

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 to add one or two Latin Tironian letters to the UCS  
**Source:** Michael Everson and Andrew West  
**Status:** Individual Contribution  
**Date:** 2019-04-26  
**Replaces:** N4841 (L2/17-300)

**0. Background.** In 1999 the character U+204A TIRONIAN SIGN ET was encoded (in Unicode 3.0) as a punctuation character, based on its use in early insular manuscripts as a symbol representing the word for ‘and’ (as well as Latin ‘et’ and Irish ‘agus’ or the earlier Irish ‘ocus’). It is still used today in Irish-language texts for *agus* ‘and’, in texts and indeed is used on public signage. What was not known in 1998 was that this character was considered to be, and used as, a *letter* in the Middle Ages (being described as such and having a place at the end of the alphabet), and that the letter in fact has been treated as a *casing character* both in medieval texts and in modern representations of them. The UCS cannot at present support this usage, and three options have been described which could be considered to make up this shortcoming. In our view, only one is the right way to do it.

**1.1. Option 1: Encode one new capital letter and change the properties of U+204A.** This is our preferred solution to the problem. In order to enact it, only one character needs to be added, and the properties of one character need to be changed. Ken Whistler provided, in 2017, the requisite information needed to alter the properties of U+204A. *We do not see any technical reason why these properties should remain unchanged.* Existing users of U+204A are not likely to be particularly numerous, and none of those users would be adversely affected if a casing relationship to that character were to be added, or if its character property was changed from Po to Ll. (Many potential users of U+204A are still using DIGIT SEVEN anyway, in e-mails and other contexts, due to insufficient keyboard support for the proper character. In fact, some older databases have used U+2510 BOX DRAWINGS LIGHT DOWN AND LEFT for this—quite wrongly.)

**1.2 The required changes to U+204A if only its matching capital is added.** The current property values for U+204A are:

gc=Po, Other_Punctuation	Mirrored=No
ccc=0, Not_Reordered	Script=Common
bc=N, Other_Neutral	Grapheme_Base=Y [derived]
lb=AL (Alphabetic)	

The specific changes Ken Whistler gave for the necessary change are:

gc=Ll, Lowercase_Letter	Changes_When_Titlecased=Y [derived]
ccc=0, Not_Reordered	Lowercase=Y [derived]
bc=L, Left_To_Right	Grapheme_Base=Y [derived]
lb=AL (Alphabetic)	Changes_When_Casemapped=Y [derived]
Mirrored=No	Cased=Y [derived]
Script=Latin	
Changes_When_Uppercased=Y [derived]	
Alphabetic=Y [derived]	

With regard to the Pattern\_Syntax value, it is True for U+204A because it is in the General Punctuation block. This value cannot be changed due to stability policy. Because of the Pattern\_Syntax value for U+204A, even after the change of General\_Category to Ll, U+204A will not be usable in identifiers. *This is satisfactory to us.* There is no requirement to use this character in identifiers,.

The Script Ad Hoc group to the UTC reviewed this (L2/17-384) and came up with a most peculiar, and to us unacceptable suggestion:

Because the TIRONIAN SIGN ET character is not used in modern texts, and no evidence was given to require a case pairing (such as need for automatic capitalization), no change should be made to the property of U+204A, in our opinion.

Because the evidence was strong for the uppercase form of ET, we recommend one new character be added, TIRONIAN SIGN CAPITAL ET, and located in the Supplemental Punctuation block at U+2E4F, with properties as given for U+204A TIRONIAN SIGN ET (Po;0;ON;;;;;N;;;;;).

While we are pleased that the Script Ad Hoc group recognized that the evidence for the uppercase form was strong enough to add a new character, the suggestion to encode it without supporting its natural behaviour makes little sense. In the first place, the property it has is accidental. It was classified as a punctuation character because one of its functions is as a *logograph* meaning ‘and’ — just as U+0026 AMPERSAND is. And so it was put into the General Punctuation block. But it is never used as punctuation. See §4 below.

