

INTERNET-DRAFT
Intended Status: Informational

H. Flanagan
RFC Series Editor
N. Brownlee
Independent Submissions Editor
October 18, 2012

Expires: April 21, 2013

RFC Series Format Development
draft-rfc-format-flanagan-01

Abstract

This document describes the current requirements for the format of RFCs and recent requests for enhancements as understood from community discussion and various proposals for new formats including HTML, XML, PDF and EPUB. Terms are defined to help clarify exactly which stages of document production are under discussion for format changes. The requirements described in this document will determine what changes will be made to RFC format.

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at
<http://www.ietf.org/lid-abstracts.html>

The list of Internet-Draft Shadow Directories can be accessed at
<http://www.ietf.org/shadow.html>

Copyright and License Notice

Copyright (c) 2012 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
1.1	Terminology	3
2.	History and Goals	4
2.1.	Issues driving change	5
2.2.	Further considerations	9
2.3.	RFC Editor goals	10
3.	Format requirements	10
4.	Security Considerations	12
5.	IANA Considerations	12
6.	References	12
6.1.	Normative References	12
6.2.	Informative References	12
	Acknowledgements	12
	Authors' Addresses	12

1 Introduction

Over 40 years ago, the RFC Series began in an environment that included handwritten RFCs, typewritten RFCs, RFCs produced on mainframes, and more. This resulted in an understanding that however they were published, a common format that could be read and revised long in the future was required. US-ASCII was chosen as that format, and since that time that format has proved to be persistent and reliable across a large variety of devices, operating systems, and editing tools. That stability has been a continuing strength of the Series. However, as new technology such as small devices and advances in display technology come in to common usage, there is a growing desire to see the format of the RFC Series adapt to take advantage of these different ways to communicate information.

Since the earliest days of the series, authors and readers have suggested enhancements to the format. However, no suggestion developed clear consensus in the Internet technical community. As always, some individuals see no need for change and some press for specific enhancements.

This document takes a look at the current requirements for RFCs as described in RFC 2223 [RFC2223] and more recently in 2223bis [2223bis], and recent requests for enhancements as understood from community discussion and various proposals for new formats including HTML, XML, PDF and EPUB. Terms are defined to help clarify exactly which stages of document production are under discussion for format changes.

1.1 Terminology

ASCII = 7-bit ASCII

Submission format = the format submitted for editorial revision and publication to the RFC Editor.

- * might not be the same as the canonical formats (though it would make the workflow somewhat simpler for the RFC Editor if it were);
- * will be converted to another format for further processing and publication if necessary
- * Currently: .txt (required), XML (optional), NROFF (optional)

Revisable format = the format that will provide the information for conversion into an Publication format; it is used or created by the

RFC Editor

- * Currently: XML (optional), NROFF (required)

Publication format = display and distribution format as it may be read or printed after publication process has completed.

- * Currently published by the RFC Editor: .txt, PDF, PDF that contains figures (rare)
- * Currently made available by other sites: HTML, PDF, others

Canonical format = the authorized, recognized, accepted, and archived version of the document.

Metadata = Information associated with a document, such as defining its structure, presentation, topic or author.

- * Currently: .txt

Metadata = Information associated with a document, such as defining its structure, presentation, topic or author

2. History and Goals

Current RFC format rules as defined in [RFC2223] and clarified in 2223bis.

- * The character codes are ASCII.
- * Each page must be limited to 58 lines followed by a form feed on a line by itself.
- * Each line must be limited to 72 characters followed by carriage return and line feed.
- * No overstriking (or underlining) is allowed.
- * These "height" and "width" constraints include any headers, footers, page numbers, or left side indenting.
- * Do not fill the text with extra spaces to provide a straight right margin.
- * Do not do hyphenation of words at the right margin.
- * Do not use footnotes. If such notes are necessary, put them at the end of a section, or at the end of the document.

- * Use single spaced text within a paragraph, and one blank line between paragraphs.
- * Note that the number of pages in a document and the page numbers on which various sections fall will likely change with reformatting. Thus cross references in the text by section number usually are easier to keep consistent than cross references by page number.
- * RFCs in plain ASCII-text may be submitted to the RFC Editor in e-mail messages (or as online files) in either the finished publication format or in nroff. If you plan to submit a document in nroff please consult the RFC Editor first.

Precedent for multiple output formats is described in RFC 2223 and has been used for a few RFCs:

Note that since the ASCII text version of the RFC is the primary version, the PostScript version must match the text version. The RFC Editor must decide if the PostScript version is "the same as" the ASCII version before the PostScript version can be published.

Neither RFC 2223 or 2223bis use the term metadata, though we currently refer to components of the text such as the Stream, Status (e.g. Updates, Obsoletes), Category and ISSN as metadata.

