

RFC 6949 : RFC Series Format Requirements and Future Development

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 15 mai 2013

Date de publication du RFC : Mai 2013

<https://www.bortzmeyer.org/6949.html>

Cela fait longtemps que le format des RFC est critiqué, à l'IETF ou ailleurs. Pour d'excellentes raisons (détaillées plus loin), ce format est très ancien et très rigide, sans concessions aux gadgets modernes. Il n'a pas encore été possible de le modifier, à la fois en raison des difficultés de prise de décision au sein de la communauté, et aussi parce que le problème est réellement difficile. Contrairement à ce que prétendent certains yakafokon, trouver un format idéal n'est pas trivial. Ce RFC est la première étape d'une tentative de réforme, commençant par un cahier des charges sur les changements souhaitables. Les règles rigides de formatage (longueur des pages et des lignes) disparaissent, l'Unicode fait son entrée dans les RFC, et la possibilité d'inclure des images apparaît.

Une des raisons de la stagnation du projet de changement depuis des années était la réorganisation de la fonction de "*RFC Editor*". Désormais, les nouvelles règles sur cette fonction sont écrites (RFC 6635¹), une nouvelle "*RFC Editor*" a pris ses fonctions, jeune et dynamique (Heather Flanagan, un des auteurs de ce RFC), et il va peut être être enfin possible de faire évoluer le format traditionnel. (Depuis cet article, le nouveau format des RFC a été adopté, cf. RFC 7990.)

Le format actuel est documenté dans le RFC 7322 mais à l'époque de notre RFC, c'était le RFC 2223 qui s'appliquait : texte brut sans enrichissement, en ASCII, pages de 58 lignes, lignes de 72 caractères au maximum, pas de notes de base de page, références croisées par numéro de section et pas de page, etc. PostScript était autorisé et quelques RFC ont été publiés ainsi. Une tentative de mise à jour de ce RFC, avec écriture d'une proposition de nouvelle version, avait eu lieu en 2004 mais avait été abandonné. J'avais déjà écrit un article détaillé <<https://www.bortzmeyer.org/rfc-en-texte-brut.html>> expliquant le pourquoi du format traditionnel et les raisons pour lesquelles il est difficile à changer.

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc6635.txt>

La section 1 de ce RFC commence par rappeler la longue histoire des RFC. Ils ont commencé il y a plus de quarante ans, et dans les débuts, certains étaient écrits à la main (!), d'autres tapés à la machine, d'autres encore produits par les outils de formatage invraisemblables qu'on trouvait sur les "mainframes" de l'époque. Résultat, certains sont devenus illisibles avec le temps, les outils tombant en désuétude. Ce problème de permanence est un des problèmes les plus cruciaux pour les RFC et explique pourquoi le "RFC Editor" n'a jamais été très chaud à l'idée d'utiliser le logiciel à la mode du moment : l'informatique en a déjà enterré beaucoup.

Au bout d'un moment, un format standard a émergé, fondé sur du texte brut et un encodage en ASCII. À partir de là, la série des RFC a vécu une longue période de calme et de stabilité et cette permanence (songez aux formats à la mode vite disparus comme WordPerfect) a beaucoup contribué à son succès. On peut relire aujourd'hui sans problème un RFC de 1980 comme le RFC 768. Ce n'est pas que le format ait été jugé parfaitement satisfaisant par tout le monde, ni qu'aucune proposition de changement n'ait été faite, bien au contraire. Mais aucun des ces propositions n'a connu de consensus : toutes résolvaient un problème précis... en en créant plusieurs autres.

Avant de regarder les propositions actuellement sur la table, il faut bien comprendre qu'il y a plusieurs endroits dans le cycle de vie d'un RFC où un format standard est nécessaire ou utile. Cette terminologie est très importante car l'incompréhension de ces étapes du cycle de vie est l'une des causes principales de la confusion des débats sur la réforme des RFC :

- Un format de **soumission**, celui dans lequel le futur RFC est remis par l'auteur au "RFC Editor". Aujourd'hui, c'est obligatoirement du texte brut, que l'auteur peut accompagner de XML (RFC 7749) s'il le souhaite.
- Un format de **révision**, celui avec lequel travaille le "RFC Editor". C'est actuellement du nroff, pour lequel il est de plus en plus dur de trouver des outils.
- Un format de **publication** dans lequel le RFC va être distribué au monde. Cette fois, il n'est pas obligé d'être unique. Le même RFC peut être distribué sous plusieurs formats, automatiquement produits. Aujourd'hui, le "RFC Editor" publie le texte brut et du PDF, d'autres publient du HTML (sur l'excellent <<http://tools.ietf.org>> par exemple voici ce RFC <<http://tools.ietf.org/html/rfc6949>>), du PDF formaté différemment, etc.
- Un format **canonique** qui fait référence en cas de désaccord (rappelez-vous que pas mal de RFC sont des documents **normatifs**). Actuellement, c'est la version en texte brut.

Évidemment, les choses ont changé depuis (le RFC 2223 date de seize ans...) Quels sont les points qui suscitent le plus de protestations ? D'abord, l'utilisation d'art ASCII et lui seul pour les graphiques. Elle a l'avantage d'encourager les auteurs à écrire des descriptions textuelles claires, plutôt que de passer du temps à faire de jolis dessins. Sans les limitations de l'art ASCII, les diagrammes seraient probablement moins concis. D'un autre côté, des choses comme les automates finis sont vite illisibles en art ASCII et les protocoles où il y a plus de trois acteurs qui interagissent sont difficiles à dessiner. Des graphiques plus élaborés que l'art ASCII permettraient également de représenter des équations mathématiques complexes.