In fact, both Old and Middle English texts show numerous examples of the capital letter J beginning a new sentence after a full stop, or a new line of verse, and automatic capitalization in such a context is a very commonly expected behaviour in modern inputting software. (In fact some software allows the user to choose whether to turn this function on and off.) Moreover, ordinary searching tools — not ad-hoc scripted tools which are not available to all non-programmer linguists and medievalists — commonly permit users to search text looking for only lowercase letters, or for only uppercase letters, or for both together. Case insensitive searching is extremely important for scholars studying large corpora of texts or long individual texts such as the *Ormulum*. Medievalists can do this with LATIN LETTER INSULAR T (ƿ); why not with TIRONIAN ET? We see no technical reason these property changes could not be made; it is no more or less expensive to do so than other similar changes have been.

Here are all the examples of punctuation marks changing to a letter after Unicode 2.0:

0374	GREEK NUMERAL SIGN	Po => Sk (Unicode 2.1)
0374	GREEK NUMERAL SIGN	Sk => Lm (Unicode 5.1) — changed twice!
0E2F	THAI CHARACTER PAIYANNOI	Po => Lo (Unicode 2.1)
0EAF	LAO ELLIPSIS	Po => Lo (Unicode 2.1)
17D7	KHMER SIGN LEK TOO	Po => Lm (Unicode 3.2)
17DC	KHMER SIGN AVAKRAHASANYA	Po => Lo (Unicode 3.2)
3006	IDEOGRAPHIC CLOSING MARK	Po => Lo (Unicode 2.1)

There are no examples of a punctuation mark changing directly to a casing letter, but a number of these are examples of punctuation marks changing to non-casing letters. There are, however, numerous examples of non-casing letters changing to either uppercase or lowercase:

01AA	LATIN LETTER REVERSED ESH LOOP	Lo => Ll (Unicode 3.0)
01BE	LATIN LETTER INVERTED GLOTTAL STOP WITH STROKE	Lo => Ll (Unicode 3.0)
03F3	GREEK LETTER YOT	Lo => Ll (Unicode 3.0)
04C0	CYRILLIC LETTER PALOCHKA	Lo => Lu (Unicode 3.0)
01BF	LATIN LETTER WYNN	Lo => Ll (Unicode 3.0)

Indeed, we have seen the casing property of entire alphabets be changed:

13A0..13F4	CHEROKEE LETTER A..CHEROKEE LETTER YV	Lo => Lu (Unicode 8.0)
10D0..10FF	GEORGIAN LETTER AN..GEORGIAN LETTER LABIAL SIGN	Lo => Ll (Unicode 11.0)

The logic is plain:

- It is possible to change a character's property from Po to Lo;
- It is possible to change a character's property from Lo to either Ll or Lu;
- Therefore it is certainly possible to change the property of U+204A from Po to Ll.

The Script Ad Hoc group's report says "the TIRONIAN SIGN ET is not used in modern texts"—but this makes no sense. All users of the Universal Character Set, including those who study medieval European literature in palaeographic readings, are modern, living people. Characters not in use in modern orthographies of living languages are still processed by scholars using modern tools which allow casing and case-sensitive searching. In this respect there is no such thing as "not in modern use" because scholars are modern users.

There are, in fact, many characters encoded for use in medievalist palaeography which have been encoded as casing characters, and in fact, many of those characters are used in Old English and Middle English documents whose palaeography is studied. In Bodleian Library Junius 1—the manuscript containing the *Ormulum*, an early Middle English text of particular palaeographic interest due to its unique orthography—all of the text is written densely and closely, and in fact the use of capital letters is regularly used throughout to indicate the beginnings of lines of verse. Orm, the English author (who devised the orthography and wrote the manuscript) *knew* what a capital letter and a small letter was, and he used them regularly and distinctively, and with intention. In analyses of this important text, the ability to sort and search and otherwise analyse case-sensitively or case-insensitively is a reasonable function for researchers to expect, as is the ability to input text in a way that takes the actual functionality of the characters into account. The TIRONIAN ET was considered, and treated, as a casing character when it was being used to write the local language of the time.

