Internet Architecture Board (IAB)
Request for Comments: 7994
Category: Informational
ISSN: 2070-1721

H. Flanagan
RFC Editor
December 2016

Requirements for Plain-Text RFCs

Abstract

In 2013, after a great deal of community discussion, the decision was made to shift from the plain-text, ASCII-only canonical format for RFCs to XML as the canonical format with more human-readable formats rendered from that XML. The high-level requirements that informed this change were defined in RFC 6949, "RFC Series Format Requirements and Future Development". Plain text remains an important format for many in the IETF community, and it will be one of the publication formats rendered from the XML. This document outlines the rendering requirements for the plain-text RFC publication format. These requirements do not apply to plain-text RFCs published before the format transition.

Status of this Memo

This document is not an Internet Standards Track specification; it is published for informational purposes.

This document is a product of the Internet Architecture Board (IAB) and represents information that the IAB has deemed valuable to provide for permanent record. It represents the consensus of the Internet Architecture Board (IAB). Documents approved for publication by the IAB are not a candidate for any level of Internet Standard; see Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/rfc7994[1].

Copyright Notice

Copyright © 2016 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 (http://trustee.ietf.org/license-info[2]) 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.

[1] http://www.rfc-editor.org/info/rfc7994
[2] http://trustee.ietf.org/license-info

Table of Contents

1 Introduction **3**

2 Character Encoding **4**

3 Figures and Artwork **5**

4 General Page Format Layout **6**

4.1 Headers and Footers 6

4.2 Table of Contents 6

4.3 Line Width 6

4.4 Line Spacing 6

4.5 Hyphenation 6

5 Elements from the xml2rfc v3 Vocabulary **7**

6 Security Considerations **8**

7 References **9**

7.1 Normative References 9

7.2 Informative References 9

Author's Address **12**

1. Introduction

In 2013, after a great deal of community discussion, the decision was made to shift from the plain-text, ASCII-only canonical format for RFCs to XML as the canonical format [XML-ANNOUNCE]. The high-level requirements that informed this change were defined in [RFC6949], "RFC Series Format Requirements and Future Development". Plain text remains an important format for many in the IETF community, and it will be one of the publication formats rendered from the XML. This document outlines the rendering requirements for the plain-text RFC publication format. These requirements do not apply to plain-text RFCs published before the format transition.

The Unicode Consortium defines "plain text" as "Computer-encoded text that consists only of a sequence of code points from a given standard, with no other formatting or structural information. Plain#text interchange is commonly used between computer systems that do not share higher-level protocols." [UNICODE-GLOSSARY]. In other words, plain-text files cannot include embedded character formatting or style information. The actual character encoding, however, is not limited to any particular sequence of code points.

A plain-text output for RFCs will continue to be required for the foreseeable future. The process of converting xml2rfc version 2 (xml2rfc v2) into text documents is well understood [RFC7749]. We plan to rely on the practice to date to inform the requirements for converting xml2rfc version 3 (xml2rfc v3) to text [RFC7991]. This document calls out those requirements that are changed from v2 or otherwise deserve special attention, such as the requirements around the character encoding that may be used; changes in the page layout; and changes in how figures, artwork, and pagination may be handled. For more details on general style, see "RFC Style Guide" [RFC7322].

The following assumptions drive the changes in the plain-text output for RFCs:

- The existing tools used by the RFC Editor and many members of the author community to create the text file are complicated to change and support; manual manipulation is often required for the final output. In particular, handling page breaks and associated widows and orphans for paginated output is tricky [WIDOWS].
- Additional publication formats -- for example, PDF, HTML -- will be available that will offer features such as markup and pretty printing.
- There is an extensive tool chain in existence among the community to work with plain-text documents. Similar functionality may be possible with other publication formats, but the workflow that uses the existing tool chain should be supported as much as is considered practical.

Where practical, the original guidance for the structure of a plain#text RFC has been kept (e.g., with line lengths, with lines per page [INS2AUTH]). Other publication formats, such as HTML and PDF, will include additional features that will not be present in the plain text (e.g., paragraph numbering, typographical emphasis).

The details described in this document are expected to change based on experience gained in implementing the new publication toolsets. Revised documents will be published capturing those changes as the toolsets are completed. Other implementers must not expect those changes to remain backwards-compatible with the details described in this document.

2. Character Encoding

Plain-text files for RFCs will use the UTF-8 [RFC3629] character encoding. That said, the use of non-ASCII characters will be only allowed in a limited and controlled fashion.

Many elements within the xml2rfc v3 vocabulary have an attribute for the ASCII equivalent to a non-ASCII character string. The ASCII equivalent will be rendered within the plain text as per the guidance in "The Use of Non-ASCII Characters in RFCs" [RFC7997]; please view the PDF version of that document.

The plain-text file will include a Byte Order Mark (BOM) to provide text reader software with in-band information about the character encoding scheme used.

