RFC Format Framework
draft-iab-rfc-framework-02

Abstract

   The canonical format for the RFC Series has been plain-text, ASCII-
   encoded for several decades. After extensive community discussion
   and debate, the RFC Editor will be transitioning to XML as the
   canonical format using the XML2RFC version 3 vocabulary. Different
   publication formats will be rendered from that base document. These
   changes are intended to increase the usability of the RFC Series by
   offering documents that match the needs of a wider variety of
   stakeholders. With these changes, however, comes an increase in
   complexity for authors, consumers, and the publisher of RFCs. This
   document serves as the framework that describes the problems being
   solved and summarizes the many documents that capture the specific
   requirements for each aspect of the change in format.

Editorial Note (To be removed by 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.

Copyright Notice

Table of Contents

1. Introduction

   [RFC6949], "RFC Series Format Requirements and Future Development,"
   discussed the need for additional features within RFCs such as non-
   ASCII characters to respect author names, more advanced artwork than
   ASCII art, and documents that could display properly on a wide
   variety of devices. Based on the discussions with the IETF community
   as well as other communities of interest, the RFC Series Editor
   decided to explore a change to the format of the Series
   [XML-ANNOUNCE]. This document serves as the framework that describes
   the problems being solved and summarizes the documents created to-
   date that capture the specific requirements for each aspect of the
   change in format.

   Key changes to the publication of RFCs are highlighted, and a
   transition plan that will take the Series from a plain-text, ASCII-
   only format to the new formats is described [RFC-INTEREST].

   This document is concerned with the production of RFCs, focusing on
   the published formats. It does not address any changes to the
   processes each stream uses to develop and review their submissions
   (specifically, how Internet-Drafts will be developed). While I-Ds
   have a similar set of issues and concerns, directly addressing those
   issues for I-Ds will be discussed within each document stream.

   The details described in this document are expected to change based
   on experience gained in implementing the RFC production center's
   toolset. Revised documents will be published capturing those changes
   as the toolset is completed. Other implementers must not expect
   those changes to remain backwards-compatible with the details
   described this document.

2. Problem Statement

   There are nearly three billion people connected to the Internet, and
   individuals from 45 countries or more regularly attending IETF
   meetings over the last 5 years [ISTATS]. The Internet is now global,
   and while the world has changed from when the first RFCs were
   published, the Series remains critical to defining protocols,
   standards, best practices, and more for this global network that
   continues to grow. In order to make RFCs easily viewable to the
   largest number of people possible, across a wide array of devices,
   and to respect the diversity of authors and reference materials, it
   is time to update the tightly prescribed format of the RFC Series.

All changes to the format of the RFC Series must consider the
requirements of a wide set of communities, over an extended length of
time. For example, existing authors and implementers, lawyers that
argue Intellectual Property Rights (IPR), educators, managers, and
policy-makers that need to know what to list in potential RFPs for
their organizations, all have preferences and requirements for their
specific needs. The immediate needs of today's communities must be
balanced with the needs for long-term archival storage.

3. Terminology

The following terminology is used as described in RFC 6949:

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

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

   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

   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

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

4. Overview of the Decision Making Process

Requirements, use cases, concerns, and suggestions were collected
from the communities of interest at every stage of the RFC format
update project. Input was received through the rfc-interest mailing
list, as well as in several face-to-face sessions at IETF meetings.
Regular conversations were held with the IETF, IRTF, IAB, and IAOC
chairs, and the Independent Stream Editor, to discuss high-level
stream requirements. Updates regarding the status of the project
were provided to the IETF community during the IETF Technical Plenary

as well as Format BoFs or IAB sessions at IETF 84, IETF 85, IETF 88,
IETF 89, and IETF 90 [IETF84] [IETF85] [IETF88] [IETF89] [IETF90].

The first document published, RFC 6949, provided the first solid
documentation on what the requirements were for the Series and in
effect was the output from the first year of discussion on the topic
of RFC format. That RFC, as with all of the RFCs that informed the
format update work, was published as an IAB stream document, thus
following the process described in RFC 4845, "Process for Publication
of IAB RFCs" [RFC4845].

