Internet Architecture Board (IAB)
Request for Comments: 6949
Updates: 2223
Category: Informational
ISSN: 2070-1721

H. Flanagan
RFC Series Editor
N. Brownlee
Independent Submissions Editor

May 2013

RFC Series Format Requirements and Future Development

Abstract

This document describes the current requirements and requests for enhancements for the format of the canonical version of RFCs. 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. This document updates RFC 2223.

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 5741.

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/rfc6949.

Table of Contents

1 Introduction

Over 40 years ago, the RFC Series began as a collection of memos in an environment that included handwritten RFCs, typewritten RFCs, RFCs produced on mainframes with complicated layout tools, and more. As the tools changed and some of the source formats became unreadable, the core individuals behind the Series realized that a common format that could be read, revised, and archived long in the future was required. US-ASCII was chosen for the encoding of characters, and after a period of variability, a well-defined presentation format was settled upon. 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, comes into 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 format stabilized, 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, while others press strongly 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]. Section 2 reviews recent requests for enhancements as understood from community discussion and various proposals for new formats including HTML, XML, PDF, and EPUB. The actual requirements are then captured in Section 3. The focus of this document is on the Canonical format of RFCs, but some mention of other phases in the RFC publication process and the document formats associated with these phases is also included. Terms are defined to help clarify exactly which stages of document production are under discussion for format changes.

1.1 Terminology

ASCII: Coded Character Set -- 7-bit American Standard Code for Information Interchange, ANSI X3.4-1986 [ASCII]

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

* Currently: formatted plain text

Metadata: information associated with a document so as to provide, for example, definitions of its structure, or of elements within the document such as its topic or author

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

* Currently published by the RFC Editor: formatted plain text, PDF of the formatted plain text, PDF that contains figures (rare)
* Currently made available by other sites: HTML, PDF, others

Reflowable text: text that automatically wraps to the next line in a document as the user moves the margins of the text, either by resizing the window or changing the font size

Revisable format: the format that will provide the information for conversion into a Publication format; it is used or created by the RFC Editor (see Section 2.3 for an explanation of current practice)

* Currently: XML (optional), nroff (required)

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

* Currently: formatted plain text (required), XML (optional), nroff (optional)

2. History and Goals

Below are the 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.

The precedent for additional formats, specifically PostScript, is described in RFC 2223 and has been used for a small number of 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 nor 2223bis uses the term 'metadata', though the RFC Editor currently refers 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 report that they find the strict limits of character encoding, line limits, and so on to be acceptable, others claim to find those limitations a significant obstacle to their desire to communicate and read the information via an RFC. With a broader community of authors currently producing RFCs and a wider range of presentation devices in use, the issues being reported by the community indicate limitations of the current Canonical format that must be reviewed and potentially addressed in the Canonical RFC format.

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

* 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. These points are not part of the list of requirements; they are the inputs that informed the requirements discussed in Section 3 of this document.

2.1.1. ASCII Art

Arguments in favor of limiting all diagrams, equations, tables, and charts to ASCII art depictions only 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.

* Use of the ASCII character set forces design of diagrams that are simple and concise.

Arguments in favor of allowing the use of more complex diagrams in place of the current use of ASCII art include:

* State diagrams with multiple arrows in different directions and labels on the lines will be more understandable.

* Protocol flow diagrams in which 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 grayscale or color to be introduced into the diagrams.

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

2.1.2. 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 expanding to UTF-8.

Arguments for retaining the ASCII-only requirement:

* It is the most easy to search and display across a variety of platforms.

* In extreme cases of having to retype or 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 which 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 properly rendered and 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".

* It will provide the ability to denote protocol examples using the character sets those examples support.

* It 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 strictly prescribed UTF-8 use:

* 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).

2.1.3. Pagination

Arguments for continuing the use of discrete pages within RFCs:

* Ease of reference and 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.

* Removing pagination results in people often using subsections rather than page number for reference purposes, forcing what would otherwise be long sections to be broken into subsections. This would encourage documents that are better organized and simpler.

2.1.4. Reflowable Text

Arguments against requiring text to be reflowable:

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

Arguments for requiring text to be reflowable:

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

2.1.5. 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 Canonical RFC format. These tags would allow for enhanced content by embedding information such as 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 RFC's accessibility or readability, several individuals have indicated that allowing metadata within the RFCs would make their reading of the documents more efficient.

Arguments for allowing metadata in the Canonical and Publication formats:

* Metadata potentially allows readers to get more detail out of a document. For example, if non-ASCII characters are allowed in the Author's Address and Reference sections, metadata must include translations of that information.

Arguments against metadata in the final Canonical and Publication formats:

* Metadata adds additional overhead to the overall process of creating RFCs and may complicate future usability as a result of requiring backward compatibility for metadata tags.

2.2. Further Considerations

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

2.2.1. Creation and Use of RFC-Specific Tools

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.

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.

2.2.2. 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 Publication formats, with a greater likelihood of similarity between them.