2.1. Issues driving change

While some authors and readers of RFCs find the strict limits of character encoding, line limits, and so on to be acceptable, others find those limitations a significant obstacle to their desire to communicate information via an RFC. With a broader base of authors currently producing RFCs and a wider range of presentation devices, the issues driving change may represent critical deficiencies in the current format.

While the specific points of concern vary, the main issues are:

- * Line art, also known as ASCII art
- * Character encoding
- * Pagination
- * Reflowable text

- * Metadata

Each area of concern has people in favor of change and people opposed to it, all with reasonable concerns and requirements. Below is a summary of the arguments for and against each major issue. The potential requirements derived from these discussions are listed later in this document.

Line art, aka ASCII art

Arguments in favor of keeping the current requirement for all diagrams, equations, tables, and charts include:

- * Dependence on advanced diagrams (or any diagrams) causes accessibility issues
- * Requiring ASCII art results in people often relying more on clear written descriptions rather than just the diagram itself.
- * By their nature ASCII forces design of diagrams that are simple and discrete.

Arguments in favor of replacing ASCII art with more complex diagrams include:

- * State diagrams with multiple arrows in different directions and labels on the lines will be more understandable.
- * Protocol flow diagrams where each step needs multiple lines of description will be clearer.
- * Scenario descriptions that involve three or more parties with communication flows between them will be clearer.
- * Given the difficulties in expressing complex equations with common mathematical notation, allowing graphic art would allow equations to be displayed properly.
- * Complex art could allow for color to be introduced in to the diagrams.

Two suggestions have been been proposed regarding how graphics should be included: one that would have graphic art referenced as a separate document the publication format and one that would allow embedded graphics in the publication format.

Character encoding For most of the history of the RFC Series, the character encoding for RFCs has been ASCII. Below are arguments for keeping ASCII as well as arguments for moving to UTF-8.

Arguments for retaining the ASCII-only requirement

- * Most easily searched and displayed across a variety of platforms.
- * In extreme cases of having to retype/scan hard copies of documents (it has been required in the past) ASCII is significantly easier to work with for rescanning and retaining all of the original information. There can be no loss of descriptive metadata such as keywords or content tags.
- * If we expand beyond ASCII, it will be difficult to know where to draw the line on what characters are and are not allowed. There will be issues with dependencies on local file systems and processors being configured to recognize any other character set.
- * The IETF works in ASCII (and English). The Internet research, design and development communities function almost entirely in English. That strongly suggests that an ASCII document can be read by everyone in the communities and audiences of interest.

Arguments for expanding to allow UTF-8:

- * In discussions of internationalization, actually being able to illustrate the issue is rather helpful, and you can't illustrate a Unicode code point with "U+nnnn".
- * Want the ability to denote protocol examples using the character sets those examples support.
- * Will allow better support for international character sets, in particular allowing authors to spell their names in their native character sets.
- * Certain special characters in equations or quoted from other texts could be allowed.
- * Citations of web pages using more international characters are possible.

Arguments for mixed ASCII and UTF-8

- * In order to keep documents as searchable as possible, ASCII-only should be required for the main text of the document and some broader UTF-8 character set allowed under clearly prescribed circumstances (e.g. author names and references).

Pagination

Arguments for continuing the use of discrete pages within RFCs:

- * Ease of reference and clear printing; referring to section numbers is too coarse a method

Arguments for removing the pagination requirement:

- * Removing pagination will allow for a smoother reading experience on a wider variety of devices, platforms, and browsers

Reflowable text

Arguments against allowing for reflowable text:

- * Reflowable text may impact the usability of graphics and tables within a document.

Arguments for allowing reflowable text

- * RFCs are more readable on a wider variety of devices and platforms, including mobile devices and a wide variety of screen layouts.

Metadata and tagging

While metadata requirements are not part of RFC 2223, there is a request that descriptive metadata tags be added as part of a revision of the RFC format. These tags would allow for enhanced content by embedding information like links, tags, or quick translations and could help control the look and feel of the publication format. While the lack of metadata in the current RFCs does not impact an RFCs accessibility or readability, if other requirements are accepted, such as allowing UTF-8 in any part of an RFC, then having the ability to use metadata to provide an ASCII "translation" of the UTF-8 letters is also a requirement.

Arguments for allowing metadata in the final output format:

- * Allowing metadata in the final output format allows readers to

potentially get more detail out of a document. For example, if non-ASCII characters are allowed in the Author and Reference sections, Metadata must include translations of that information.

Arguments against metadata in the final output format:

- * Metadata adds additional overhead to the overall process of creating RFCs and may complicate future usability.

2.2. Further considerations

Some of the discussion beyond the issues described above went into potential solutions. Those solutions and the debate around them added a few more points to the potential requirements for a change in RFC Format. In particular, discussing whether a change in format should also include the creation and ongoing support of specific RFC authoring and/or rendering tools and whether the Publication format should be in a format that must go through a rendering agent to be readable.

