Internet Architecture Board (IAB)
Request for Comments: 7995
Category: Informational
ISSN: 2070-1721

T. Hansen, Editor
AT&T Laboratories
L. Masinter
M. Hardy
Adobe

December 2016

PDF Format for RFCs

Abstract

This document discusses options and requirements for the PDF rendering of RFCs in the RFC Series, as outlined in RFC 6949. It also discusses the use of PDF for Internet-Drafts, and available or needed software tools for producing and working with PDF.

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 7841.

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/rfc7995[1].

Copyright Notice

[1] http://www.rfc-editor.org/info/rfc7995
[2] http://trustee.ietf.org/license-info

Table of Contents

1. Introduction

The RFC Series is evolving, as outlined in [RFC6949]. Future documents will use a canonical format, XML, with renderings in various formats, including PDF.

Because PDF has a wide range of capabilities and alternatives, not all PDFs are "equal". For example, visually similar documents could consist of scanned or rasterized images, or include text layout options, hyperlinks, embedded fonts, and digital signatures. (See [APP-PDF] for a history of PDF.)

This document explains some of the relevant options and makes recommendations, for both the RFC Series and Internet-Drafts.

The PDF format and the tools to manipulate it are not as well known as those for the other RFC formats, at least in the IETF community. This document discusses some of the processes for creating and using PDFs using both open source and commercial products.

The details described in this document are expected to change based on experience gained in implementing the new publication toolsets. Revised documents will be published capturing those changes as the toolsets are completed. Other implementers must not expect those changes to remain backwards-compatible with the details described in this document.

2. Choosing PDF Versions and Standards

PDF [PDF] has gone through several revisions, primarily for the addition of features. PDF features have generally been added in a way that older viewers "fail gracefully", but even so, the older the PDF version produced, the more legacy viewers will support that version but the fewer features will be enabled.

As PDF has evolved a broad set of capabilities, additional standards for PDF files are applicable. These standards establish ground rules that are important for specific applications. For example, PDF/X was specifically designed for Prepress digital data exchange, with careful attention to color management and printing instructions. The PDF/E standard was designed for engineering documents with dynamic workflows (where a document continues to be revised after publication) and allows interactive media (including animation and 3D).

Two additional standards families are important to the RFC format, though: long-term preservation (PDF/A), and user accessibility (PDF/UA [PDFUA]). These then have sub-profiles (PDF/A-1, PDF/A-2 [PDFA2], PDF/A-3 [PDFA3]), each of which has conformance levels. These standards are then supported by various software libraries and tools.

It is effective and useful to use these standards to capture PDF for RFC requirements, and they will make the PDF files useful in workflows that expect them.

Recommendations:

- Use PDF 1.7; although relatively recent, it is well supported by widely available viewers.
- For RFCs, require PDF/A-3 with conformance level "U". This captures the archivability and long-term stability of PDF 1.7 files, mandatory Unicode mapping (Sections 14.8.2.4.2 ("Unicode Mapping in Tagged PDF") and 9.10.2 ("Mapping Character Codes to Unicode Values") of [PDF]), and many of the requirement features.
- Use PDF/A-3 for embedding additional data (including the XML source file) in RFCs and Internet-Drafts.
- Use PDF/UA for user accessibility.

3. Options and Requirements for PDF RFCs

This section lays out options and requirements for PDFs produced by the RFC Editor for RFCs. There are two subsections: Section 3.1 covers "visible" requirements related to how the PDF normally appears when it is viewed with a PDF viewer; Section 3.2 covers "invisible" options and requirements, which primarily affect the ability to process PDFs in other ways but do not ordinarily control the way the document appears. (Of course, a viewer UI might display processing capabilities, such as showing whether a document has been digitally signed.)

In many cases, the choice of PDF requirements is heavily influenced by the capabilities of available tools to create PDFs. Most of the discussion of tooling is to be found in Appendix C.

3.1. "Visible" Requirements

PDF supports rich visible layout of fixed-sized pages.

3.1.1. General Visible Requirements

For a consistent "look" of RFCs and good style, the PDFs produced by the RFC Editor should have a clear, consistent, identifiable, and easy-to-read style. They should print well on the widest range of printers and should look good on displays of varying resolution.