**2. Option 2: Ignore the existing punctuation character and encode two new casing letters for medievalist use.** This solution is conservative (in terms of not altering existing character properties for an infrequently used character), but we do not believe that this solves the problem well. This option does not prevent confusion—in fact it could cause genuine confusion because it will require an *artificial* distinction between U+204A and the new lower-case character that would be added here. There is no distinction between an insular TIRONIAN ET being used either as a letter, or as a

standalone logograph. We recall some discussion about possible confusion that could occur if both U+00B7 · MIDDLE DOT and U+A78F • SINOLOGICAL DOT were encoded. Those two characters, however, *do* enjoy a genuine and non-artificial glyph distinction.

There is, however, precedent for a similar encoding strategy. Let us explore this to see if it can apply to the Tironian character.

In N2962R, the option was given to the UTC and SC2 to deal with a particular problem a natural language presented to the UCS. The existing U+0294 LATIN LETTER GLOTTAL STOP had been encoded for a long time as a character used in the International Phonetic Alphabet (IPA). Users of Canadian aboriginal languages (Chipewyan, Dogrin, and Slavey), who had been users of the original U+0294, however, had innovated a casing pair. Reasonable arguments on the conservative side were given against making any change to U+0294, in order to respect its inherent caselessness—which, significantly, also involved its shape—and therefore a new U+0241 Ꞥ LATIN CAPITAL LETTER GLOTTAL STOP and a new U+0242 ꞥ LATIN SMALL LETTER GLOTTAL STOP were added. But here, there was a genuine good argument given to leave U+0294 alone. No such argument has been given by the Script Subcommittee for U+204A.

Thus we have these three:

Ꞥ	0294	LATIN LETTER GLOTTAL STOP	(Unicode 1.1)
Ꞥ	0241	LATIN CAPITAL LETTER GLOTTAL STOP	(Unicode 4.1)
ꞥ	0242	LATIN SMALL LETTER GLOTTAL STOP	(Unicode 4.1)

No one is genuinely unhappy about this. The ordinary phonetic use of U+0294 has not been adversely affected by the change. Comparing Options 1 and 2, however, it is clear to us that only first is the more economical and preferable solution.

Ꞥ	A7C9	LATIN CAPITAL LETTER TIRONIAN ET
Ꞥ	204A	TIRONIAN SIGN ET

Option 2 would require an entirely *artificial* glyph distinction in the code charts and in fonts—which is plainly objectionable since there is no genuine distinction between a Tironian logograph (‘and’ is a word, not a punctuation mark) and a Tironian letter (polyvalent just as U+A753 ꝥ LATIN SMALL LETTER P WITH FLOURISH is, being used for *pro* or *por*).

Ꞥ	A7C8	LATIN CAPITAL LETTER TIRONIAN ET
Ꞥ	A7C9	LATIN SMALL LETTER TIRONIAN ET (glyph shape <i>quite</i> uncertain)
Ꞥ	204A	TIRONIAN SIGN ET

Moreover it is a waste of a code point, since the arguments for keeping U+0294 distinct from the Canadian Aboriginal characters do not apply here: the LATIN SMALL LETTER TIRONIAN ET is in fact identical to TIRONIAN SIGN ET—there is only one character (in a case pair) used in Britain and Ireland in the insular script. To encode it would be to encode a duplicate character—which is not what the UCS is supposed to do.



Today's default UCS order is somewhat different. The Ț TIRONIAN SIGN ET is ordered as punctuation, after & AMPERSAND:

... / < \ < & < Ț < # < % ...

The relative ordering of the TIRONIAN ETs should be changed if U+204A is changed to be a letter and it seems sensible to put them, interfiled, between Norse ƿ VEND (which derives from Þ WYNN) and 3 LATIN LETTER ET:

A << (Æ) < B < C < D << Ð < E < F < G < H < I < J < K < L < M < N < O < P < Q <  
R < S < T < U < V < X < Y < Z < Þ < ƿ < Ț <<< J < 3 < ƒ ...

