Network Working Group
Internet-Draft
Obsoletes: 7997 (if approved)
Intended status: Informational
Expires: 17 August 2026

P. Hoffman

13 February 2026

Text in RFCs
draft-rswg-rfc7997bis-09

Abstract

This document sets policy for the inclusion of characters in the definitive versions and publication formats of RFCs. The policy for the RFC Series is that all displayable text is allowed as long as there is a high expectation that readers of an RFC will be able to interpret its text as intended. This document obsoletes RFC 7997.

[[A repository for this draft can be found here (https://github.com/paulehoffman/7997bis).]]

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 17 August 2026.

Table of Contents

1. Introduction

The early policy for the RFC Series was that RFCs could only contain characters from the ASCII character set. Later policies, from [RFC7997], allowed more characters, set the language of the RFC Series to be English, and set the encoding for RFCs of UTF-8. In the time since [RFC7997] was published, the IETF community has had much more experience of using non-ASCII characters in RFCs.

This document obsoletes [RFC7997]. This document makes substantial changes to the policies in [RFC7997] based on the positive experience since its publication.

The RFC Publication Center (RPC) is responsible for implementing the policies in this document, as described in [RFC9720]. The RPC style guides may define which characters authors may use and how they are used.

1.1. Terminology

The term "non-ASCII characters" means characters outside the set that was defined in ASCII. ASCII is described in [RFC20].

The term "Unicode characters" means characters defined in [UnicodeLatest].

"U+ notation" means using the characters "U+" and a hexadecimal number to represent a Unicode code point. See [BCP137] for more on U+ notation.

More terminology about characters and encoding formats can be found in [RFC6365].

2. Basic Requirements for Text in RFCs

RFCs should only contain text that can be displayed correctly across a wide range of readers and browsers. People whose systems do not have the fonts needed to display part of a particular RFC still need to be able to read the definitive versions and publication formats correctly in order to understand and implement the information described in the document.

The ability to use non-ASCII characters in RFCs in a clear and consistent manner will allow the correct display of proper names and improve the ability to describe internationalized protocols. Apart from their role in proper names, non-ASCII characters should be used only when they enhance the technical content and accuracy of the document.

3. Policy for Text in RFCs

English is the required language of RFCs. However, because non-ASCII characters are often required for instances including proper names and examples, the policy for the RFC Series is that all displayable text is allowed as long as there is a high expectation that readers of an RFC will be able to interpret its text as intended. Apart from their role in proper names, non-ASCII characters should be used only when they enhance the technical content and accuracy of the document.

There are many Unicode characters that obviously cannot be displayed (such as control characters), and many whose ability to be displayed is debatable. If an RFC includes such characters in normative or descriptive text, the RFC needs to also clearly describe the character.

The preferred method for describing such characters is using the U+ notation from [BCP137] and/or using the character’s official name from the Unicode Standard [UnicodeLatest]. [BCP137] describes the pros and cons of different options for identifying Unicode characters and may help authors decide how to represent the non-ASCII characters in their documents.

Note that this policy only applies to normative or descriptive text; text such as names does not need character description. Further, some RFC authors might choose to use something other than the U+ notation to describe characters, such as if the RFC already covers a different syntax that the reader will understand from the rest of the RFC.

Characters in an RFC will generally appear in Normalization Form C (NFC) as defined in [UnicodeLatest]. If the RFC would be more correct and more understandable with particular characters not in NFC, the RPC can use unnormalized text. In such a case, a note should be included to describe why unnormalized text was used.

3.1. Names

Authors of RFCs whose names include non-ASCII characters will likely have preferences for how their names are displayed. Authors can give their names using only Latin script characters, or using non-Latin script and an equivalent in Latin script. Authors' preferences for display of their names should be honored.

Company names and geographic names generally do not need Latin equivalents, but they can be included at the discretion of the author and the RPC.