3.1.2. Page Size and Margins

PDF files are laid out for a particular size of page and margins. There are two paper sizes in common use: "US Letter" (8.5x11 inches, 216x279 mm, in popular use in North America) and "A4" (210x297 mm, 8.27x11.7 inches, standard for the rest of the world). Usually, PDF printing software is used in a "shrink to fit" mode where the printing is adjusted to fit the paper in the printer. There is some controversy, but the argument that A4 is an international standard is compelling. However, if the margins and header positioning are chosen appropriately, the document can be printed without any scaling.

Recommendation: The Internet-Draft and RFC processors should produce A4 size by default. However, the margins and header positioning need to be chosen to look good on both paper sizes without scaling. Following the advice found in [RFC2346], this means that we should use A4 portrait mode with left and right margins of 20 mm, and top and bottom margins of 33 mm.

3.1.3. Headers and Footers

Page headers and footers are part of the page layout. There are a variety of options. Note that page headers and footers in PDF can be typeset in a way that the entire (longer) title might fit.

Recommendation: Page headers and footers should contain information similar to the headings in the current text versions of documents, including page numbers, title, author, and date. However, the page headers and footers should be typeset in a way so as to be unobtrusive. The page headers and footers should be placed into the PDF in such a way that they do not interfere with screen readers.

3.1.4. Paragraph Numbering

One common feature of the Internet-Draft output formats is optional visible paragraph numbers, to aid in discussions. In the PDF, and thus in the printed rendition, it is possible to make paragraph numbers unobtrusive and even to impinge on the margins.

Recommendation: When the XML "editing=yes" option has been chosen, show paragraph numbers in the right margin, typeset in a way so as to be unobtrusive. (The right margin instead of the left margin prevents the paragraph numbers from being confused

with the section numbers.) If possible, the paragraph numbers should be coded in such a way that they do not interfere with screen readers.

3.1.5. Paged Content Layout

By its nature, PDF is paginated, so pagination issues must be considered. This is reflected in two areas: running headers and footers, and how text is laid out on a page for optimal reading.

Appendix B describes the process of creating a paged document from running text such that related material is present on the same page together and artifacts of pagination don't interfere with easy reading of the document.

Layout engines differ in the quality of the algorithms used to automate these processes. In some cases, the automated processes require some manual assistance to ensure, for example, that a text line intended as a heading is "kept" with the text for which it is a heading.

Recommendations:

- Headers and footers should be printed on each page. The information should include the RFC number or Internet-Draft name, the page number, the category (e.g., Informational), a shortened version of the authors' names, the date of the RFC or Internet#Draft, and the short form of the document title.
- Choose a layout engine so that
  - manual intervention is minimized
  - widow and orphan processing is automatic
  - heading and title contiguation is automatic

3.1.6. Typeface Choices

A PDF may refer to a font by name, or it may use an embedded font. When a font is not embedded, a PDF viewer will attempt to locate a locally installed font of the same name. If it cannot find an exact match, it will find a "close match". If a close match is not available, it will fall back to something implementation dependent and usually undesirable.

In addition, the PDF/A standards mandate the embedding of fonts. Instead of using additional software to embed the fonts, the software generating the PDF files should produce PDF/A-conforming files directly, thus ensuring that all glyphs include Unicode mappings and embedded fonts from the outset.

If the HTML version of the document is being visually mimicked, the font(s) chosen should have both variable-width and constant-width components, as well as bold and italic representations.

The typefaces used by Internet-Drafts and by RFCs need not be identical.

Few fonts have glyphs for the entire repertoire of Unicode characters; for this purpose, the PDF generation tool may need a set of fonts and a way of choosing them. The RFC Editor is defining where Unicode characters may be used within RFCs [RFC7997].

Typefaces are typically licensed, and in many cases there is a fee for use by PDF creation tools; however, there is usually no fee for display or print of the embedded fonts.

