

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 13, 2013

P. Hoffman
VPN Consortium
July 12, 2012

Proposal for Canonical and Other Formats for RFCs
draft-hoffman-rfcformat-canon-others-03

Abstract

This document proposes a new, XML-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 accommodate 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 January 13, 2013.

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 both a canonical version of the RFC that is in XML format and multiple additional forms of the RFC, most notably at least in one or more HTML formats. The XML format used will likely be an updated version of that from [RFC2629], most notably to include in-line graphic art.

It is noted that XML files are not easily readable. However, it is also noted that the canonical version of an RFC doesn't need to be easily readable: only the non-canonical formats derived from the canonical version need to be readable.

Today, all RFCs are easily retrievable by all readers. In the future, all of the versions of an RFC and its art 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 and art; 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 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; such metadata can be contained in either XML or HTML formats. All of these people can be accommodated by the RFC Editor publishing multiple non-canonical versions of RFCs. The canonical version of the RFC and all the non-canonical versions of the RFC should have predictable URLs so that tools can easily find (for example) an RFC in the reader's preferred HTML style just by knowing the RFC number.

The method that the RFC Editor uses to create the non-canonical formats for RFCs is left up to the RFC Editor. For example, they might generate it directly from the input files, through an intermediate format, or something else.

2. Canonical RFC Format and Content

Canonical RFCs are in XML format. The most salient rules for the format and content of those files are:

- o The format for the XML will be specified by the RFC Editor. It is likely that the XML format will be an improvement to that which is now referred to as "xml2rfc" ([RFC2629] and its informal successors).
- o The XML format will allow for art to be contained in the file. This art might be instead of text art in a document (such as for a diagram that is too complex to render well in text), or might be better variants for text art. The RFC Editor will determine which graphic formats are allowed, and it is likely that at least one vector format and one pixel-based format will be permitted.
- o The XML format will contain all of the metadata needed to produce any of the non-canonical formats for an RFC.
- o The text encoding for the document is UTF-8.
- o The RFC Editor can decide where it is and is not appropriate to use non-ASCII characters from the Unicode repertoire 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.

Text art encompasses many types of content. The unifying feature is that it is one or more lines of text that must be rendered with a monospace font in order to be fully understood in the context of the document. Thus, text art includes graphical representations such as packet diagrams, flow diagrams, and flow charts, but it also includes other text that needs to have column alignment such as multi-line ABNF.

This proposal does not deal with how mathematical equations might be included in the canonical RFC format. An author can do it as text or as art in an external file. The RFC Editor might allow an equation-specific format from external art files.

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. The RFC Editor will oversee the development of the tools needed to produce the non-canonical 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).

4. Input to the RFC Series

The RFC Editor will allow submission of RFCs in the same XML format as the canonical version of an RFC. This allows an author to use differencing tools to track all changes that are made to the document that they submitted to the RFC Editor.

The RFC Editor will also possibly allow additional formats for submission based on agreement with the RFC streams. If other submission formats are allowed, the RFC Editor will convert the submission to the canonical format before performing any editing so that all editorial changes are easily tracked within the canonical format. This is similar to what they do currently with submissions that are not in the format the RFC Editor uses for its internal tooling.

This proposal in this document does not affect the allowed format for the publication of Internet-Drafts. The IETF Chair has indicated that such a change might happen after the choices are made for RFC format.

5. Metadata Needed to Create RFCs

The canonical format for RFCs must contain all of the body text for the RFC as well as all of the metadata that is used to mark up the RFC. RFC metadata is useful for many things such as finding RFCs with particular types of content and for making it clearer to a reader what the RFC author intended.

The following is a list of the metadata that needs to be part of the canonical RFC format. This list will probably be controversial, but the eventual list needs to contain all of the metadata that is intended for the final RFC format so that the format can be fully specified. Items marked with an asterisk are especially likely to need much more work.

Status

- Category
- RFC stream
- Obsoletes
- Updates
- Date published
- Draft derived from
- Title and short title

Per-author

- Name
- Initials
- Company
- Postal
- Email
- URL
- Role (editor, other)

Front content

- Abstract

- Copyright
- Other legal *
- Headings
 - Level
 - Number
- Paragraphs
 - Indented for quoting or emphasis
- Lists
 - Style (bulleted, numbered, unmarked)
 - Nesting
 - Elements
 - Definitions
- Validatable formats
 - ABNF
 - XML
 - JSON
 - ASN.1
- Inline art
 - Datagram layout
 - State diagram
 - Flow chart
 - Pseudocode
 - Table *
 - Non-validatable fragments of validatable formats
 - Other
- Tables *
- Special sections
 - Security Considerations
 - IANA Considerations
 - Normative References
 - Informative References
- Cross-references
 - To internal sections
 - To reference
 - To section in reference
- References *
 - RFCs
 - Specific versions of IDs
 - Non-specific versions of IDs
 - Common non-IETF documents (IEEE, ITU, ISO, ANSI, Unicode)
 - Corner cases

6. IANA Considerations

None

7. Security Considerations

None

8. Informative References

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

[RFC2629] Rose, M., "Writing I-Ds and RFCs using XML", RFC 2629, June 1999.

Author's Address

Paul Hoffman
VPN Consortium

Email: paul.hoffman@vpnc.org