After the high-level requirements were published, the RFC Series
Editor (RSE) brought together an RFC Format Design Team to start
working out the necessary details to develop the code needed to
create new and changed formats. The design team discussed moving
away from the existing xml2rfc vocabulary, but with such a strong
existing support base within the community and no clear value with
other XML vocabularies or schemas, the decision was made to work with
the XML2RFC version 2 (xml2rfc v2) model and use it as the base for
the new format world [I-D.iab-xml2rfcv2]. Part of this discussion
included a decision to stop using an XML document type definition
(DTD) in favor of a Regular Language for XML Next General (Relax NG)
model using a defined vocabulary. While the bi-weekly calls for this
team were limited to Design Team members, review of the decisions as
documented in the drafts produced by this team were done publicly
through requests for feedback on the rfc-interest mailing list.
Several of the drafts produced by the Design Team, including the
xml2rfc v2 and v3 drafts and the SVG profile drafts, were sent
through an early GenART review before starting the process to be
accepted as an IAB stream draft [GEN-ART] [I-D.iab-xml2rfc].

While the IETF community provided the majority of input on the
process, additional outreach opportunities were sought to gain input
from an even broader audience. Informal discussions were held with
participants at several International Association of Scientific,
Technical, and Medical Publisher events, and presentations made at
technical conferences such as the TERENA Networking Conference 2014
and NORDUnet 2014 [TNC2014] [NDN2014].

In order to respond to concerns regarding responses to subpoenas and
to understand the requirements for lawyers, advice was requested from
the IETF Trust legal team regarding what format or formats would be
considered reasonable when responding to a subpoena request for an
RFC.

Given that several other standards development organizations (SDOs)
do not offer plain-text documents, and in fact may offer more than
one format for their standards, informal input was sought from them

regarding their experience with supporting one or more non-plain-text formats for their standards.

Finally, the entire process was reviewed regularly with the RFC Series Oversight Committee and regular updates provided to the IAB and IESG [RSOC]. They have offered support and input throughout the process.

Where consensus was not reached during the process, the RSE made any necessary final decisions, as per the guidance in RFC 6635, "RFC Editor Model (Version 2)" [RFC6635].

5. Key Changes

At the highest level, the changes being made to the RFC Format involve breaking away from a pure-ASCII plain text and moving to canonical format that includes all the information required for rendering a document into a wide variety of publication formats. The RFC Editor will become responsible for more than just the plain-text file and the PDF-from-text format created at time of publication; they will be creating several different formats in order to meet the diverse requirements of the community.

The final XML file produced by the RFC Editor will be considered the canonical format for RFCs; it is the lowest common denominator that holds all the information intended for an RFC. PDF/A-3 will be the publication format offered in response to subpoenas for RFCs published through this new process, and will be developed with an eye towards long-term archival storage. HTML will be the focus of providing the most flexible set of features for an RFC, including JavaScript to provide pointers to errata and other metadata. Plain-text will continue to be offered in order to support existing tool chains where practicable and the individuals who prefer to read RFCs in this format.

6. Canonical Format Documents

6.1. XML for RFCs

Key points regarding the XML format:

o The canonical format for RFCs is XML using the XML2RFC version 3 (xml2rfc v3) vocabulary. This file must contain all information necessary to render a variety of formats; any question about what was intended in the publication will be answered from this format.

o Authors may submit drafts in xml2rfc v2 vocabulary, but the final publication will convert that to xml2rfc v3 vocabulary.

o SVG is supported and will be embedded in the final XML file.

o There will be automatically generated identifiers for sections,
   paragraphs, figures, and tables in the final XML file.

o The XML file will not contain any xml2rfc v3 vocabulary elements
   or attributes that have been marked deprecated.

o A Document Type Definition (DTD) will no longer be used. The
   grammar will be defined using RelaxNG.

o The final XML file will contain, verbatim, the appropriate
   boilerplate as applicable at time of publication specified by RFC
   5741 or its successors [RFC5741].

o The final XML will be self-contained. For instance, all features
   that reference externally-defined input will be expanded. This
   includes all uses of xinclude, src attributes (such as in
   <artwork> or <sourcecode> elements), include-like processing
   instructions, and externally defined entities.

o The final XML will not contain comments or processing
   instructions.

o The final XML will not contain src attributes for <artwork> or
   <sourcecode> elements.

[I-D.iab-xml2rfcv2] Describes the xml2rfc v2 vocabulary. While in
wide use today, this vocabulary had not been formally documented. In
order to understand what needed to change in the vocabulary to allow
for a more simple experience and additional features for authors, the
current vocabulary needed to be fully described. This document, when
published, will be obsoleted by the RFC published from draft-iab-
xml2rfc.