Recommendations:

- For consistent viewing, all fonts should be embedded. The fonts used must be available for use by the IETF community. Some discussion of available typefaces can be found in Appendix C.4.
- The choice of typefaces with respect to serif, sans-serif, monospace, etc., should follow the recommendations for HTML and CSS renderings ("CSS" refers to a Cascading Style Sheet) [RFC7992] and [RFC7993].
- The range of Unicode characters allowed in the XML source for Internet-Drafts and RFCs may be bounded by the availability of embeddable fonts with appropriate glyphs [RFC7997].

3.1.7. Hyphenation and Line Breaks

Typically, when doing page layout of running text, especially with narrow page width and long words, layout processors of English text often have the option of either hyphenating words or using existing hyphens as a place to introduce word breaks. However, inserting line breaks mid-word can be harmful when the "word" is actually a sequence of characters representing a protocol element or protocol sequence.

Recommendation: Avoid introducing hyphenated line breaks mid-word into the visual display, consistent with requirements for plain text and HTML.

3.1.8. Hyperlinks

PDF supports hyperlinks to sections of the same document and also to sections of other documents.

The conversion to PDF can generate:

- hyperlinks within the document
- hyperlinks to other RFCs and Internet-Drafts
- hyperlinks to external locations
- hyperlinks within a table of contents
- hyperlinks within an index

Recommendations:

- All hyperlinks available in the HTML rendition of the RFC should also be visible and active in the PDF produced. This includes both internal hyperlinks and hyperlinks to external resources.
- The table of contents, including page numbers, is useful when printed. Section numbers and page numbers in the table of contents should also be hyperlinked to their respective sections in the body of the document.
- As specified in Section 4.8.6.2 ("Referencing RFCs") of [RFC7322], hyperlinks to RFCs from the references section should point to the RFC "info" page (e.g., <https://www.rfc-editor.org/info/rfc7322>), which then links to the various formats available.
- Hyperlinks to Internet-Drafts from the references section should point to the Datatracker entry page for the draft, which then links to the various formats available.

3.1.9. Similarity to Other Outputs

There is some advantage to having the PDF files look like the text or HTML renderings of the same document. Even so, there are several options. The PDF

1. could look like the text version of the document, or
2. could look like the text version of the document but with pictures rendered as pictures instead of using their ASCII art equivalent, or
3. could look like the HTML version.

Recommendation: The PDF rendition should look like the HTML rendition, at least in spirit. Some differences from the HTML rendition would include different typeface and size (chosen for printing), page numbers in the table of contents and index, and the use of page headers and footers.

Most of the choices used for the renderings per [RFC7992] and [RFC7993] are thus applicable. See those documents for specifics on the rendering of the specific XML elements. Some notes:

- Every place in the document that would receive an HTML ID would be given an identical PDF named destination. In addition, a named destination will be created for each page with the form "pg-#", as in "pg-35".
- No pilcrows are generated or made visible.
- The table of contents (generated if the XML's <rfc> element's tocInclude attribute has the value "true") [RFC7991] will have the section number linked to the section start but will also include a page number that

is linked to the corresponding page. The section title and the page number will be separated by a visually appropriate separator, and the page numbers will be aligned with each other.

- The index (generated if the XML's <rfc> element's indexInclude attribute has the value "true") will have the section number linked to that section named destination but will also include a page number that is linked to the page named destination.
- The running header in one line (on page 2 and all subsequent pages) has the RFC number on the left (RFC NNNN), the (possibly shortened form) title centered, and the date (Month Year) on the right. The text is rendered in a way that is visually unobtrusive.
- The running footer in one line (on all pages) has the author's last name on the left, category centered, and the page number on the right ([Page N]). The text is rendered in a way that is visually unobtrusive.
- We should not attempt to replicate in PDF the feature of the HTML format that includes a dynamic block that displays up-to-date information on updates, obsoletions, and errata.

3.2. "Invisible" Options and Requirements

PDF offers a number of features that improve the utility of PDF files in a variety of workflows, at the cost of extra effort in the xml2rfc conversion process; the trade-offs may be different for the RFC Editor production of RFCs and for Internet-Drafts.

