

RFC 4395 : Guidelines and Registration Procedures for New URI Schemes

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 11 Mai 2009

Date de publication du RFC : Février 2006

<http://www.bortzmeyer.org/4395.html>

Les URI, ces identificateurs normalisés dans le RFC 3986¹ commencent tous par un **plan** ("*scheme*"), la partie de l'URI avant le premier deux-points. Ce RFC explique les procédures d'enregistrement dans le registre des plans (<http://www.iana.org/assignments/uri-schemes.html>).

Il existe de nombreux plans, du très connu `http:` au moins fréquent `tag:` (RFC 4151) en passant par bien d'autres, souvent assez confidentiels comme le `dns:` du RFC 4501 ou le `dict:` du RFC 2229. Le RFC 3986 décrit seulement la syntaxe générique des URI, celle commune à tous les URI et qui se limite largement à « un plan, deux points, puis du texte » (par exemple, `http://www.bortzmeyer.org/608.html` ou `tag:bortzmeyer.org,2006-02:Blog/608`). La grande majorité du contenu de l'URI a une signification qui dépend du plan et un logiciel d'usage très général doit donc connaître ces plans et leurs particularités. Et si j'ai une idée géniale de nouveaux URI et que je veux réserver un plan, je fais comment ? Je lis ce RFC 4395. Que me dit-il ?

La section 2 du RFC fournit des critères pour déterminer si un nouveau plan d'URI est une bonne idée. Ainsi, pour que son enregistrement soit accepté, le plan doit présenter une utilité à long terme (section 2.1). Cette restriction est justifiée par le fait que tout nouveau plan peut nécessiter une modification de tous les logiciels qui traitent des URI et agissent différemment selon le plan. (Le RFC note également que, bien que l'espace de nommage des plans soit infini, en pratique, il pourrait y avoir une concurrence trop forte pour les noms courts et facilement mémorisables et que cela justifie donc de ne pas accepter toutes les candidatures.)

Il y a aussi des contraintes plus techniques. Ainsi, la section 2.2 impose que le nouveau plan respecte les règles syntaxiques existantes. Ainsi, le `//` a une signification bien précise dans un URI, il précède

¹Pour voir le RFC de numéro NNN, <http://www.ietf.org/rfc/rfcNNN.txt>, par exemple <http://www.ietf.org/rfc/rfc3986.txt>

le nom de la machine qui sert d'**autorité de nommage**, pour attribuer le reste de l'URI (section 3.2 du RFC 3986). En l'absence d'une telle machine de référence, le // ne doit donc pas être utilisé. (c'est pour cela que `dict://dict.example.org/d:chocolate:` a un //, car il contient le nom d'un serveur, ici `dict.example.org` alors que `mailto:echo@generic-nic.net` n'en a pas, car les adresses de courrier sont globales, elles ne dépendent pas d'un serveur particulier).

Il faut bien sûr que le plan soit correctement et complètement défini (sections 2.3 et 2.4). Par exemple, si l'URI est résolvable, sa description doit expliquer comment. Autre exemple, la description doit expliquer ce qu'on peut faire de l'URI. Accéder (et parfois modifier) à une ressource (cas du `http:`)? À une machine (cas de `telnet:`)?

Enfin, le plan doit recevoir un nom (section 2.8), qui soit à la fois assez court pour être pratique et assez long pour être descriptif. Pire, comme ces plans sont visibles (des URI sont souvent affichés dans des publications, sur les cartes de visite, sur les publicités), le nom du plan ne doit pas interférer avec des marques déposées défendues par des bataillons d'avocats. Il vaut donc mieux ne pas essayer de normaliser le plan `coca-cola:` (qui permettrait d'écrire des choses utiles comme `coca-cola:light`)... Le RFC recommande aussi d'éviter des noms trop marketing comme tout ce qui contient « *"universal"* » ou « *"standard"* ».

Ces règles de la section 2 sont obligatoires pour les plans enregistrés de manière permanente. Mais le registre (<http://www.iana.org/assignments/uri-schemes.html>) contient aussi des plans enregistrés à titre provisoire et les règles en question ne sont qu'indicatives pour eux (section 3).

Bref, une fois qu'on a bien lu toutes ces considérations, on peut passer à l'enregistrement proprement dit. Pour cela, il faut suivre la procédure exposée en section 5 (qui utilise les termes du RFC 5226). On remplit un formulaire (section 5.4), il est examiné (dans la plupart des cas) sur une liste de diffusion comme `uri@w3.org`, puis par un expert (section 5.2).

C'est par exemple le processus qu'a suivi le plan `geo` (<http://geouri.org/>), qui permet des URI comme `geo:48.208333,16.372778` (la cathédrale Saint-Étienne à Vienne). La section 3 de la norme sur `geo:`, le RFC 5870, contient le formulaire d'enregistrement :

3. IANA Registration of the 'geo' URI Scheme

This section contains the fields required for the URI scheme registration, following the guidelines in Section 5.4 of [RFC4395].

3.1. URI Scheme Name

`geo`

3.2. Status

`permanent`

3.3. URI Scheme Syntax

The syntax of the 'geo' URI scheme is specified below in Augmented Backus-Naur Form (ABNF) [RFC5234]:

```

geo-URI      = geo-scheme ":" geo-path
geo-scheme   = "geo"
geo-path     = coordinates p
coordinates  = coord-a "," coord-b [ "," coord-c ]

coord-a     = num...

```

<http://www.bortzmeyer.org/4395.html>

À noter que, bien qu'il existe une norme pour les URI en Unicode (le RFC 3987, sur les IRI), il n'y a pas de plan en caractères non-ASCII.

Ce RFC remplace les anciens RFC 2717 et RFC 2718. Ces deux RFC avaient été écrits avec l'idée d'une stricte séparation entre **localisateurs** (les URL) et **noms** (les URN). À l'usage, cette distinction est apparemment bien moins claire qu'on ne le pensait, beaucoup de gens ont utilisé, à juste titre, des URL (par exemple des URI `http:`) comme noms et, inversement, ont développé des mécanismes pour accéder à une ressource à partir d'un nom. Notre RFC 4395 ne parle donc plus que d'URI. Un autre changement est que le RFC 2717 avait tenté d'organiser hiérarchiquement les plans alors que nous sommes revenus à un espace de nommage plat.