

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 Latin characters for Marshallese
Source: Michael Everson
Status: Individual Contribution
Date: 2013-06-25

This proposal begins by requesting the encoding of four Latin characters comprising two casing pairs, for support of the Marshallese language. This proposal is in response to N4458 “Cedillas and commas below” and N4459 “Latvian and Livonian glyphs with commaaccent in the Unicode Standard”; the four characters listed below are proposed here in accordance with N4456 “Latvian and Marshallese Ad Hoc Report”. If this proposal is accepted, the following characters will exist:

Ł	A7BA	LATIN CAPITAL LETTER MARSHALLESE L WITH CEDILLA x 013B Ł latin capital letter l with cedilla • character has no decomposition
ł	A7BB	LATIN SMALL LETTER MARSHALLESE L WITH CEDILLA x 013C ł latin small letter l with cedilla • character has no decomposition
Ń	A7BC	LATIN CAPITAL LETTER MARSHALLESE N WITH CEDILLA x 0145 Ń latin capital letter l with cedilla • character has no decomposition
ń	A7BD	LATIN SMALL LETTER MARSHALLESE N WITH CEDILLA x 0146 ń latin small letter l with cedilla • character has no decomposition

Issues. This proposal succeeds in solving a particular problem: that Marshallese Ł, ł, Ń, and ń cannot currently be correctly displayed in the UCS because the implemented glyphs for U+004C + U+0327, U+006C + U+0327, U+004E + U+0327, U+006E + U+0327, U+013B, U+013C, U+0145, U+0146 are (and should be for stability) be Ł, ł, Ń, ń. The problem for the UCS is that the other letters used in Latvian and Livonian also prevent “proper” cedillas to be displayed with them: Đ, đ, Ģ, ģ, Ķ, ķ, Ŗ, and ŗ. At least some of these characters, with “proper” cedillas, are known to be specified in bibliographic and geographic transliteration and transcription systems. In light of this, because it makes no sense for users of the UCS to be unable to represent these letters with actual cedillas, I suggest that instead of the four letters and names given above, the following 12 characters be encoded:

Ł	A7BA	LATIN CAPITAL LETTER L WITH INVARIANT CEDILLA <ul style="list-style-type: none"> • used in Marshallese x 013B Ł latin capital letter l with cedilla <ul style="list-style-type: none"> • character has no decomposition
ł	A7BB	LATIN SMALL LETTER L WITH INVARIANT CEDILLA <ul style="list-style-type: none"> • used in Marshallese x 013C ł latin small letter l with cedilla <ul style="list-style-type: none"> • character has no decomposition
Ń	A7BC	LATIN CAPITAL LETTER N WITH INVARIANT CEDILLA <ul style="list-style-type: none"> • used in Marshallese x 0145 Ń latin capital letter n with cedilla <ul style="list-style-type: none"> • character has no decomposition
ń	A7BD	LATIN SMALL LETTER N WITH INVARIANT CEDILLA <ul style="list-style-type: none"> • used in Marshallese x 0146 ń latin small letter n with cedilla <ul style="list-style-type: none"> • character has no decomposition
Đ	A7BE	LATIN CAPITAL LETTER D WITH INVARIANT CEDILLA x 1E10 Đ latin capital letter d with cedilla <ul style="list-style-type: none"> • character has no decomposition
đ	A7BF	LATIN SMALL LETTER D WITH INVARIANT CEDILLA x 1E11 đ latin small letter d with cedilla <ul style="list-style-type: none"> • character has no decomposition
Ĝ	A7C0	LATIN CAPITAL LETTER G WITH INVARIANT CEDILLA x 0122 Ĝ latin capital letter g with cedilla <ul style="list-style-type: none"> • character has no decomposition
ĝ	A7C1	LATIN SMALL LETTER G WITH INVARIANT CEDILLA x 0123 ĝ latin small letter g with cedilla <ul style="list-style-type: none"> • character has no decomposition
Ķ	A7C2	LATIN CAPITAL LETTER K WITH INVARIANT CEDILLA x 0136 Ķ latin capital letter k with cedilla <ul style="list-style-type: none"> • character has no decomposition
ķ	A7C3	LATIN SMALL LETTER K WITH INVARIANT CEDILLA x 0137 ķ latin small letter k with cedilla <ul style="list-style-type: none"> • character has no decomposition

- R** A7C4 LATIN CAPITAL LETTER R WITH INVARIANT CEDILLA
 x 0156 **R̥** latin capital letter r with cedilla
 • character has no decomposition
- r̥** A7C5 LATIN SMALL LETTER R WITH INVARIANT CEDILLA
 x 0157 **r̥** latin small letter r with cedilla
 • character has no decomposition

A. Administrative

1. Title

Proposal for the addition of Latin characters for Marshallese

2. Requester's name

Michael Everson

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

Individual contribution.

4. Submission date

2013-06-25

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

4, or 12.

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.)?

No.

2b. If YES, with whom?

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?

Speakers of Marshallese; users of transliteration and transcription systems.

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

Ordinary use.

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?