3.2.1. Internal Text Representation

The contents of a PDF file can be represented in many ways. The PDF file could be generated:

- as an image of the visual representation, such as a JPEG image of the word "IETF". That is, there might be no internal representation of letters, words, or paragraphs at all.
- placing individual characters in position on the page, such as saying "put an 'F' here," then "put a 'T' before it," then "put an 'E' before that," then "put an 'I' before that" to render the word "IETF". That is, there might be no internal representation of words or paragraphs at all.
- placing words in position on the page, such as keeping the characters of the word "IETF" together. That is, there might be no internal representation of paragraphs at all.
- ensuring that the running order of text in the content stream matches the logical reading order. That is, a sentence such as "The Internet Engineering Task Force (IETF) supports the Internet." would be kept together as a sentence, and multiple sentences within a paragraph would be kept together.

All of these end up with essentially the same visual representation of the output. However, each level has trade-offs for auxiliary uses, such as searching or indexing, commenting and annotation, and accessibility (text-to-speech). Keeping the running order of text in the content stream in the proper order supports all of these auxiliary uses.

In addition, the "role map" feature of PDF (Section 14.7.3 ("Structure Types") of [PDF]) would allow for the mapping of the logical tags found in the original XML into tags in the PDF.

Recommendations:

- Text in content streams should follow the XML document's logical order (in the order of tags) to the extent possible. This will provide optimal reuse by software that does not understand Tagged PDF. (PDF/UA requires this.)
- It might be possible to use the "role map" annotation to capture enough of the xml2rfc source structure, to the point where it is possible to reconstruct the XML source structure completely. However, there is not a compelling case to do so over embedding the original XML, as described in Section 3.2.7.

3.2.2. Unicode Support

PDF itself does not require the use of Unicode. Text is represented as a sequence of glyphs that can then be mapped to Unicode.

Recommendations:

- PDF files generated must have the full text, as it appears in the original XML.
- Unicode normalization may occur.
- Text within SVG for SVG images should also have Unicode mappings.
- Alt-text for images should also support Unicode.

3.2.3. Image Processing (Artwork)

The XML allows both ASCII art and SVG to be used for artwork.

Recommendations:

- If both ASCII art and SVG are available for a picture, the SVG artwork should be preferred over the ASCII artwork.
- ASCII artwork must be rendered using a monospace font.

3.2.4. Text Description of Images (Alt-Text)

Guidelines for the accessibility of PDF <http://www.w3.org/TR/WCAG20#TECHS/PDF1.html> recommend that images, formulas, and other non-text items provide textual alternatives, using the "/Alt" Tag in PDF to provide human-readable text that can be vocalized by text-to-speech technology.

Recommendation: Any alt-text for artwork and figures available in the XML source should be stored using the PDF /Alt property. Internet-Draft authors and the RFC Editor should ensure that alt#text for all SVG or images is included within the XML source.

3.2.5. Metadata Support

Metadata encodes information about the document authors, the document series, date created, etc. Having this metadata within the PDF file allows it to be used by search engines, viewers, and other reuse tools. PDF supports embedded metadata in a variety of ways, including using the Extensible Metadata Platform (XMP) [XMP]. The RFC Editor maintains metadata about an RFC on its info page.

Recommendation: The PDFs generated should have all of the metadata from the XML version embedded directly as XMP metadata, including the author, date, the document series, and a URL for where the document can be retrieved. This information should be consistent with the RFC Editor info page at the time of publication.

3.2.6. Document Structure Support

PDF supports an "outline" feature where sections of the document are marked; this could be used in addition to the table of contents as a navigation aid.

The section structure of an RFC can be mapped into the PDF elements for the document structure. This will allow the bookmark feature of PDF readers to be used to quickly access sections of the document.

Recommendation: The section structure of an RFC should be mapped into the PDF elements for the document structure. This would include section headings for the boilerplate sections, such as the Abstract, the Status of This Memo section, the table of contents, and the Author's Address section, plus the obvious section headings that are normally included in the table of contents. If possible, this should be done in a way that the same fragment identifiers for the HTML version of the RFC will work for the PDF version.

3.2.7. Embedded Files

PDF has the capability of including other files; the files may be labeled by both a media type and a role, the AFRelationship key [PDFA3]. In this way, the PDF file also acts as a container.

Embedded content may be compressed.

Many PDF viewers support the ability to view and extract embedded files, although this capability is not universal.

