

## ***Appendix D: How to Submit New Characters for Consideration***

Unicode Inc. accepts proposals for inclusion of new characters or scripts in the Unicode standard. All proposals must be in writing, must include at least one picture of the proposed character (normally from a printed source), and must include documentation justifying the proposal. The identification of the sponsor(s) must be included, along with a postal or electronic mail address, and/or a phone number. Guidelines for the preparation of a proposal appear below.

The Unicode standard definition of character is stated in the Glossary. In particular, the distinction between the terms *character* and *glyph* should be noted. Because of this distinction, graphics such as ligatures, conjunct consonants, minor variant written forms, or abbreviations of longer forms are generally not acceptable as Unicode characters. Each proposal will be evaluated by the editorial subcommittee of the Unicode Technical Committee. The result of this evaluation will be communicated to the sponsor(s) of the proposal. All proposals (whether successful or not) will be retained by the Unicode Consortium. This archive will be indexed. The Unicode Consortium is also interested in obtaining information on known glyphs, minor variants, ligatures, conjunct consonants, and other such “non-characters,” mainly for cataloging and research purposes.

*Send proposals to:*

Unicode Incorporated  
c/o Metaphor Computer Systems  
1965 Charleston Avenue  
Mountain View, CA 94043  
USA

*E-mail contact:*

Rick\_McGowan@NeXT.COM  
FAX: USA 415-780-3714

The sponsor proposing the addition of a new character to the Unicode standard should follow these guidelines:

1. Determine that the proposed addition is a character according to the definition given in the Unicode standard.
2. Determine that the proposed addition does not already exist in the Unicode standard. The sponsor must both take into account different names by which the character may be known and minor glyphic variation in the Unicode illustration, and ensure that documentation supporting

the proposal states whether any Unicode characters were examined as possible equivalents for the proposed character and, if so, why each was rejected.

3. Include at least one picture of the proposed character, showing it in a printed context. Photocopies are acceptable. The sponsor must cite the source of each picture. The requirement for an illustration from a printed source is waived in the case of historical characters that do not generally appear in printed contexts and characters in modern use by cultures that don't have or use printing.
4. List the proposed name or alternative name, in the Unicode name list style. The proposed name should be typed in upper case. Each alternative name should be preceded by an equal sign, and be entirely lower case.
5. List any known commercially available fonts that include the character. Give as much information as possible on where and how such fonts may be obtained.
6. Provide any additional information, including other sources of printed examples, and works about the character which discuss it or define it. Who uses (or did use) it? When? In what context? For what purpose? List any known information about the character and its place in current culture or historical perspective.
7. If the character is a Han (ideographic) character, strong evidence for its uniqueness, indispensability, and value to users must be submitted, including printed contexts where it is used. If it could be considered a variant of a character already included in the Unicode standard, the proposed character should be shown in a context that demonstrates why it must be distinguished as a separate character and not considered a variant of the character already included.
8. If the character is part of a dead language or obsolete/rare script, cite the most important modern sources of information on the script. Names (including academic affiliation) of researchers in the relevant field are welcomed.
9. Proposals to include entire scripts (Syriac and Egyptian hieroglyphics, for example) should also cite modern, definitive sources of information regarding such scripts. Sponsorship by the relevant academic bodies (such as *The International Association of Egyptologists*) may be helpful in determining the proper scope for encoding of characters in such cases.

In the meantime, there are ways for programmers and scholarly organizations to make use of Unicode character encoding, even if the script they want to use or transmit is not yet part of the Unicode standard.

The Private Use Area can always be used to store and transmit characters, with the understanding that the encoding should be changed to the standard one, if and when the character or script ends up being included in a future version. A few living scripts are very likely to be included in the near future, including Ethiopian, Mongolian, Sinhala, Burmese, Khmer and Cree/Inuktitut.

In the case where a rare script has been turned down, individual groups that make use of rare scripts can reach a private agreement about setting aside part of the Private Use Area to encode their private set of characters. This leaves open the question of what to do with character sets that are simply too big to fit into the Private Use Area; in that case, your group might consider an entire private encoding in which to exchange data. The Unicode Consortium is generally most predisposed toward early inclusion of scripts that are in modern use, however minor or rare (such as Maldivian); behind those rank the scripts which are clearly extinct, important as they may be (such as Demotic); behind those rank a number of scripts of more or less historical significance (such as Linear B); and behind those lie the multitude of other scripts that have been used at one historical locus (such as Rongo-Rongo). The fact that a script is turned down for inclusion is not necessarily indicative of its ultimate fate. Some important scripts may be set aside for later inclusion because of other more immediate needs. Some criteria in ranking items for potential inclusion may have to do with factors such as the documented level to which the script is already used with computers, or the overall size of its literature.

