

Proposal for Canonical and Other Formats for RFCs
draft-hoffman-rfcformat-canon-others-00

Abstract

This document proposes a modernized text-based canonical format for RFCs that explicitly allows external art as a normative part of the RFC. If the RFC Editor chooses this format, they will also publish non-canonical versions of RFCs in order to accomodate the largest target audience of readers. Having a simple, stable canonical format and a varying number of non-canonical formats that can change over time allows the RFC Editor to add useful formats, particularly in HTML, that can keep up with the needs of all RFC readers.

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 November 13, 2012.

Copyright 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.

1. Introduction

A clear result of the decades-long discussion about the format of published RFCs is that different RFC readers have different needs and desires. No single format will be sufficient, or even useful, to all people who read RFCs. Another clear result is that the format described in [[RFC2223](#)] and its follow-ons is no longer the best format for publishing protocols, process descriptions, research findings, and the many other types of documents that are part of the modern RFC series.

This document proposes to deal with these issues in a way that meets the needs and desires of the widespread RFC-reading community. For every RFC, the RFC Editor will publish all of the following:

- o A canonical version of the RFC using a modernized version of the current text-based format
- o Canonical versions of any art referred to in the canonical RFC
- o Multiple additional forms of the RFC, most notably at least in one or more HTML formats
- o Metadata for the RFC

Today, all RFCs are easily retrievable by all readers. In the future, all of the versions of an RFC, its art, and its metadata must be easily accessible as well. To make this easier, the RFC Editor will establish a permanent URL template for each RFC that leads to a page that lists all of the versions, art, and metadata; a copy of that URL will be included near the beginning of the RFC in the boilerplate so that new RFC readers can find it. Further, it will be easy for advanced RFC users to mirror the entire collection of RFC material.

A major motivation behind the "one canonical, many non-canonical" proposal here is to allow the RFC Editor to easily change the non-canonical formats in the future without having to change the canonical format. For example, the recent discussion of RFC formats has shown that many people strongly desire good HTML versions of RFCs, but there is not agreement of exactly what format the HTML should take. Further, it is completely clear that the HTML standard will evolve in the coming years and decades, and some of the new

features that will be added will be quite useful in RFCs. Allowing the RFC Editor to add additional HTML formats to the RFC collection, even for RFCs that have been published in the past, gives the greatest value to RFC readers without forcing any changes on the canonical RFC format.

Similarly, it is clear that HTML is not the only useful format for RFCs. Some people really like plain text; others want PDFs or other printer-ready paginated formats; still others want different formats that can be converted to different reading devices. Some people want detailed metadata for RFCs so that they can better mine them for relevant information. All of these people can be accommodated by the RFC Editor publishing multiple non-canonical versions of RFCs.

2. Canonical RFC Format

Canonical RFCs are text files. The format of those files is:

- o Text paragraphs consist of a single line with no line wrapping.
- o The document has no page breaks and thus no page headers and footers.
- o It is explicitly allowed to have art in external files. Those files are referred to in figures in the RFC as URLs that point to the RFC Editor's web site. The RFC Editor will determine which graphic formats are allowed, and it is likely that at least SVG and GIF will be permitted.
- o The text encoding for the document is UTF-8. The RFC Editor can decide where it is and is not appropriate to use non-ASCII characters in the RFC. For example, the RFC Editor might make rules about using non-ASCII characters in people's names, reference titles, examples in text, and so on.
- o Text art that internal to the document is limited to 95 columns. This is reasonable for printing on laser printers from the past 25 years, and allows much more expressive art than the current maximum of approximately 70 columns.

3. Additional Formats Provided by the RFC Editor

The RFC Editor will derive and publish non-canonical documents in multiple formats from the RFC. If the RFC-reading community agrees on a single HTML format, that will certainly be published. If the RFC-reading community cannot agree on just one HTML format, the RFC

Editor might publish non-canonical versions of an RFC in multiple HTML formats.

Depending on interest from the RFC-reading community, the RFC Editor will also publish non-canonical versions in other formats. For example, it is likely that the RFC Editor will publish in at least one format of PDF. Because some tools in widespread use rely on the current RFC format, the RFC Editor might also publish a non-canonical version in using the rules in [RFC 2333](#) (line lengths, page headers, and so on).

The RFC Editor will also publish metadata for each canonical RFC. It will include document-wide metadata including title, author information, date of publication, list of references, and so on. The metadata will also include position-dependent information such as a structured table of contents, the start and end lines for each piece of text art, and so on.

The metadata will be available in a file separate from the canonical RFC to which it refers. Some of the non-canonical formats (notably, HTML) allow rich internal metadata; in those cases, the RFC Editor will simply include the metadata in the non-canonical documents themselves.

4. Input Format for RFCs

This document does not suggest an input format for RFCs. The RFC Editor needs to decide that in coordination with all of the representatives of the various RFC streams. Fortunately, the decision of the input format does not need to be tied to the decision of the canonical RFC format or the non-canonical versions available at any time. Even if the canonical RFC format is the one described in this document, there is no reason that the input format to the RFC Editor has to be the same format: it could be XML, HTML, and so on. The conversion from the input format to the canonical and non-canonical formats used by the RFC Editor is done by the RFC Editor with their own tools and human-based processes.

It is important to note that the RFC format described here includes external art files. The RFC Editor will have to also specify how those art files are submitted (by upload, by URL, and so on).

Many RFCs are later revised, sometimes by different authors many years after the original version was published. Also, some RFCs liberally copy a great deal of text from earlier RFCs. Whatever was given to the RFC Editor for a particular RFC should be preserved and made available to other authors, possibly as one of the non-canonical

versions of the RFC.

5. IANA Considerations

None

6. Security Considerations

None

7. Informative References

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

Author's Address

Paul Hoffman
VPN Consortium

Email: paul.hoffman@vpnc.org