Après les images, le sujet le plus chaud est évidemment l'encodage. ASCII a des tas d'avantages : standard, très répandu, facile à manipuler (outils de recherche, par exemple), plus pratique lorsqu'il faut renumériser un document (oui, le cas s'est déjà produit dans le passé)... Et la langue des RFC étant de toute façon l'anglais, les limites d'ASCII ne sont pas trop graves.

Mais il y a quand même des limites : lorsqu'un RFC parle d'internationalisation (comme le RFC 5890 ou le RFC 6530, ou comme de plus en plus de normes), c'est bien dommage de ne pas pouvoir donner d'exemple lisible dans ce RFC. Devoir écrire U+00c9 pour parler de É est pénible... Des exemples comme ceux de XMPP (RFC 6121, section 5.2.3) bénéficieraient de l'acceptation d'Unicode dans les RFC.

Et certains auteurs de RFC ont des noms qui ne peuvent pas s'écrire correctement en ASCII et ils regrettent de devoir les modifier.

Autres points sur lesquels des gens ont souvent râlé contre le format actuel : la pagination et l'ajustement du texte. La pagination en unités fixes fait vraiment rétro : elle avait un sens quand les RFC étaient souvent imprimés pour être lus mais à l'ère des tablettes, cela semble bien dépassé. Plus compliqué est le cas de l'ajustement du texte. La largeur de colonne fixe qui est la norme actuelle convenait aux VT100 mais pas aux écrans modernes, avec leurs tailles très variables.

Enfin, ce n'est pas tout de définir des formats, il faut aussi des outils qui rendent la production de RFC conforme à ces formats facile (pour avoir une idée des difficultés, lire le RFC 5385). Ces outils doivent-ils être développés à l'IETF ou bien doit-on utiliser des outils standard ? Les exigences de l'IETF sont-elles tellement spécifiques qu'on ne puisse pas utiliser des outils existants ? D'un autre côté, les outils spécifiques donnent plus de souplesse et de pouvoir et, aujourd'hui, il en existe déjà beaucoup (par exemple, la version 2 de `xml2rfc` <<http://tools.ietf.org/tools/xml2rfc2/>> - une réécriture complète, abandonnant TCL pour Python, pour produire des RFC suivant le schéma du RFC 7749, est bien meilleure que la précédente) donc on peut dire que le modèle « faisons tout nous-même » marche bien (il y a beaucoup d'informaticiens à l'IETF).

Enfin, s'il y a des réclamations des lecteurs et des auteurs, il y en a aussi du "*RFC Editor*" : il voudrait pouvoir abandonner `nrff` (65 % des RFC aujourd'hui sont soumis au format XML, qu'il faut traduire en `nrff` au lieu de l'utiliser directement comme format d'entrée pour les différentes publications).

Après les contraintes et les réclamations, les décisions (section 3). D'abord, ce qui ne change **pas** :

- Stabilité du RFC une fois publié (jamais de correction mais des errata <<http://www.rfc-editor.org/errata.php>>),
 - Format canonique stable et standard (pas question d'utiliser Word),
 - Le texte brut reste la meilleure solution dans bien des cas donc au moins un des formats de publication doit être le texte brut,
 - Le contenu, par exemple les avertissements juridiques reste comme spécifié dans le RFC 7841.
- Ensuite, les nouvelles exigences effectivement retenues, notamment :
- Documents accessibles même aux personnes en situation de handicap (notez que certaines personnes réclamaient de la couleur dans les RFC, ce qui peut être un problème pour l'accessibilité),
 - Possibilité d'utiliser l'Unicode d'une manière très encadrée, notamment l'ASCII est obligatoire pour le texte normatif (pas de passage global d'ASCII vers UTF-8, mais possibilité d'utiliser UTF-8 par exemple pour les... exemples),
 - Possibilité d'inclure des graphiques en SVG (sous une forme qui reste à déterminer) mais uniquement en noir & blanc,
 - Nécessité de garder un mécanisme permettant des polices de chasse fixe et du texte qui ne s'ajuste pas, pour l'art ASCII et pour les exemples de code source,
 - Format canonique permettant de marquer clairement le code ou assimilé (MIB, ABNF...).
 - Autant que possible, le RFC en cours de révision devrait tenir dans un petit nombre de fichiers (pas une "*tarball*" avec des tas de trucs),
 - Le nombre de formats publiés directement par le "*RFC Editor*" restera faible, pour que ledit éditeur puisse se concentrer sur le contenu et pas sur la génération de fichiers dans des formats divers.

Et il y a aussi des exigences qui ont été supprimées :

- La limite de longueur des pages disparaît (plus de pages de taille fixe),
- Même chose pour celle de la longueur des lignes. Ces deux exigences du RFC 2223 ayant disparue, cela ouvre la voie à du texte ajustable,
- Le contenu ne sera plus obligé d'être à 100 % en ASCII.

Voilà, le cahier des charges est posé, il reste à définir le nouveau format, développer les outils, et déployer le tout... Ce sera pour le prochain épisode. À la réunion IETF 86 a été annoncé <<https://www.ietf.org/ibin/c5i?mid=6&rid=49&gid=0&k1=934&k2=11939&tid=1368216404>> le choix de s'appuyer sur le format XML comme format canonique, avec comme formats de publication le texte brut, HTML, PDF et EPUB (voir la FAQ <<http://www.rfc-editor.org/rse/FormatFAQ.html>> sur ces décisions.) Deux "*Internet-Drafts*" sont en cours, un sur le format, `draft-rfc-format-flanagan`, et un sur le style, `draft-flanagan-style`. (Depuis, le premier "*draft*" a été publié, RFC 7990.)