                    Requirements for Plain-Text RFCs
                      draft-flanagan-plaintext-05

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. The high-level requirements
   that informed this change were defined in RFC6949, "RFC Series Format
   Requirements and Future Development." Plain text is still an
   important format for many in the IETF community, and will still 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 cut over.

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 July 16, 2015.

Copyright Notice

Table of Contents

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
   is still an important format for many in the IETF community, and will

still 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 for a plain-text
file, however, is not limited to ASCII or any other character
encoding system.

A plain-text output for RFCs will continue to be required for the
foreseeable future. The details of what that means for RFCs in terms
of which character encoding may be used, what the page layout will
look like, how to handle figures and artwork, and how pagination will
be handled, are documented in this memo. 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:

o 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].

o Additional publication formats--for example: PDF, HTML-- will be
   available that will offer features such as markup, pretty
   printing, etc.

o 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]

Note that this document provides guidance for plain-text RFCs;
Internet-Drafts are out of scope.

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. The details regarding
the guidance for how to include non-ASCII characters is under
development and documented in "The Use of Non-ASCII Characters in
RFCs" [I-D.flanagan-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

Authors may continue to include 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 does not attempt to render all metadata from the XML.
   URIs that occur in targets may be placed into parenthetical
   expressions or reference entries. [I-D.hoffman-xml2rfc].

   Since plain-text cannot include any rich-text-style formatting, such
   as bold or italic fonts, tags such as <em> and <strong> will not be
   rendered in any way in the plain-text draft.

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

    TBD.

9. Change Log for the Draft

9.1. -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

9.2. -03 to -04

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

9.3. -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.

9.4. -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.flanagan-nonascii]
          Flanagan, H., "The Use of Non-ASCII Characters in RFCs",
          draft-flanagan-nonascii-03 (work in progress), July 2014.

[I-D.hoffman-xml2rfc]
          Hoffman, P., "The 'XML2RFC' version 3 Vocabulary", draft-
          hoffman-xml2rfc-15 (work in progress), January 2015.

[RFC3629] Yergeau, F., "UTF-8, a transformation format of ISO
          10646", STD 63, RFC 3629, November 2003.

[RFC5741] Daigle, L., Kolkman, O., and IAB, "RFC Streams, Headers,
          and Boilerplates", RFC 5741, December 2009.

[RFC6949] Flanagan, H. and N. Brownlee, "RFC Series Format
          Requirements and Future Development", RFC 6949, May 2013.

[RFC7322] Flanagan, H. and S. Ginoza, "RFC Style Guide", RFC 7322,
          September 2014.

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>.

[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", message to the rfc-interest@rfc-
          editor.org mailing list", 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