanists, Americanists, Celticists.

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

Used historically and in modern editions.

4b. Reference

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

Yes.

5b. If YES, where?

Various publications.

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

Yes.

6b. If YES, is a rationale provided?

Yes.

6c. If YES, reference

Accordance with the Roadmap. Keep with other Latin phonetic characters.

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

No.

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?

No.

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?