3.2. Examples

Where the use of non-ASCII characters is purely part of an example or not otherwise required for correct protocol operation, giving the Unicode code points and Unicode names of the non-ASCII characters is not required, but it can improve the readability of the RFC. An RFC can use either or both forms, whichever is sensible in the circumstance. For example, for text that might just say "The value can be followed by a monetary symbol such as ¥ or ", text that says one of the following is likely more beneficial to the reader:

* The value can be followed by a monetary symbol such as ¥ (U+00A5) or (U+20AC)

* The value can be followed by a monetary symbol such as ¥ (YEN SIGN) or (EURO SIGN)

* The value can be followed by a monetary symbol such as ¥ (U+00A5, YEN SIGN) or (U+20AC, EURO SIGN)

RFCs may be viewed using only black and white or grayscale, particularly when printed. Because of this, examples should generally use characters that do not specify a color. However, some examples might require text with color due to the nature of the examples. If so, those examples need to also include U+ notation. For example, "A color display should be able to differentiate (U+1F534, LARGE RED CIRCLE), 泙 (U+1F7E2, LARGE GREEN CIRCLE), and 鳩 (U+1F535, LARGE BLUE CIRCLE)."

4. RFC Publication Language

The RFC publication language is English.

5. RFC Publication Encoding

The encoding format for the RFC Series is UTF-8 [STD63].

6. IANA Considerations

This document contains no IANA considerations.

7. Security Considerations

Authors and the RPC should cross-check that the used characters match their code point numbers or Unicode character names.

8. References

8.1. Normative References

[BCP137] Best Current Practice 137, <https://www.rfc-editor.org/info/bcp137>. At the time of writing, this BCP comprises the following:

Klensin, J., "ASCII Escaping of Unicode Characters", BCP 137, RFC 5137, DOI 10.17487/RFC5137, February 2008, <https://www.rfc-editor.org/info/rfc5137>.

[RFC7997] Flanagan, H., Ed., "The Use of Non-ASCII Characters in RFCs", RFC 7997, DOI 10.17487/RFC7997, December 2016, <https://www.rfc-editor.org/rfc/rfc7997>.

[RFC9720] Hoffman, P. and H. Flanagan, "RFC Formats and Versions", RFC 9720, DOI 10.17487/RFC9720, January 2025, <https://www.rfc-editor.org/rfc/rfc9720>.

[STD63] Internet Standard 63, <https://www.rfc-editor.org/info/std63>. At the time of writing, this STD comprises the following:

Yergeau, F., "UTF-8, a transformation format of ISO 10646", STD 63, RFC 3629, DOI 10.17487/RFC3629, November 2003, <https://www.rfc-editor.org/info/rfc3629>.

[UnicodeLatest] The Unicode Consortium, "The Unicode Standard", n.d., <http://www.unicode.org/versions/latest/>.

8.2. Informative References

[RFC20] Cerf, V., "ASCII format for network interchange", STD 80, RFC 20, DOI 10.17487/RFC0020, October 1969, <https://www.rfc-editor.org/rfc/rfc20>.

[RFC6365] Hoffman, P. and J. Klensin, "Terminology Used in Internationalization in the IETF", BCP 166, RFC 6365, DOI 10.17487/RFC6365, September 2011, <https://www.rfc-editor.org/rfc/rfc6365>.

Appendix A. Acknowledgements

This document is based on [RFC7997] which was authored by Heather Flanagan.

The acknowledgements from [RFC7997] are to the members of the IAB i18n program, to the RFC Format Design Team: Nevil Brownlee, Tony Hansen, Joe Hildebrand, Paul Hoffman, Ted Lemon, Julian Reschke, Adam Roach, Alice Russo, Robert Sparks, and Dave Thaler.

Writing this document was greatly helped by contributions from the RFC Series Working Group (RSWG), including: Brian Carpenter, Carsten Bormann, Eliot Lear, John Klensin, John Levine, Martin Dテシrst, Martin Thomson, Pete Resnick, Rob Sayre, and Russ Housley.

Author’s Address

Paul Hoffman
Email: paul.hoffman@icann.org