Embedding content in the PDF file allows the PDF to act as a complete package that can be transformed, archived, and digitally signed. (Some sample code illustrating how items can be attached to a PDF file and subsequently extracted can be found at <https://github.com/Aiybe/xmptest>.) Useful possibilities:

- Embed the source XML input file itself within the PDF. If the source SVG and images for illustrations are also embedded, this would make the PDF file totally self-referential.
- Embed directly extractable components that are useful for independent processing, including ABNF, MIBs, and source code for reference implementations. This capability might be supported through other mechanisms from the XML source files but could also be supported within the PDF.
- Finding, extracting, and embedding other components may require additional markup to clearly identify them and additional review to ensure the correctness of embedded files that are not visible.

Recommendations:

- Embed the XML source and all illustrations, for RFCs, as a standard feature for xml2rfc's PDF output.
- If possible, make this a standard feature for Internet-Drafts as well.
- Named <sourcecode> entries should be embedded.
- Bitmap images (SVG sources, JPEGs, PNGs, etc.) should be embedded.

3.3. Digital Signatures

The RFC Editor and staff are at times called to provide evidence that a particular RFC is the "original" and has not been modified; digital signatures can provide that verification. As signatures also apply to embedded content, embedding the XML source will provide a way of signing the source XML that was used to produce the PDF file as well.

PDF has supported digital signatures since PDF 1.2, and there are multiple methods and options available for signing PDF files. The method chosen for the signing of Internet#Drafts and RFCs will be determined by separate policy.

If PDF digital signatures are chosen, the authors suggest the following:

- PDF documents generated by the Internet-Draft upload tools should be signed with no restrictions on what can be done to the documents afterwards.
- If Internet-Drafts are allowed to be uploaded in PDF form by an individual, the signature being added should be set in the same way as that noted in the previous paragraph. A PDF that would not allow the IETF Secretariat to re#sign it in that fashion should be rejected.
- PDF documents generated by the RFC Editor should be signed and certified, and restrictions placed on them to only allow additional signatures and comments (markup) to be added.

4. Security Considerations

The following security considerations apply:

Threats:

- There is a risk that user-submitted Internet-Drafts in PDF might contain malware that targets a vulnerability in one of the deployed PDF consumers (readers, printers, validation tools, etc.) in use.
- There is a small risk that a PDF production toolset might itself have some vulnerability by which it could be tricked into producing malware-bearing PDF files.
- Section 7 of [RFC3778] describes some additional security considerations for PDF, although this specification is intended to avoid features (like scripting) that might trigger some of those concerns.

Mitigations:

- The toolsets for producing PDFs need careful security reviews before deploying broadly.
- If users are allowed to submit Internet-Drafts in PDF, such PDF files should be examined carefully for conformance to this specification, as well as any known exploits of deployed PDF software.

5. References

5.1. Normative References

[PDF] ISO, "Document management -- Portable document format -- Part 1: PDF 1.7", ISO 32000-1, 2008. Also available free from Adobe.

[XMP] ISO, "Graphic technology -- Extensible metadata platform (XMP) specification -- Part 1: Data model, serialization and core properties", ISO 16684-1, 2012.
Not available free, but there are a number of descriptive resources, e.g., <http://en.wikipedia.org/wiki/Extensible_Metadata_Platform>.

[PDFA2] ISO, "Document management -- Electronic document file format for long-term preservation -- Part 2: Use of ISO 32000-1 (PDF/A-2)", ISO 19005-2, 2011.

[PDFA3] ISO, "Document management -- Electronic document file format for long-term preservation -- Part 3: Use of ISO 32000-1 with support for embedded files (PDF/A-3)", ISO 19005-3, 2012.

[PDFUA] ISO, "Document management applications -- Electronic document file format enhancement for accessibility -- Part 1: Use of ISO 32000-1 (PDF/UA-1)", ISO 14289-1, 2014.

[RFC3778] Taft, E., Pravetz, J., Zilles, S., and L. Masinter, "The application/pdf Media Type", RFC 3778, DOI 10.17487/RFC3778, May 2004, <https://www.rfc-editor.org/info/rfc3778>.