Arguments against a markup language as the Revisable format:

* Having the Revisable format be in a markup language 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

Currently, each RFC has an nroff file created prior to publication. For RFCs revised using an XML file, the nroff file is created by converting XML to nroff at the final step. As more documents are submitted with an XML file (of the RFCs published in 2012, approximately 65% were submitted with an XML file), this conversion is problematic in terms of time spent and data lost from XML. Making the publication process for the RFC Editor more efficient is strongly desired.

3. Format Requirements

Understanding the major pain points and balancing them with the expectation of long-term viability of the documents brings us to a review of what must be kept of the original requirements, what new requirements may be added, and what requirements may be retired. Detailed rules regarding how these changes will be implemented will be documented in a future RFC.

3.1. Original Requirements to Be Retained

Several components of the original format requirements must be retained to ensure the ongoing continuity, reliability, and readability of the Series:

1. The content of an RFC must not change, regardless of format, once published.

2. The Canonical format must be persistent and reliable across a large variety of devices, operating systems, and editing tools for the indefinite future. This means the format must be both readable and editable across commonly used devices, operating systems, and platforms for the foreseeable future.

3. While several Publication formats must be allowed, in order to continue support for the most basic reading and search tools and to provide continuity for the Series, at least one Publication format must be plain text.

4. 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 [Style].

3.2. Requirements to Be Added

In addition to those continuing requirements, discussions with various members of the wider Internet community have yielded the following general requirements for the RFC Series.

5. The documents must be made accessible to people with visual disabilities through such means as including alternative text for images and limiting the use of color. See the W3C's accessibility documents [WCAG20] and the United Nations "Convention on the Rights of Persons with Disabilities" [UN2006] for guidance. Appropriate authoring tools are highly desirable but focus on the creation of Internet-Drafts, a topic outside the scope of the RFC Editor.

6. The official language of the RFC Series is English. ASCII is required for all text that must be read to understand or implement the technology described in the RFC. Use of non-ASCII characters, expressed in a standard Unicode Encoding Form (such as UTF-8), must receive explicit approval from the document stream manager and will be allowed after the rules for the common use cases are defined in the Style Guide.

7. The Submission and Publication formats need to permit extending the set of metadata tags, for the addition of labeled metadata. A predefined set of metadata tags must be created to make use of metadata tags consistent for the life of the Series.

8. Graphics may include ASCII art and a more complex form to be defined, such as SVG line art [SVG]. Color and grayscale will not be accepted. RFCs must correctly display in monochromatic black-and-white to allow for monochrome displays, black-and-white printing, and support for visual disabilities.

9. The Canonical format must be renderable into self-contained Publication formats in order to be easily downloaded and read offline.

10. Fixed-width fonts and non-reflowable text are required for ASCII-art sections, source code examples, and other places where strict alignment is required.

11. At least one Publication format must support readable print to standard paper sizes.

12. The Canonical format should be structured to enable easy program identification and parsing of code or specifications, such as MIB modules and ABNF.

The requirements of the RFC Editor regarding RFC format and the publication process include:

13. The final conversion of all submitted documents to nroff should be replaced by using an accepted Revisable format throughout the process.

14. In order to maintain an efficient publication process, the RFC Editor must work with the minimal number of files required for each submission (not a tar ball of several discrete components).

15. In order to maintain the focus of the RFC Editor on editing for clarity and consistency rather than document layout details, the number of Publication formats produced by the RFC editor must be limited.

16. Tools must support error checking against document layout issues as well as other format details (diagrams, line breaks, variable- and fixed-width fonts).

3.3. Requirements to Be Retired

Some of the original requirements will be removed from consideration, but detailed rules regarding how these changes will be implemented will be documented in a future RFC.

* 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

This document sets out requirements for RFCs in their various formats; it does not concern interactions between Internet hosts. Therefore, it does not have any specific security considerations.

5. 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.

[ASCII] American National Standard for Information Systems - Coded Character Sets - 7-Bit American National Standard Code for Information Interchange (7-Bit ASCII), ANSI X3.4-1986, American National Standards Institute, Inc., March 26, 1986.

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

[Style] Flanagan, H. and S. Ginoza, "RFC Style Guide", Work in Progress, October 2012.

[SVG] Dahlstrom, E., Dengler, P., Grasso, A., Lilley, C., McCormack, C., Schepers, D., and J. Watt, "Scalable Vector Graphics (SVG) 1.1 (Second Edition)", W3C Recommendation, 16 August 2011, <http://www.w3.org/TR/SVG11/>.

[WCAG20] Caldwell, B., Cooper, M., Reid, L., and G. Vanderheiden, "Web Content Accessibility Guidelines (WCAG) 2.0", 11 December 2008, <http://www.w3.org/TR/WCAG20/>.

[UN2006] United Nations, "Convention on the Rights of Persons with Disabilities", December 2006.

6. 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 Internet-Draft 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