

Network Working Group
Internet-Draft
Obsoletes: 7997 (if approved)
Updates: 7322 (if approved)
Intended status: Informational
Expires: 17 March 2026

P. Hoffman
13 September 2025

Text in RFCs
draft-rswg-rfc7997bis-04

Abstract

The early policy for the RFC Series was that RFCs could only contain characters from the ASCII character set. Later policy, from RFC 7997, allowed more characters and enforced an encoding for RFCs of UTF-8. Since RFC 7997 was published, the IETF community has had much more experience of using non-ASCII characters in RFCs.

The policy for the RFC Series is that all displayable text is allowed as long as the reader of an RFC can interpret that text. This policy does not change language policy of the RFC Series, namely that English is the required language for the series.

This document obsoletes RFC 7997 and updates the RFC Style Guide (RFC 7322).

[[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 March 2026.

Copyright Notice

Copyright (c) 2025 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document.

Table of Contents

1. Introduction	2
1.1. Terminology	2
2. Basic Requirements for Text in RFCs	3
3. Policy for Text in RFCs	3
3.1. Names	4
3.2. Examples	4
4. IANA Considerations	4
5. Security Considerations	4
6. References	4
6.1. Normative References	4
6.2. Informative References	5
Appendix A. Acknowledgements	5
Author's Address	6

1. Introduction

This document sets policy for the inclusion of characters in the definitive versions and publication formats of RFCs. It also reaffirms the policy that the encoding format for the RFC Series is UTF-8, [STD63]. This document obsoletes [RFC7997] and updates the RFC Style Guide [RFC7322]. 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].

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 define in [UnicodeCurrent].

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

2. Basic Requirements for Text in RFCs

RFCs should 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.

As stated in the RFC Style Guide [RFC7322], the language of the RFC Series is English.

Searches whose results might include RFCs should return accurate results and support appropriate Unicode string matching behaviors.

3. Policy for Text in RFCs

The policy for the RFC Series is that all displayable text is allowed as long as the reader of an RFC can interpret that text.

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+NNNN" syntax from [BCP137]. [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+NNNN" syntax to describing 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 [UnicodeNorm]. If the RFC would be more correct and more understandable with particular characters not in NFC, the RFC can use unnormalized text. In such a case, a text 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 based on their lived experiences. These authors can give their names using only ASCII characters, or as Unicode characters and an ASCII interpretation of their name. The RPC policy should be that authors' preferences for display of their names be honored.

Company names and geographic names generally do not need ASCII interpretations, 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 and not otherwise required for correct protocol operation, giving the Unicode equivalent of the non-ASCII characters is not required, but it can improve the readability of the RFC. For example, for text that might just say "The value can be followed by a monetary symbol such as ¥ or ", it is likely more beneficial to the reader to instead say "The value can be followed by a monetary symbol such as ¥ (U+00A5) or (U+20AC)".

RFCs are often displayed on systems that use only black and white, 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 the "U+NNNN" syntax. For example, "A color display should be able to differentiate (U+1F534), 评 (U+1F7E2), and 鳩 (U+1F535)."

4. IANA Considerations

This document contains no IANA considerations.

5. Security Considerations

Valid Unicode that matches the expected text must be verified in order to preserve expected behavior and protocol information.

6. References

6.1. 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>>.

[UnicodeCurrent] The Unicode Consortium, "The Unicode Standard", 2023, <<http://www.unicode.org/versions/latest/>>.

6.2. 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>>.

[RFC7322] Flanagan, H. and S. Ginoza, "RFC Style Guide", RFC 7322, DOI 10.17487/RFC7322, September 2014, <<https://www.rfc-editor.org/rfc/rfc7322>>.

[UnicodeNorm] The Unicode Consortium, "Unicode Standard Annex", 2023, <<http://www.unicode.org/reports/tr15/>>.

Appendix A. Acknowledgements

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

The acknowledgements from [RFC7997] are to the members of the IAB il8n 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.

This current document was greatly helped by contributions from the RFC Series Working Group (RSWG), including from Brian Carpenter, Carsten Bormann, Eliot Lear, John Levine, and Martin Thomson.

Author's Address

Paul Hoffman
Email: paul.hoffman@icann.org