5.2. Informative References

[RFC7991] Hoffman, P., "The "xml2rfc" Version 3 Vocabulary", RFC 7991, DOI 10.17487/RFC7991, December 2016, <http://www.rfc-editor.org/info/rfc7991>.

[RFC7997] Flanagan, H., Ed., "The Use of Non-ASCII Characters in RFCs", RFC 7997, DOI 10.17487/RFC7997, December 2016, <http://www.rfc-editor.org/info/rfc7997>.

[RFC7993] Flanagan, H., "Cascading Style Sheets (CSS) Requirements for RFCs", RFC 7993, DOI 10.17487/RFC7993, December 2016, <http://www.rfc-editor.org/info/rfc7993>.

[RFC7992] Hildebrand, J., Ed. and P. Hoffman, "HTML Format for RFCs", RFC 7992, DOI 10.17487/RFC7992, December 2016, <http://www.rfc-editor.org/info/rfc7992>.

[APP-PDF] Hardy, M., Masinter, L., Markovic, D., Johnson, D., and M. Bailey, "The application/pdf Media Type", Work in Progress, draft-hardy-pdf-mime-04, September 2016.

[RFC2346] Palme, J., "Making Postscript and PDF International", RFC 2346, DOI 10.17487/RFC2346, May 1998, <https://www.rfc-editor.org/info/rfc2346>.

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

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

A. History and Current Use of PDF with RFCs and Internet-Drafts

NOTE: This section is meant as an overview to give some background.

A.1. RFCs

The RFC Series has for a long time accepted Postscript renderings of RFCs, either in addition to or instead of the text renderings of those same RFCs. These have usually been produced when there was a complicated figure or mathematics within the document. For example, consider the figures and mathematics found in RFCs 1119 and 1142, and compare the figures found in the text version of RFC 3550 with those in the Postscript version. The RFC Editor has provided a PDF rendering of RFCs. Usually, this has been a print of the text file that does not take advantage of any of the broader PDF functionality, unless there was a Postscript version of the RFC, which would then be used by the RFC Editor to generate the PDF.

A.2. Internet-Drafts

In addition to PDFs generated and published by the RFC Editor, the IETF tools community has also long supported PDF for Internet-Drafts. Most RFCs start with Internet-Drafts, edited by individual authors. The Internet-Drafts submission tool at <https://datatracker.ietf.org/submit/> accepts PDF and Postscript files in addition to the (required) text submission and (currently optional) XML. If a PDF wasn't submitted for a particular version of an Internet-Draft, the tools would generate one from the Postscript, HTML, or text.

B. Paged Content Layout Quality

The process of creating a paged document from running text typically involves ensuring that related material is present on the same page together and that artifacts of pagination don't interfere with easy reading of the document. Typical high-quality layout processors do several things:

Widow and Orphan Management:
: Widows and orphans (<https://en.wikipedia.org/wiki/Widows_and_orphans>) should be avoided automatically (unless the entire paragraph is only one line). Ensure that a page break does not occur after the first line of a paragraph (orphans), if necessary, using slightly longer page sizes. Similarly, ensure that a page break does not occur before the last line of a paragraph (widows).

Keep Section Heading Contiguous:
: Do not insert a page break immediately after a section heading. If there isn't room on a page for the first (two) lines of a section after the section heading, insert a page break before the heading.

Avoid Splitting Artwork:
: Figures should not be split from figure titles. If possible, keep the figure on the same page as the (first) mention of the figure.

Headers for Long Tables after Page Breaks:
: Another common option in producing paginated documents is to include the column headings of a table if the table cannot be displayed on a single page. Similarly, tables should not be split from the table titles.

keepWithNext and keepWithPrevious:
: The XML attributes "keepWithNext" and "keepWithPrevious" should be used and followed whenever possible.

Whitespace Preservation:
: The Unicode Points for XML entities such as Non-Breaking Space (nbsp) and

Non-Breaking Hyphen (nbhy) should be followed as directed whenever possible.

C. Tooling

This section discusses tools for viewing, comparing, creating, manipulating, and transforming PDF files, including those currently in use by the RFC Editor and Internet-Drafts, as well as outlining available PDF tools for various processes.

