

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

**Doc Type:** Working Group Document  
**Title:** Revised proposal to add two characters for Middle English to the UCS  
**Source:** Michael Everson and Andrew West  
**Status:** Individual Contribution  
**Date:** 2023-06-01  
**Replaces:** N5145 (L2/20-268), N5043 (L2/19-178)

This proposal requests the encoding of two spacing characters used for the early Middle English Ormulum. If this proposal is accepted, the following characters will exist:

ᚷ	A7D2	LATIN CAPITAL LETTER DOUBLE THORN
ᚹ	A7D4	LATIN CAPITAL LETTER DOUBLE WYNN

Although the proposal made in N5145 was successful, the two capital letters here were not encoded, although space was made for them in the code chart. There has been no correspondence or discussion about this, though we assume that this was done due to the absence of 12th-century examples of the characters in use in the document.

The author of the early Middle English Ormulum, Orm, who called himself both Orrm and Orrmin, had devised an orthography for English which expressed distinctions between long and short vowels, and expressed precise distinctions between certain consonants. His orthography is remarkable in its accuracy; it precedes more formal phonetic analyses by centuries. Though the work has some lacunae, it consists of over 20,000 lines of verse; the metrical nature of the work also assists in our understanding of the phonology of this dialect of Middle English.

The manuscript also contains numerous passages in Latin, and the scribe (who was Orm himself) used two distinctive styles of writing for each. When writing Latin, Orm uses a Carolingian hand with numerous signs of abbreviation and with a number of standard ligatures. He does not use these ligatures in his Middle English text, which is written in Insular script and is quite distinct from the standard Latin hand and orthography. It is for this reason that the palaeography of Orm's orthography is both interesting and important. A palaeographic edition is being prepared, and in the preparation a number of characters missing from the UCS have been identified. The missing characters are proposed for encoding here.

1. DOUBLE WYNN AND DOUBLE THORN. The two runic borrowings into Insular script, WYNN ᚹᚹ and THORN ᚷᚷ, get special treatment in Orm's orthography. Instead of trying to squeeze a combining letter atop (in order to indicate a short vowel), Orm devised double letters where the two bowls share a single vertical stem: proposed here as ᚷᚷ and ᚹᚹ. These letters are extremely frequent throughout the manuscript (DOUBLE THORN being the more frequent). Orm writes *piᚷ* as easily as he writes *pipᚷ*

‘with’, and *tropþe* as easily as *troppþe* ‘belief’. He does not write a \*COMBINING THORN over *þ* or a \*COMBINING WYNN over *p*; neither of those combining characters is encoded and neither is required for the *Ormulum*.

Þ þ Æ æ ƿ ƿ ß ß

**2. Capitalization.** Capitalization in Orm’s manuscript is quite rigorously applied at the beginning of every verse. In fact, since the manuscript is really rather cramped, the capitalization provides an important cue to reading the text.

In N5145, casing pairs for all of the letters used in Orm’s orthography were proposed, as they have been for other palaeographic letters encoded in the Latin Extended-D block. The reason unattested casing forms for old orthographies have been proposed (and accepted) for encoding in the past has been for the use of modern scholars. Obviously the Universal Character Set is not just for reproduction of the words written down by people long ago. Modern people may wish to use the words from medieval documents in a variety of contexts. Here are two examples of this.

In the first line of each group, an article title is given. In the second two lines, the same is given as an editor might require it to be typeset in the header of the book or journal.

“Reflexes of the voiced velar in the Ormulum: ʒiff, ʒodd, and seggen”

REFLEXES OF THE VOICED VELAR IN THE ORMULUM: ʒIFF, ʒODD, AND SEGGEN

REFLEXES OF THE VOICED VELAR IN THE ORMULUM: SIFF, ʒODD, AND SEGGEN

Marking short vowels in the Ormulum: ƿiþþ and ƿiþ; troppþe and troþþe”

MARKING SHORT VOWELS IN THE ORMULUM: ƿIþþ AND ƿIþ; TROPPþE AND TROþþE

MARKING SHORT VOWELS IN THE ORMULUM: ƿIþþ AND ƿIþ; TROPPþE AND TROþþE

With respect this second case, note that the following example—with no capital DOUBLE THORN OR DOUBLE WYNN—would be incorrect and unacceptable.

MARKING SHORT VOWELS IN THE ORMULUM: ƿIþþ AND ƿþ; TROPPþE AND TROþþE.