Creation and use of RFC-specific tools

Arguments against community-supported RFC-specific tools:

- * We cannot be so unique in our needs that we can't use commercial tools.
- * Ongoing support for these tools adds a greater level of instability to the ongoing availability of the RFC Series through the decades.
- * The community that would support these tools cannot be relied on to be as stable and persistent as the Series itself.

Arguments in support of community-supported RFC-specific tools:

- * Given the community that would be creating and supporting these tools, there would be greater control and flexibility over the tools and how they implement the RFC format requirements.
- * Community supported tools currently exist and are in extensive use within the community, so it would be most efficient to build on that base.

Markup Language

Arguments in support of a markup language as the revisable

format:

- * Having a markup language such as XML or HTML allows for greater flexibility in creating a variety of display formats, with a greater likelihood of similarity between them.

Arguments against a markup language as the revisable format:

- * Having the publication format be in code instead of in a simple text-formatting structure ties us in to specific tools and/or tool support going forward.

2.3. RFC Editor goals

Today, each RFC has an nroff file created prior to publication. For RFCs revised using an XML file, this file is created by converting XML to nroff at the final step. As more documents are submitted with an XML file (so far in 2012, 66% of approved I-Ds were submitted with an XML file), this conversion is problematic in terms of time spent and data lost from XML. Making this process more efficient is strongly desired by the RFC Editor.

3. Format requirements

Understanding the major pain points and balancing them with the goal of long-term viability of the documents and the requirements of a broad community base brings us to a review of what must be kept of the original requirements, what new requirements must be added, and what requirements can be retired.

There are several components of the original format requirements that must be retained to ensure the ongoing continuity, reliability and readability of the Series:

- * While several publication formats must be allowed, the publication formats must include support for plain-text printing.
- * RFCs must not change, regardless of format, once published
- * The Boilerplate and overall structure of the RFC must be in accordance with current RFC and Style Guide requirements (see [RFC5741])

Issues such as overstriking, page justification, hyphenation, and spacing will be defined in the RFC Style Guide. [link required?]

In addition to those ongoing requirements, the following must be added:

- * There must be support for Accessibility (this has significant implications) [WCAG20]
- * The Submission and Publication formats need to permit extensible encoding, for the addition of labeled metadata.
- * UTF-8/Unicode in the header and references and associated ASCII translations recorded in the metadata (that translation will be used in the ASCII text version of the RFC)
- * If descriptive metadata within a document is allowed, defining the accepted set of metadata tags will become part of the RFC Style Guide.
- * Graphics may include ASCII art and SVG line art. Color will not be accepted; RFCs must correctly display in monochrome.
- * RFC must be readable self-contained (i.e. must not contain normative external links, figures, etc.).
- * Fonts are restricted to fixed-width fonts.

The goals of the RFC Editor in considering how the formats for Submission and Publication should change include:

- * The final conversion of all submitted documents to nroff should be replaced by using an accepted revisable format throughout the process.
- * Support for a limited set of submission and publication formats as opposed to supporting all formats possible.
- * Some proposed solutions allow for multiple files to be submitted for later consolidation. The RFC Editor would prefer to work with the minimal number of files required for each submission (not a tar ball of several discrete components).
- * Tools should support the submission and format error checking on the part of the authors of Submission Format documents

Some of the original requirements may be removed from consideration:

- * Pagination (" Each page must be limited to 58 lines followed by a form feed on a line by itself.")

- * Maximum line length ("Each line must be limited to 72 characters followed by carriage return and line feed.")
- * Limitation to 100% ASCII text ("The character codes are ASCII.")

4. Security Considerations

<Security considerations text>

5. IANA Considerations

<IANA considerations text>

6. References

6.1. Normative References

[WCAG20] W3C, "Web Content Accessibility Guidelines (WCAG) 2.0", December 11 2008, <http://www.w3.org/TR/WCAG20/>

6.2. Informative References

[RFC2223] Postel, J. and J. Reynolds, "Instructions to RFC Authors", RFC 2223, October 1997.

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

[2223bis] Reynolds, J. Bradon, R., "Instructions to Request for Comments (RFC) Authors", Work In Progress, August 2004.

Acknowledgements

The authors received a great deal of helpful input from the community in pulling together these requirements and wish to particularly acknowledge the help of Joe Hildebrand, Paul Hoffman and John Klensin, who each published an I-D on the topic of potential format options before the IETF 84 BOF.

Authors' Addresses

Heather Flanagan
RFC Series Editor

EMail: rse@rfc-editor.org

Nevil Brownlee
Independent Submissions Editor

Email rfc-ise@rfc-editor.org