

RFC 7992 : HTML Format for RFCs

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 16 décembre 2016

Date de publication du RFC : Décembre 2016

<https://www.bortzmeyer.org/7992.html>

Depuis la sortie du RFC 7990¹ et ses copains, le format canonique (le format de référence) des RFC n'est plus le texte seul mais un langage XML, normalisé dans le RFC 7991. Les formats « de publication » seront produits automatiquement à partir de ce XML. C'est ainsi que notre RFC 7992 décrit la sortie HTML, qui permettra de publier des beaux RFC sur le Web.

HTML, trop riche et trop mouvant, est en effet mal adapté à l'écriture et à la maintenance de documents. Il n'était donc pas envisageable de le choisir comme format canonique. En revanche, il est incontournable comme **format de publication**. C'est en effet le langage du Web, et il y a donc logiquement une forte demande pour pouvoir lire les RFC en HTML. Avant, lorsque le format canonique était le texte brut <<https://www.bortzmeyer.org/rfc-en-texte-brut.html>>, des versions non officielles étaient publiées en HTML (voir un exemple <<https://tools.ietf.org/html/rfc7285>>) mais, le texte brut n'ayant pas de formatage précis, ces versions n'avaient pas vraiment l'air de vraies pages Web...

(Notez que ce blog que vous êtes en train de lire est produit par un mécanisme analogue à celui que les RFC suivront désormais : tapé en XML, avec le HTML produit automatiquement <<https://www.bortzmeyer.org/blog-implementation.html>>.)

La section 1 de notre RFC résume les principes de l'HTML utilisé. D'abord, ce sera un sous-ensemble de HTML (HTML a bien trop de fonctions). Ensuite, la présentation sera largement délégué à une feuille de style CSS, dont les caractéristiques sont mentionnées dans le RFC 7993.

La section 2, elle, est le « cahier des charges » du HTML des RFC. Elle précise les exigences du RFC 6949. Elle concerne les auteurs du ou des logiciels de production des RFC (pour ces logiciels, voir le RFC

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc7990.txt>

7998). Les auteurs de RFC, eux, n'ont pas à s'en soucier, ils écrivent du XML, et le HTML sera produit par les outils.

Le but principal est que l'HTML produit soit parfaitement lisible sur la grande majorité des navigateurs utilisés. Pas question bien sûr d'ajouter une de des ridicules mentions « optimisé pour Internet Explorer » qui étaient si communes sur les sites Web d'amateurs, dans les années 2000. Notre RFC mentionne explicitement l'exigence que les textes soient lisibles avec au moins un navigateur « texte », comme Lynx, certaines personnes accédant au Web ainsi (par obligation ou par goût). C'est l'une des raisons de la décision de ne pas accepter la totalité de HTML.

Le fichier HTML devra être autonome (ne pas dépendre de fichiers extérieurs), de manière à pouvoir être transmis facilement par des mécanismes tels que rsync ou le courrier électronique.

Le JavaScript est accepté mais à condition qu'il ne modifie en aucun cas le texte du RFC. (Il peut, par exemple, ajouter des éléments de navigation, ou afficher des métadonnées.)

On l'a dit, CSS sera utilisé pour la présentation, mais le cahier des charges exige qu'on puisse facilement remplacer la feuille de style par une de son choix, pour s'adapter aux goûts locaux.

Le Web étant fondé sur la notion de lien hypertexte, il y aura évidemment des liens, aussi bien ceux mis explicitement par l'auteur (« ce point est développé en section N »), que ceux ajoutés automatiquement (de la table des matières vers les différentes sections, par exemple).

Un point crucial est évidemment l'accessibilité. Comme le savent tous ceux et toutes celles qui vont régulièrement à Paris Web, c'est un point essentiel mais souvent oublié. Notre RFC note donc que les publications en HTML des futurs RFC devront être accessibles aux malvoyants, aux daltoniens, et aux utilisateurs de petits écrans, par exemple les "smartphones". (Note personnelle : ce dernier point ne devrait pas être dans une section sur l'accessibilité. Le Web est prévu - contrairement aux formats du monde du papier, comme PDF - pour être visible sur tous les périphériques.)

Au fait, quelle version de HTML sera utilisée (section 3 de notre RFC)? Ce sera HTML5 (et pas, et je le déplore, XHTML; l'inconvénient, souvent cité contre XHTML, de la difficulté à l'écrire n'a pas de sens ici, puisque le HTML sera produit automatiquement).