MARKING SHORT VOWELS IN THE ORMULUM: ƿIþþ AND ƿþ; TROPPþE AND TROþþE.

This is as unacceptable as writing \*MAßSTAB or \*MAßSTAB rather than MAßSTAB or MAßSTAB would be (MASSSTAB and MASSSTAB are orthographically equivalent to ƿIþþ, ƿIþþ, TROPPþE, and TROPPþE in this context). Without the encoded casing pair, it is not possible to accommodate such an ordinary editorial and design decision.

In Orm’s orthography. DOUBLE THORN and DOUBLE WYNN are typically used following a vowel, as doubled consonants are used to indicate vowel length. Orm is scrupulous about capitalization; indeed, his casing of TIRONIAN ET is regular and helped to get it encoded. It has not been possible to examine all 110,000 words in the 20,000 lines of verse in the Ormulum. It is certainly possible that Orm didn’t write any words in all-capitals in this manuscript (or if he did, it might be in Latin only). But had he done so, he would certainly have written both CAPITAL DOUBLE THORN and CAPITAL DOUBLE WYNN. He knew what capital letters are, and how to use them. *But Orm is long dead.* Orm doesn’t care about using Unicode or ISO/IEC 10646. It is modern scholars like us who wish to

produce palaeographic or expanded editions of the Ormulum, to cite words and phrases from the Ormulum, and indeed to use them in titles as given in the completely likely examples given above.

Latin letters used in natural orthographies are naturally casing. This is a standard structural function of the Latin script. We have encoded many casing pairs for natural orthographies in the past and we should not have to have the same argument every time new characters are proposed for encoding.

**3. Ordering.** We recommend the following.

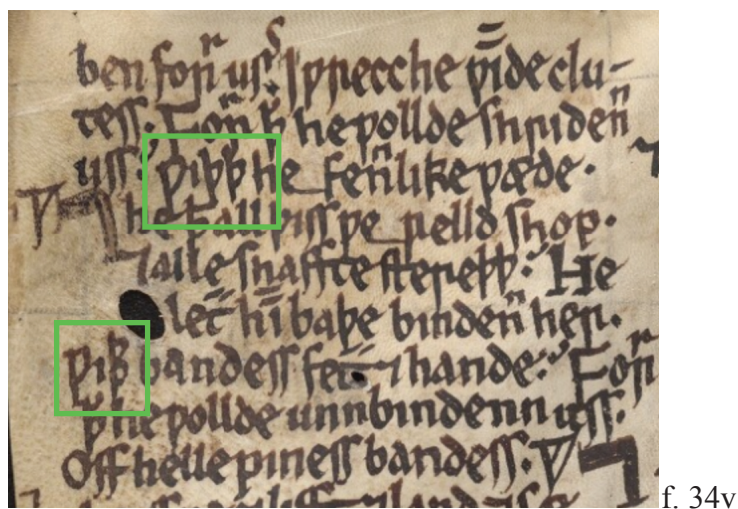
... þ << Þ < **þ** << **Þ** < þ << Þ < þ << Þ < p << P < **p** << **P** < ʝ << ʝ ...

**4. Security.** Neither of these characters is required in identifiers.

**5. Unicode Character Properties.** Character properties are proposed here.

```
A7D2;LATIN CAPITAL LETTER DOUBLE THORN;L1;0;L;;;;N;;;A7D3;  
A7D3;LATIN SMALL LETTER DOUBLE THORN;L1;0;L;;;;N;;;A7D2;;A7D2  
A7D4;LATIN CAPITAL LETTER DOUBLE WYNN;L1;0;L;;;;N;;;A7D5;  
A7D5;LATIN SMALL LETTER DOUBLE WYNN;L1;0;L;;;;N;;;A7D4;;A7D4
```

**6. Figure.**



**Figure 1.** Text from the Ormulum f. 34v, lines 3674–3683, showing a SMALL DOUBLE THORN in the word *piþ* ‘with’ and two thorns in *piþþ*.

If a user of the standard can write WITH and *PIÞÞ* alongside *with* and *piþþ*, a user of the standard should be able to write *PIÞ* alongside *piþ*. We have encoded this sort of thing many times.

## A. Administrative

1. Title

**Revised proposal to add two characters for Middle English 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

**2023-06-01**

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

**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. 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.

**Yes. N5145.**

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?

**The authors of this proposal 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?

**Germanicists, Anglicists, dialectologists, linguists.**

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

**Used in the Ormulum, a unique but very important record of Early Middle English; also used in publications about it and extracts from it.**

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?

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?