3. Figures and Artwork

Artwork, such as network diagrams or performance graphs, must be tagged by the XML <artwork> element (see Section 2.5 of "The 'xml2rfc' Version 3 Vocabulary" [RFC7991]. Where this artwork is comprised of an ASCII art diagram, it must be tagged as "type='ascii#art'". The plain-text format will only include ASCII art. If the canonical format includes figures or artwork other than ASCII art, then the plain-text output must include a pointer to the relevant figure in the HTML version of the RFC to allow readers to see the relevant artwork.

Authors who wish to include ASCII art for the plain-text file and SVG art for the other outputs may do so, but they should be aware of the potential for confusion to individuals reading the RFC with two unique diagrams describing the same content. If there is conflicting information between the publication formats, please review the XML and PDF files to resolve the conflict.

4. General Page Format Layout

One plain-text output will be created during the publication process with basic pagination that includes a form feed instruction every 58 lines at most, including blank lines. Instructions or a script will be made available by and for the community to strip out pagination as per individual preference.

4.1. Headers and Footers

The front matter on the front page (such as the RFC number and category) and the back matter on the last page (the authors' full names and contact information) will continue with the structure described in RFC 7841 [RFC7841], "RFC Streams, Headers, and Boilerplates". Running headers and footers will no longer be added.

4.2. Table of Contents

In order to retain similar content wherever possible between the various publication formats, the table of contents will list section and subsection numbers and titles but will not include page numbers.

4.3. Line Width

Each line must be limited to 72 characters followed by the character sequence that denotes an end-of-line (EOL). The EOL sequence used by the RFC Editor will be the two-character sequence CR LF (Carriage Return followed by Line Feed). This limit includes any left-side indentation.

Note that the EOL used by the RFC Editor may change with different transports and as displayed in different display software.

4.4. Line Spacing

Use single-spaced text within a paragraph, and one blank line between paragraphs.

4.5. Hyphenation

Hyphenated words (e.g., "Internet-Draft") should not be split across successive lines.

5. Elements from the xml2rfc v3 Vocabulary

The plain-text formatter uses the relevant tags from the xml2rfc v3 source file to build a document conforming to the layout and structure described by the full RFC Style Guide [RFC7322] (including the updates in the web portion of the Style Guide) [STYLEWEB].

6. Security Considerations

The requirements of the plain-text format involve no significant security considerations. As part of the larger format project, however, unintended changes to the text as a result of the transformation from the base XML file could in turn corrupt a standard, practice, or critical piece of information about a protocol.

7. References

7.1. Normative References

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

[RFC6949] Flanagan, H. and N. Brownlee, "RFC Series Format Requirements and Future Development", RFC 6949, DOI 10.17487/RFC6949, May 2013, <http://www.rfc-editor.org/info/rfc6949>.

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

[RFC7749] Reschke, J., "The "xml2rfc" Version 2 Vocabulary", RFC 7749, DOI 10.17487/RFC7749, February 2016, <http://www.rfc-editor.org/info/rfc7749>.

[RFC7841] Halpern, J., Ed., Daigle, L., Ed., and O. Kolkman, Ed., "RFC Streams, Headers, and Boilerplates", RFC 7841, DOI 10.17487/RFC7841, May 2016, <http://www.rfc-editor.org/info/rfc7841>.

[RFC7991] Hoffman, P., "The "xml2rfc" Version 3 Vocabulary", RFC 7991, DOI 10.17487/RFC7991, December 2016, <http://www.rfc-editor.org/info/rfc7991>.

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

7.2. Informative References

[INS2AUTH] RFC Editor, "Instructions to Request for Comments (RFC) Authors", August 2004, <https://www.rfc-editor.org/materials/instructions2authors.txt>.

[STYLEWEB] RFC Editor, "Web Portion of the Style Guide", February 2016, <http://www.rfc-editor.org/styleguide/part2/>.

[UNICODE-GLOSSARY] The Unicode Consortium, "Glossary of Unicode Terms", September 2016, <http://www.unicode.org/glossary/>.

[WIDOWS] Wikipedia, "Widows and orphans", September 2016, <https://en.wikipedia.org/w/index.php?title=Widows_and_orphans&oldid=738356204>.

[XML-ANNOUNCE] Flanagan, H., "Subject: Direction of the RFC Format Development effort", May 2013, <http://www.rfc-editor.org/pipermail/rfc-interest/2013-May/005584.html>.

IAB Members at the Time of Approval

The IAB members at the time this memo was approved were (in alphabetical order):

Jari Arkko
Ralph Droms
Ted Hardie
Joe Hildebrand
Russ Housley
Lee Howard
Erik Nordmark
Robert Sparks
Andrew Sullivan
Dave Thaler
Martin Thomson
Brian Trammell
Suzanne Woolf

Acknowledgements

This document owes a great deal of thanks to the efforts of the RFC Format Design Team: Nevil Brownlee, Tony Hansen, Joe Hildebrand, Paul Hoffman, Ted Lemon, Julian Reschke, Adam Roach, Alice Russo, Robert Sparks, and David Thaler.

Author's Address

Heather Flanagan
RFC Editor
EMail: rse@rfc-editor.org
URI: http://orcid.org/0000-0002-2647-2220