Internet Architecture Board
Internet-Draft
Intended status: Informational
Expires: August 13, 2016

H. Flanagan
RFC Editor

February 10, 2016

Requirements for Plain-Text RFCs
draft-iab-rfc-plaintext-02

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 RFC6949, "RFC Series Format Requirements and Future Development." Plain text remains an important format for many in the IETF community, and will be one of the publication formats rendered from the XML. This draft documents the rendering requirements for the plain-text RFC publication format. These requirements do not apply to plain-text RFCs published before the format transition.

Editorial Note (To be removed by the RFC Editor)

Discussion of this draft takes place on the rfc-interest mailing list (rfc-interest@rfc-editor.org), which has its home page at https://www.rfc-editor.org/mailman/listinfo/rfc-interest.

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 http://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 August 13, 2016.

Copyright Notice

Copyright (c) 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) 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. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction 3
2. Character Encoding 4
3. Figures and Artwork 4
4. General Page Format Layout 5
4.1. Headers and Footers 5
4.2. Table of Contents 5
4.3. Line Width 5
4.4. Line Spacing 5
4.5. Hyphenation 6
5. Elements from the xml2rfc v3 vocabulary 6
6. Acknowledgements 6
7. IANA Considerations 6
8. Security Considerations 6
9. Change Log for the Draft 6
9.1. draft-iab-rfc-plantext-01 to -02 6
9.2. draft-iab-rfc-plantext-00 to -01 6
9.3. draft-flanagan-plaintext-09 to draft-iab-rfc-plaintext-00 7
9.4. -08 to -09 7
9.5. -07 to -08 7
9.6. -06 to -07 7
9.7. -05 to -06 7
9.8. -04 to -05 7
9.9. -03 to -04 7
9.10. -02 to -03 8
9.11. -01 to -02 8
10. References 8
10.1. Normative References 8
10.2. Informative References 9
Author's Address 9

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 will be one of the publication formats rendered from the XML. This draft documents 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 [I-D.iab-xml2rfcv2]. We plan to rely on the practice to date to inform the requirements for converting XML2RFC version 3 (xml2rfc v3) to text [I-D.iab-xml2rfc]. This document calls out those requirements that are changed from v2 or otherwise deserve special attention, such as the requirements around character encoding may be used, changes in the page layout, changes in handling figures, artwork, and pagination. For more details on general style, see "The 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, pretty printing, etc.

- 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, such as with line lengths, lines per page, etc. [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 RFC production center's toolset. Revised documents will be published capturing those changes as the toolset is completed. Other implementers must not expect those changes to remain backwards-compatible with the details described 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" [I-D.iab-rfc-nonascii]. Please view the PDF version of that draft.

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 is defined as anything marked by the XML >artwork< element (see Section 2.5 of "The 'XML2RFC' version 3 Vocabulary" [I-D.iab-xml2rfc]. Only the 'type=ascii-art' will be rendered within the plain-text format. This marks figures drawn with ASCII characters.

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 author's full names and contact information) will continue with the structure described in RFC 5741 [RFC5741], "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 xml2rfcv3 source file to build a document conforming to the layout and structure described by the full RFC Style Guide (including the updates in the web portion of the Style Guide). [STYLEWEB] In particular, the formatter will use the "submissionType", "seriesInfo", "author", "address", "title", "reference", "referencegroup", and "references" to build the front and back matter of the document.

6. Acknowledgements

This draft 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.

7. IANA Considerations

This memo includes no requests to IANA.

8. Security Considerations

The requirements of the plaintext 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.

9. Change Log for the Draft

9.1. draft-iab-rfc-plantext-01 to -02

nits fixed

9.2. draft-iab-rfc-plantext-00 to -01

Introduction: removed sentence restricting this format to RFCs only; clarified that plaintext will be based on existing practice (except where otherwise called out)

Elements from the xml2rfc v3 vocabulary: clarified what xml2rfcv3 tags will render the front and back matter of a document.

9.3. draft-flanagan-plaintext-09 to draft-iab-rfc-plaintext-00

Figures and Artwork, Character Encoding: included additional detail regarding how these items will be flagged within the XML.

9.4. -08 to -09

Security Considerations: added text

9.5. -07 to -08

Change log: forgot to update the change log for the -06 to -07 changes.

9.6. -06 to -07

Introduction: updated to state that this document does not require backwards compatibility.

9.7. -05 to -06

Abstract: Changed "cut over" to "transition"

Elements from xml2rfc v3: emphasized that doc structure is guided by the RFC Style Guide

9.8. -04 to -05

Abstract and Introduction: Revised for better readability; clarified the definition and implications of the term "plain-text"

General Page Format Layout: Added explicit EOL detail and added some clarification regarding pagination

Elements from the xml2rfc v3 vocabulary: section added

9.9. -03 to -04

Change Log for the Draft: forgot to complete the change log between the various revisions of the draft

9.10. -02 to -03

Abstract: expanded

Introduction: adjusted language of assumptions

Figures and Artwork: adjusted to indicate where to go in case information for the images conflicts between different formats

General Page Layout: switched back to producing one basic paginated format, with an expectation of instructions and/or a script to create local, unpaginated copies for individual use.

9.11. -01 to -02

Introduction: added pointer to original page layout information

Character encoding: clarified language around encoding and use of BOMs

General Page Format Layout: removed increased line width requirement; added sections on Line Width, Line Spacing, and Hyphenation (pulled from 2223-bis

10. References

10.1. Normative References

[I-D.iab-rfc-nonascii]
Flanagan, H., "The Use of Non-ASCII Characters in RFCs", draft-iab-rfc-nonascii-00 (work in progress), January 2016.

[I-D.iab-xml2rfc]
Hoffman, P., "The "xml2rfc" version 3 Vocabulary", draft-iab-xml2rfc-02 (work in progress), January 2016.

[I-D.iab-xml2rfcv2]
Reschke, J., "The 'XML2RFC' version 2 Vocabulary", draft-iab-xml2rfcv2-02 (work in progress), September 2015.

[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>.

[RFC5741] Daigle, L., Ed., Kolkman, O., Ed., and IAB, "RFC Streams, Headers, and Boilerplates", RFC 5741, DOI 10.17487/RFC5741, December 2009, <http://www.rfc-editor.org/info/rfc5741>.

[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>.

10.2. Informative References

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

[STYLEWEB] RFC Editor, "Web Portion of the Style Guide", May 2015, <http://www.rfc-editor.org/styleguide/part2/>.

[UNICODE-GLOSSARY] The Unicode Consortium, "Glossary of Unicode Terms", 2014, <http://www.unicode.org/glossary/>.

[WIDOWS] Wikipedia, "Widows and orphans", October 2014, <http://en.wikipedia.org/wiki/Widows_and_orphans>.

[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 >.

Author's Address

Heather Flanagan
RFC Editor

Email: rse@rfc-editor.org
URI: http://orcid.org/0000-0002-2647-2220