## 7. Unicode Character Properties. Character properties are proposed here.

A7C9;LATIN CAPITAL LETTER TIRONIAN ET;Lu;0;L;;;;;N;;;204A;  
204A;TIRONIAN SIGN ET;Ll;0;L;;;;;N;;;A7C9;;A7C9

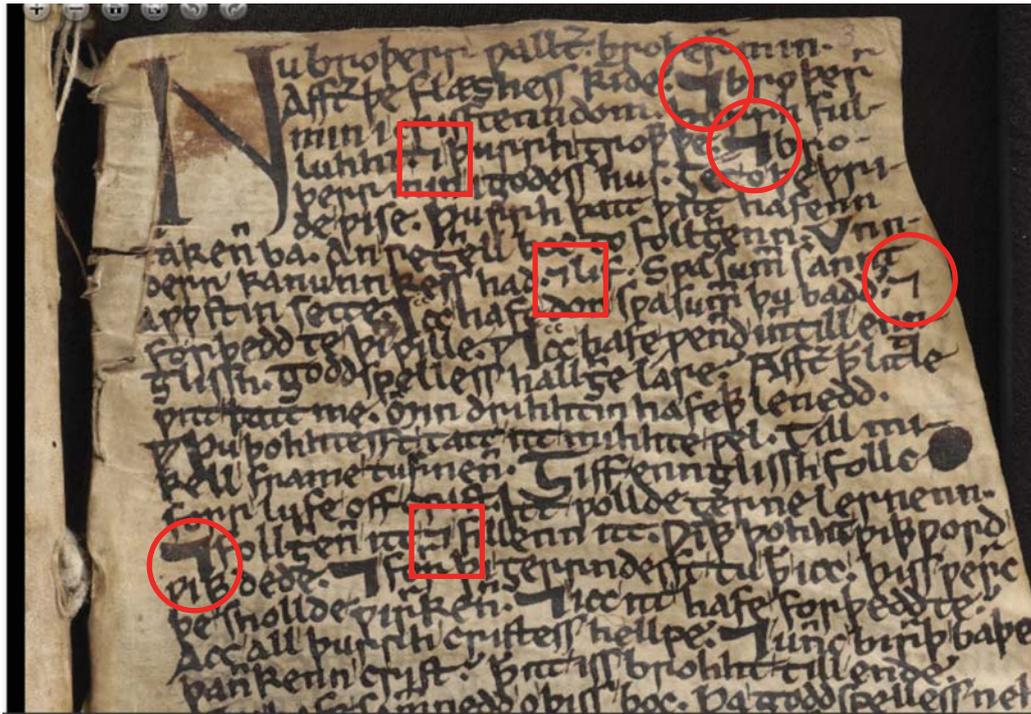
or (in the case of Option 2):

A7C8;LATIN CAPITAL LETTER TIRONIAN ET;Lu;0;L;;;;;N;;;A7C9;  
A7C9;LATIN SMALL LETTER TIRONIAN ET;Ll;0;L;;;;;N;;;A7C8;;A7C8