[I-D.iab-xml2rfc] Describes the xml2rfc v3 vocabulary. The design
goals in this vocabulary were to make the vocabulary more intuitive
for authors, and to expand the features to support the changes being
made in the publication process. This draft, when published, will
obsolete the RFC published from draft-iab-xml2rfcv2.

7. Publication Format Documents

7.1. HTML

[I-D.iab-html-rfc] - Describes the semantic HTML that will be
produced by the RFC Editor from the xml2rfc v3 files.

Key points regarding the HTML output:

o The HTML will not be derived from the plain-text publication
   format.

o The body of the document will use a subset of HTML. The documents
   will include CSS for default visual presentation; it can be
   overwritten by a local CSS file.

o SVG is supported and will be included in the HTML file.

o Text will be reflowable.

o JavaScript will be supported only as an additional option for
   presentation of specific publication formats to provide up-to-date
   links to post-publication metadata, such as errata or obsoletion.
   Documents will be complete and readable as at time of publication,
   when JavaScript is disabled.

7.2. PDF

[I-D.hansen-rfc-use-of-pdf] - Describes the tags and profiles that
will be used to create the new PDF format, including both the
internal structure and the visible layout of the file. A review of
the different versions of PDF is offered, with a recommendation of
what PDF standard should apply to RFCs.

Key points regarding the PDF output:

o The PDF file will not be derived from the plain-text publication
   format.

o The PDF publication format will conform to the PDF/A-3 standard
   and will embed the canonical XML source.

o The PDF will look more like the HTML publication format than the
   plain-text publication format.

o The PDF will include a rich set of tags and metadata within the
   document

o SVG is supported and will be included in the PDF file.

7.3. Plain Text

[I-D.iab-rfc-plaintext] - Describes the details of the plain text
format, focusing in particular on what is changing from the existing
plain-text output.

Key points regarding the plain-text output:

o The plain-text document will no longer be the canonical version of
   an RFC.

o The plain-text format will be UTF-8 encoded; non-ASCII characters
   will be allowed.

o A Byte Order Mark (BOM) will be added at the start of each file.

o Widow and orphan control for the plain-text publication format
   will not have priority for the developers creating the rendering
   code [TYPOGRAPHY].

o Authors may choose to have pointers to line art in other
   publication formats in place of ASCII art in the .txt file.

o Both a paginated and an unpaginated plain-text file will be
   created.

o Running headers and footers will not be used.

7.4. Potential Future Publication Formats

7.4.1. EPUB

This format is intended for use by ebook readers and will be
available for RFCs after the requirements have been defined. No
draft is currently available.

8. Figures and Artwork

8.1. SVG

[I-D.iab-svg-rfc] Describes the profile for SVG line art. SVG is an
XML-based vocabulary for creating line drawings; SVG information will
be embedded within the canonical XML at time of publication.

9. Content and Page Layout

9.1. Non-ASCII Characters

There are security and readability implications to moving outside the
ASCII range of characters. [I-D.iab-rfc-nonascii] focuses on exactly
where and how non-ASCII characters may be used in an RFC, with an eye
towards keeping the documents as secure and readable as possible
given the information that needs to be expressed.

9.2. Style Guide

The RFC Style Guide [RFC7322] was revised to remove as much page
formatting information as possible, focusing instead on grammar,
structure, and content of RFCs. Some of the changes recommended,
however, informed the XML v3 vocabulary.

9.3. CSS Requirements

[I-D.iab-rfc-css] describe how the CSS classes mentioned in the HTML
format draft, "HyperText Markup Language Request for Comments
Format", should be used to create an accessible and responsive design
for the HTML format.

10. Transition Plan

10.1. Statement of Work and RFP for Tool Development

Existing tools for the creation of RFCs will need to be updated, and
new tools created, to implement the updated format. As the
requirements gathering effort, described in the various documents
described earlier int this draft, finishes the bulk of the work, the
Tools Development Team of the IETF will work with the RSE to develop
Statements of Work (SoWs). Those SoWs will first be reviewed within
the Tools Development Team, the Tools Management Committee, and go
out for a public comment period. After public review, the SoWs will
be attached to a Request for Proposal (RFP) and posted as per the
IASA bid process [IASA-RFP].

Once bids have been received, reviewed, and awarded, coding will
begin.

10.2. Testing and Transition

During the I-D review and approval process, authors and stream-
approving bodies will select drafts to run through the proposed new
publication process. While the final RFCs published during this time
will continue as plain-text and immutable once published, the
feedback process is necessary to bootstrap initial testing. These
early tests will target finding issues with the proposed xml2rfc v3
vocabulary that result in poorly formed publication formats as well
as issues that prevent proper review of submitted drafts.