C.1. PDF Viewers

As with most file formats, PDF files are experienced through a reader or viewer of PDF files. For most of the common platforms in use (iOS, OS X, Windows, Android, ChromeOS, Kindle) and for most browsers (Edge, Safari, Chrome, Firefox), PDF viewing is built in. In addition there are many PDF viewers available for download and installation.

PDF viewers vary in capabilities, and it is important to note which PDF viewers support the features utilized in PDF RFCs and Internet#Drafts (features such as links, digital signatures, Tagged PDF, and others mentioned in Section 3).

C.2. Printers

While almost all viewers also support the printing of PDF files, printing is one of the most important use cases for PDFs. Some printers have direct PDF support.

C.3. PDF Generation Libraries

Because the xml2rfc format is a unique format, software for converting XML source documents to the various formats will be needed, including PDF generation.

One promising direction is suggested in <http://greenbytes.de/tech/webdav/rfc2629xslt/rfc2629xslt.html#output.pdf.fop>: using XSLT (Extensible Stylesheet Language Transformations) to generate XSL#FO (XSL Formatting Objects); XSL-FO is then processed by a FOP (Formatting Objects Processor) such as Apache FOP.

Several libraries are also available for generating PDF signatures. The choice of library to use for xml2pdf will depend on many factors: programming language, quality of implementation, quality of PDF generated, support, cost, availability, and so forth.

C.4. Typefaces

Various typefaces are available that might satisfy the requirements of this document. Google's Noto typeface family <https://www.google.com/get/noto/> supports a significant subset of Unicode and includes fixed-width, serif, and sans-serif styles. Another potentially useful set of typefaces (without extensive Unicode support, however) includes:

- Source Sans Pro <https://en.wikipedia.org/wiki/Source_Sans_Pro>
- Source Serif Pro <https://en.wikipedia.org/wiki/Source_Serif_Pro>
- Source Code Pro <https://en.wikipedia.org/wiki/Source_Code_Pro>

Another font that looks promising for its broad Unicode support is Skolar <https://www.rosettatype.com/Skolar>, but it requires licensing.

C.5. Other Tools

In addition to generating and viewing PDF, other categories of PDF tools are available and may be useful both during specification development and for published RFCs. These include tools for comparing two PDFs, checkers that could be used to validate the results of conversion, reviewing and commentary tools that attach annotations to PDF files, and digital signature creation and validation.

Validation of an arbitrary author-generated PDF file would be quite difficult; there are few PDF validation tools. However, if RFCs and Internet-Drafts are generated by conversion from XML via xml2rfc, then explicit

validation of PDF and adherence to expected profiles would mainly be useful to ensure that xml2rfc has functioned properly.

Recommendation:	Discourage (but allow) submission of a PDF representation for Internet-Drafts. In most cases, the PDF for an Internet-Draft should be produced automatically when XML is submitted, with an opportunity to verify the conversion.

IAB Members at the Time of Approval

The IAB members at the time this memo was approved were (in alphabetical order):

Jari Arkko
Ralph Droms
Ted Hardie
Joe Hildebrand
Russ Housley
Lee Howard
Erik Nordmark
Robert Sparks
Andrew Sullivan
Dave Thaler
Martin Thomson
Brian Trammell
Suzanne Woolf

Acknowledgements

The input of the following people is gratefully acknowledged: Nevil Brownlee (ISE), Brian Carpenter, Chris Dearlove, Martin Duerst, Heather Flanagan (RSE), Joe Hildebrand, Paul Hoffman, Duff Johnson, Ted Lemon, Sean Leonard, Henrik Levkowetz, Julian Reschke, Adam Roach, Leonard Rosenthol, Alice Russo, Robert Sparks, Andrew Sullivan, and Dave Thaler.

Authors' Addresses

Tony Hansen (editor)
AT&T Laboratories
200 Laurel Ave. South
Middletown, NJ 07748
United States of America
EMail: tony@att.com

Larry Masinter
Adobe
345 Park Ave.
San Jose, CA 95110
United States of America
EMail: masinter@adobe.com
URI: http://larrymasinter.net

Matthew Hardy
Adobe
345 Park Ave.
San Jose, CA 95110
United States of America
EMail: mahardy@adobe.com