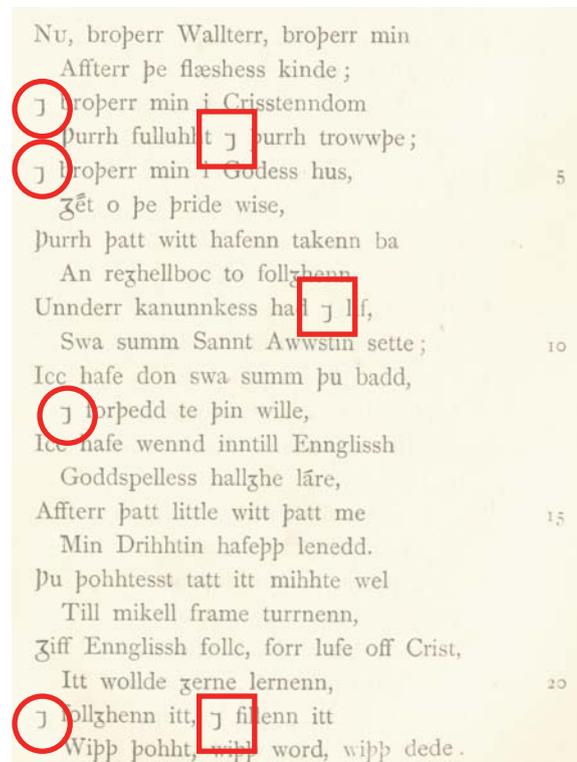
## 8. Bibliography

- Conner, Patrick W. 1993. *Anglo-Saxon Exeter: A Tenth-century Cultural History*. Woodbridge: Boydell Press. ISBN 978-0-85115-307-0
- S. J. Crawford, S. J. 1929. *Byrhtferth's Manual (A.D. 1011), now edited for the first time from MS. Ashmole 328 in the Bodleian Library*. (Early English Text Society: 177). London: Oxford University Press.
- EM 3501, Exeter Cathedral Library.
- Hall, Joseph. 1920. *Selections from Early Middle English, 1130–1250*. Part I: Texts. Oxford: Clarendon Press.
- Robinson, A. J. 1956. *Anglo-Saxon Charters*. Cambridge: Cambridge University Press. 2009 reprint ISBN 978-0-521-17832-7
- White, R. M., and Robert Holt, eds. 1878. *The Ormulum, with the notes and glossary of Dr R. M. White*. 2 volumes. Oxford: Clarendon Press.
- Wright, C. E. 1960. *English vernacular hands from the twelfth to the fifteenth centuries*. (Oxford Palaeographical Handbooks.) Oxford: Clarendon Press.

Figures.



**Figure 1.** Old English text, the *Ormulum*, written early 13th century, Oxford Bodleian Library, Junius MS 1, fol. 3r. At the beginnings of sentences the letters N, A, P, S, V, S, I, S (for which see N5043), M, T, and Þ are given alongside capital J. Capital J is circled and small 7 boxed in red. The *Ormulum* is written in a very precise orthography and is an extraordinarily important source of our knowledge of Middle English pronunciation.



**Figure 2.** Old English text, the *Ormulum*, written early 13th century, Oxford Bodleian Library, Junius MS 1, fol. 3r. In this transcription a distinction between capital J and small 7 has not been made, likely due to typographic constraints. From White & Holt 1878:xcvii.

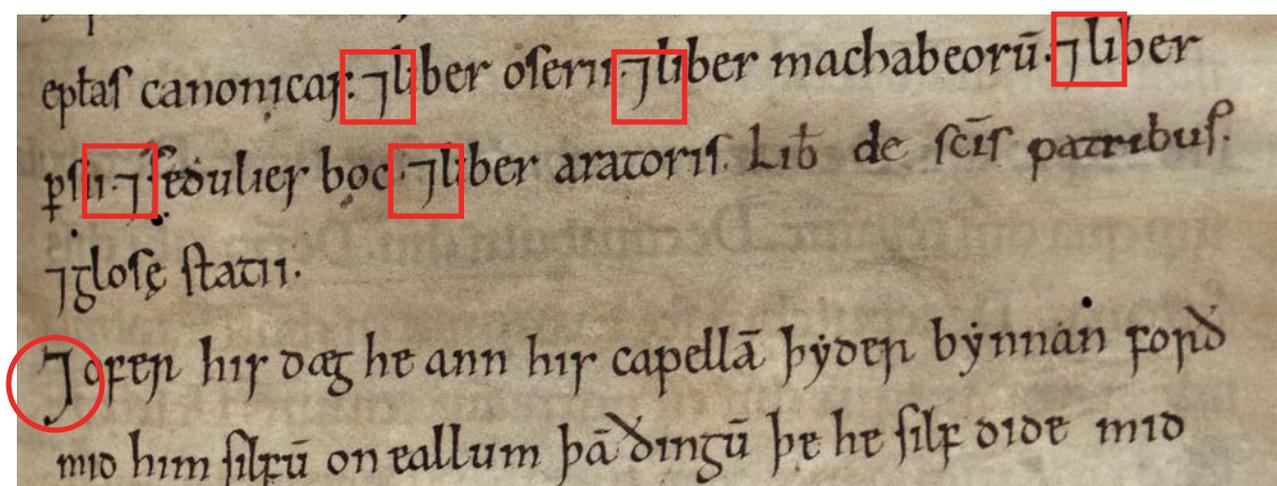


Figure 3. Latin and Old English text, “The Gifts of Bishop Leofric”, written early 13th century, Exeter Cathedral Library, MS 3501, fol. 2r. At the beginnings of sentences the capital J can clearly be seen. (There are other examples of this in the manuscript as well.)