La section 4 précise la syntaxe utilisée (rappelez-vous qu'on n'accepte pas la totalité de HTML5) : encodage en UTF-8, sauts de ligne en style Unix (un U+000A et rien d'autre), pas de caractères de contrôle comme la tabulation (U+0009). Les éventuels commentaires du source XML ne seront pas mis dans le HTML (l'idée est qu'ils sont pour les auteurs, pas pour les lecteurs).

Il y a des objets divers qu'on retrouve souvent dans le source XML. Ils sont rassemblés dans la section 5. Par exemple, on trouve les identificateurs qui seront mis comme valeur des attributs `id` dans le HTML produit. Ce sont parfois des identificateurs mis explicitement par l'auteur, et parfois des identificateurs produits par le logiciel, par exemple pour que les entrées de la table des matières pointent vers la section correspondante.

Autre objet récurrent, la marque de paragraphe ("*pilcrow*" pied-de-mouche, caractère Unicode U+00B6, celui-ci : [Caractère Unicode non montré ²]), qui sera mise automatiquement derrière chaque paragraphe, mais pas affiché par défaut (il faudra promener le pointeur dessus pour le voir).

2. Car trop difficile à faire afficher par \LaTeX

Maintenant, attaquons les différentes parties du RFC rendu en HTML. D'abord (section 6), les premiers objets HTML qu'on rencontrera, notamment les métadonnées du RFC. Il y aura évidemment un DOCTYPE identifiant le document comme du HTML5. L'élément racine sera `<html>`, avec une étiquette de langue qui sera bien sûr en, l'anglais. L'élément `<head>` qui suivra contiendra une déclaration de jeu de caractère, un titre, et diverses métadonnées :

```
<meta charset="utf-8">
<title>The Mother of all RFCs</title>
<meta name="author" content="Joe Hildebrand">
<meta name="author" content="Heather Flanagan">
<meta name="description" content="This document defines...">
<meta name="generator" content="xmljade v0.2.4">
<meta name="keywords" content="html,css,rfc">
```

(Rappelez-vous que le HTML produit n'est hélas pas du XHTML donc il est normal que les `<meta>` ne soient pas explicitement fermés.) Il y aura aussi un lien vers la licence des RFC, en utilisant le cadre général des liens (RFC 8288) :

```
<link rel="license" href="https://www.rfc-editor.org/copyright/">
```

Cette première partie du RFC produit contiendra aussi une feuille de style, ainsi qu'un lien vers une éventuelle feuille locale, au cas où un lecteur souhaiterait lire le RFC selon un style différent :

```
<style>
  body {}
  ...
</style>
<link rel="stylesheet" type="text/css" href="rfc-local.css">
```

Le début de la partie visible du RFC sera composée d'une `<dl>` pour les métadonnées affichées, et d'une table des matières. Les métadonnées seront donc du genre :

```
<dl id="identifiers">
  <dt>Workgroup:</dt>
  <dd class="workgroup">rfc-interest</dd>
  <dt>Series:</dt>
  <dd class="series">Internet-Draft</dd>
  <dt>Status:</dt>
  <dd class="status">Informational</dd>
  <dt>Published:</dt>
  <dd><time datetime="2014-10-25"
    class="published">2014-10-25</time></dd>
  ...
```

La partie principale du RFC sera, elle, rendue selon les principes décrits en section 9 pour chacun des éléments XML qui composent le source.

La dernière partie du RFC inclura un index (si le source XML avait un attribut `indexInclude` dans l'élément `<rfc>`), les adresses des auteurs (formatées en hCard), et les métadonnées considérées comme les moins importantes (les autres ayant été mises au début).

La section 9 de notre RFC est particulièrement longue car elle décrit le rendu en HTML de tous les éléments du vocabulaire XML du RFC 7991. Je ne vais pas tout décrire ici, juste donner quelques exemples. Ainsi, `<artwork>` sera rendu dans un élément HTML `<pre>`, si le schéma était en art ASCII, sera inclus tel quel dans le HTML si le schéma était en SVG (RFC 7996), et sera mis sous forme d'un `` (avec contenu de plan `data:`) dans les autres cas. `<sourcecode>`, lui, est toujours restitué sous forme d'un `<pre>` HTML.

La traduction de certains éléments en HTML est plus directe. Par exemple, `` est simplement rendu par le même élément HTML.

Et, pour finir, un petit mot sur la sécurité (section 11) : comme les RFC en HTML ne sont pas forcément téléchargés depuis le Web mais peuvent être lus depuis un fichier local (après, par exemple, synchronisation via `rsync`), on ne bénéficie pas forcément des protections du navigateur. Donc, prudence.