

RFC 1738 : Uniform Resource Locators (URL)

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 24 Janvier 2008

Date de publication du RFC : Décembre 1994

<http://www.bortzmeyer.org/1738.html>

En même temps qu'étaient normalisés les URI, dans le RFC 1630¹, un effort était en cours pour normaliser leurs prédécesseurs, les URL et cet effort a donné ce RFC.

Ce RFC a aujourd'hui essentiellement un intérêt historique, puisque la spécification des URI, RFC 3986, le remplace en grande partie. Il a gardé longtemps son intérêt (et son statut de norme, désormais perdu) du fait qu'il normalisait certains **plans** d'URL comme `news:`. Ceux-ci sont désormais dans le RFC 5538.

Les URL sont issus du travail dont le cahier des charges est décrit dans le RFC 1737. Contrairement aux URI, plus généraux, ils sont supposés ne servir que lorsqu'on connaît la localisation exacte d'une ressource, de façon à pouvoir la récupérer, ou la mettre à jour (section 2). En pratique, la distinction n'est pas évidente. Certains URL ne donnent pas de moyen de récupérer une ressource (comme `news:` que spécifie notre RFC ou bien `mailto:`, désormais dans le RFC 6068). Et même des URL de plan `http:` peuvent être utilisés uniquement pour identifier, pas pour récupérer une ressource (c'est notamment courant dans le monde XML).

Notre RFC décrit donc la syntaxe des URL (sections 2.1 et 5), bien connue désormais qu'on voit des URL même sur le flanc des autobus, et leur encodage (section 2.2, où est documentée la très mauvaise décision d'encoder les octets et pas les caractères Unicode).

La section 3 décrit ensuite les éléments communs à tous les URL (comme la syntaxe `host:port`), puis des **plans** ("*scheme*") pour plusieurs catégories d'URL. Ces plans ont presque tous été mis à jour dans des RFC ultérieurs mais certains restent normalisés uniquement ici comme `ftp:`. On trouve des plans très utilisés comme le célèbre `http:`, section 3.3 (désormais dans la section 3.2 du RFC 2616),

1. Pour voir le RFC de numéro NNN, <http://www.ietf.org/rfc/rfcNNN.txt>, par exemple <http://www.ietf.org/rfc/rfc1630.txt>

mais aussi d'autres désormais dépassés comme son défunt concurrent `gopher:` (section 3.4, désormais dans le RFC 4266) ou comme l'ancêtre `wais:` (section 3.9, officiellement abandonné par le RFC 4156).

Le schéma `news:`, section 3.6, permet de récupérer un article d'Usenet sans indiquer de nom de serveur NNTP.

L'IANA maintient le registre de ces plans <<http://www.iana.org/assignments/uri-schemes.html>>.

On notera que notre RFC 1738 normalisait également un mécanisme pour décrire un URL au milieu de texte brut, en utilisant un préfixe défini dans l'annexe A, <URL:<http://www.bortzmeyer.org/1738.html>>. Ce mécanisme n'a pas eu de succès et a officiellement été abandonné par le RFC 3986, annexe C.