Bede super apocalipsin,<sup>57</sup> 7 expositio Bede super .vii. epistolas canonicas,<sup>58</sup> 7 liber Oserii, 7 liber Machabeorum, 7 liber Persii,<sup>59</sup> 7 Sedulies boc, 7 liber Aratoris.<sup>60</sup> Liber<sup>61</sup> de sanctis patribus,<sup>62</sup> 7 glose Statii.<sup>63</sup>

[IV. Statement of purpose and anathema]

[12] 7 ofer his dæg he ann his capellam þyder bynnan forð mid him silfum on eallum þam dingu þe he silf dide mid Godes ðeninge, on þæt gerad þæt þa

Figure 4. Transcription of the Bishop Leofric text from Conner 1993. Here, with 7-shaped glyphs, lowercase 7 and uppercase J are distinguished by size and position.

25 7 liber Isidori de miraculis Xpi · 7 liber Oserii · 7 liber Machabeorum · 7 liber Persii · 7 Sedulies boc · 7 liber Aratoris · 7 diadema monachorum<sup>8</sup> · 7 glose Statii · 7 liber officialis Amalarii.  
 J ofer his dæg he ann his capellam þider binnan forð mid him silfum · on eallum þam þingu þe he silf dide mid Godes

Figure 5. Transcription of another MS containing the Bishop Leofric text, from an edition of the Anglo-Saxon charters (Robinson 1956:228). Here both lowercase 7 and uppercase J are clearly distinguished, just as they are in the manuscript. The shapes here exemplify good practice for glyph shapes for the two characters (though the lowercase one here is a bit too 7-like).

book *De Miraculis Christi* and the book of Oserius and the *Liber Machabeorum* and Persius' book and Sedulius' book and Arator's book and the *Diadema Monachorum* and the glosses of Statius and the *Liber Officialis* of Amalarius.

And after his death he grants his *capella* to the minster along with himself, as far as concerns all the things with which he himself performed the service of God, on condition that the

Figure 6. Translation of the text in Figure 5 (Robinson 1956:229).



**Figure 7.** Left, the TIRONIAN SIGN ET used in public signage over the door of the Warwick Hotel in Salthill Galway (it reads :”Beár 7 Bialann” ‘Bar & Restaurant’), and right, the same in use on public road signage. The glyph on the parking sign is particularly nice as there is a slight curve to the topbar and the glyph isn’t very 7-shaped.

Photo by Stan Carey <https://stancarey.wordpress.com/2014/09/18/the-tironian-et-in-galway-ireland/>



**Figure 8.** The TIRONIAN SIGN ET used on a parking sign in Ireland. Photo by Kaihsu Tai.



**Figure 9.** The TIRONIAN SIGN ET used in public signage in Ireland, on the left on a telephone manhole cover, and on the right on a postbox, indicating “P 7 T” (*Post agus Teileagraf* ‘Post and Telegraph’).

## A. Administrative

1. Title

**Proposal to add one or two Tironian Latin letters to the UCS**

2. Requester's name

**Michael Everson and Andrew West**

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

**Individual contribution.**

4. Submission date

**2019-04-26**

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

**1 or 2.**

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.

**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?

**Peter Baker, the Anglo-Saxon discussion group ansax-l@listserv.wvu.edu, the Chaucer discussion group chaucer@listserv.uic.edu, Facebook groups for Old and Middle English. The authors are members of the user community.**

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?

**Medievalists, Celticists, Anglicists, Germanicists.**

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

**Common in linguistic texts.**

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 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?