Feedback will result in regular iteration of the basic code and XML
vocabulary. In order to limit the amount of time the RFC Production
Center (RPC) spends on testing and QA, note that their priority is to
edit and publish documents, community assistance will be necessary to
help move this stage along. A mailing list and experimental source

directory on the RFC Editor website will be created for community members willing to assist in the detailed review of the XML and publication formats. Editorial checks of the publication formats by the community are out of scope; the focus will be the QA of each available output, checking for inconsistencies across formats.

The purpose of testing phase is to work with the community to identify and fix bugs in the process and the code, before producing canonical, immutable XML, and to collect additional feedback on the usability of the new publication formats.

Success will be measured by the closure of all bugs which had been identified by the RPC and the Tools Development team as fatal and consensus on the readiness of the XML vocabulary and final XML files for publication. The actual rendering engine can go through further review and iteration, as the publication formats may be republished as needed.

Authors are not required to submit their approved drafts in an XML format, though they are strongly encouraged to do so; plain-text will also remain an option for the foreseeable future. However, documents submitted as plain-text cannot include such features as SVG artwork.

A known risk at this point of the transition is the difficulty in quantifying the resources required from the RPC. This phase will require more work on the part of the RPC to support both old and new publication processes for at least six months. There is potential for confusion as consumers of RFCs find some documents published at this time with a full set of outputs, while other documents only have plain text. There may be a delay in publication as new bugs are found that must be fixed before the files can be converted into the canonical format and associated publication formats.

Final success of the transition will be measured by the closure of all bugs which had been identified by the RPC and the Tools Development team as major or critical. There must also be rough consensus from the community regarding the utility of the new formats.

10.3. Completion

Authors may submit XML (preferred) or plain text. The XML drafts submitted for publication will be converted to canonical XML format and published with all available publication formats. All authors will be expected to review the final documents as consistent with the evolving procedures for reviewing drafts.

Success for this phase will be measured by a solid understanding by
the RSE and the IAOC of the necessary costs and resources required
for long-term support of the new format model.

11. IANA Considerations

This document has no actions for IANA.

12. Security Considerations

Changing the format for RFCs involves modifying a great number of
components to publication. Understanding those changes and the
implications for the entire tool chain is critical so as to avoid
unintended bugs that would allow unintended changes to text.
Unintended changes to text could in turn corrupt a standard, practice
or critical piece of information about a protocol.

13. Acknowledgements

With many thanks to the RFC Format Design Team for their efforts in
making this transition successful: Nevil Brownlee (ISE), Tony Hansen,
Joe Hildebrand, Paul Hoffman, Ted Lemon, Julian Reschke, Adam Roach,
Alice Russo, Robert Sparks (Tools Team liaison), and Dave Thaler.

14. Appendix - Change log

To be removed by RFC Editor

14.1. draft-iab-rfc-framework-00 to -01

Decision Making Process: noted taht other XML schemas and
vocabularies were considered by the design team

XML for RFCs: "boilerplate at time of publication"

HTML: clarified that JavaScript should not impact readability of the
document as it looked at time of publication

15. References

15.1. Normative References

[RFC6949] Flanagan, H. and N. Brownlee, "RFC Series Format
           Requirements and Future Development", RFC 6949,
           DOI 10.17487/RFC6949, May 2013,
           <http://www.rfc-editor.org/info/rfc6949>.

   [I-D.iab-xml2rfc]
              Hoffman, P., "The "xml2rfc" version 3 Vocabulary", draft-
              iab-xml2rfc-02 (work in progress), January 2016.

   [I-D.iab-xml2rfcv2]
              Reschke, J., "The 'XML2RFC' version 2 Vocabulary", draft-
              iab-xml2rfcv2-02 (work in progress), September 2015.

   [I-D.iab-svg-rfc]
              Brownlee, N., "SVG Drawings for RFCs: SVG 1.2 RFC", draft-
              iab-svg-rfc-00 (work in progress), January 2016.

   [I-D.iab-html-rfc]
              Hildebrand, J. and P. Hoffman, "HyperText Markup Language
              Request For Comments Format", draft-iab-html-rfc-01 (work
              in progress), January 2016.

   [I-D.hansen-rfc-use-of-pdf]
              Hansen, T., Masinter, L., and M. Hardy, "PDF for an RFC
              Series Output Document Format", draft-hansen-rfc-use-of-
              pdf-08 (work in progress), October 2015.

   [I-D.iab-rfc-plaintext]
              Flanagan, H., "Requirements for Plain-Text RFCs", draft-
              iab-rfc-plaintext-01 (work in progress), January 2016.

   [I-D.iab-rfc-nonascii]
              Flanagan, H., "The Use of Non-ASCII Characters in RFCs",
              draft-iab-rfc-nonascii-00 (work in progress), January
              2016.

   [I-D.iab-rfc-css]
              Flanagan, H., "CSS Requirements for RFCs", draft-iab-rfc-
              css-00 (work in progress), January 2016.

15.2. Informative References

   [RFC4845] Daigle, L., Ed. and Internet Architecture Board, "Process
              for Publication of IAB RFCs", RFC 4845,
              DOI 10.17487/RFC4845, July 2007,
              <http://www.rfc-editor.org/info/rfc4845>.

   [RFC5741] Daigle, L., Ed., Kolkman, O., Ed., and IAB, "RFC Streams,
              Headers, and Boilerplates", RFC 5741,
              DOI 10.17487/RFC5741, December 2009,
              <http://www.rfc-editor.org/info/rfc5741>.

   [RFC6635] Kolkman, O., Ed., Halpern, J., Ed., and IAB, "RFC Editor
              Model (Version 2)", RFC 6635, DOI 10.17487/RFC6635, June
              2012, <http://www.rfc-editor.org/info/rfc6635>.

   [RFC7322] Flanagan, H. and S. Ginoza, "RFC Style Guide", RFC 7322,
              DOI 10.17487/RFC7322, September 2014,
              <http://www.rfc-editor.org/info/rfc7322>.

   [GEN-ART] IETF, "General Area Review Team (Gen-ART)", n.d.,
              <http://www.ietf.org/iesg/directorate/gen-art.html>.

   [IASA-RFP]
              IETF Administrative Support Activity, "RFPs and RFIs",
              n.d., <http://iaoc.ietf.org/rfps-rfis.html>.

   [IETF84]   Flanagan, H., "IETF 84 Proceedings: RFC Format (rfcform)",
              n.d., <http://www.ietf.org/proceedings/84/rfcform.html>.

   [IETF85]   Flanagan, H., "IETF 85 Proceedings: RFC Format (rfcform)",
              n.d., <http://www.ietf.org/proceedings/85/rfcform.html>.

   [IETF88]   Flanagan, H., "IETF 88 Proceedings: RFC Format (rfcform)",
              n.d., <http://www.ietf.org/proceedings/88/rfcform.html>.

   [IETF89]   Flanagan, H., "IETF 89 Proceedings: RFC Format (rfcform)",
              n.d., <http://www.ietf.org/proceedings/89/rfcform.html>.

   [IETF90]   Flanagan, H., "IETF 90 Proceedings: RFC Format (rfcform)",
              n.d., <http://www.ietf.org/proceedings/90/rfcform.html>.

   [ISTATS]   "Internet Live Stats", n.d.,
              <http://www.internetlivestats.com/internet-users/>.

   [NDN2014] "28th NORDUnet Conference 2014", 2014,
              <https://events.nordu.net/display/NORDU2014/
              BoF%27s+and+side+meetings>.

   [RFC-INTEREST]
              RFC Editor, "rfc-interest -- A list for discussion of the
              RFC series and RFC Editor functions.", n.d.,
              <https://www.rfc-editor.org/mailman/listinfo/rfc-
              interest>.

   [RSOC]   IAB, "RFC Editor Program: The RSOC", n.d.,
              <http://www.iab.org/activities/programs/
              rfc-editor-program/>.

   [TNC2014] Flanagan, H., "IETF Update - 'What's Hot?' - RFC Update",
              n.d., <https://tnc2014.terena.org/core/presentation/84>.

   [TYPOGRAPHY]
              Butterick, M., "Butterick's Practical Typography", n.d.,
              <http://practicaltypography.com/
              widow-and-orphan-control.html>.

   [XML-ANNOUNCE]
              "Subject: [rfc-i] Direction of the RFC Format Development
              effort", n.d., <http://www.rfc-editor.org/pipermail/
              rfc-interest/2013-May/005584.html>.

Author's Address

   Heather Flanagan
   RFC Editor

   Email: rse@rfc-editor.org
   URI: http://orcid.org/0000-0002-2647-2220