

RFC 7595 : Guidelines and Registration Procedures for URI Schemes

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 24 juin 2015

Date de publication du RFC : Juin 2015

<https://www.bortzmeyer.org/7595.html>

Ce n'est pas tous les jours qu'on enregistre un **plan** d'URI (le plan - "*scheme*" en anglais - est le premier composant d'un URI, la partie avant le premier deux-points). Ils sont peu nombreux et on voit rarement autre chose que le fameux `http:`. Mais, si jamais vous voulez ajouter un plan à la liste existante, ce RFC vous explique les règles d'enregistrement. Il remplace le RFC 4395¹.

Les plans d'URI sont normalisés dans le RFC 3986, section 3.1. Le plan est souvent appelé à tort « protocole » alors que, dans la grande majorité des cas, il n'a aucun rapport avec un protocole (voir la section 3.8 de notre RFC), et, même quand le nom du plan est celui d'un protocole (comme `http:`), il n'implique pas l'utilisation de ce protocole (un URI ne sert pas forcément à accéder à une ressource). 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 `acct:` du RFC 7565 ou le `dict:` du RFC 2229. La liste des plans enregistrés se trouve à l'IANA <<https://www.iana.org/assignments/uri-schemes/uri-schemes.xml>>. 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 syntaxe et une signification qui dépendent du plan et un logiciel d'usage très général doit donc connaître ces plans et leurs particularités.

Il existe un registre des plans <<https://www.iana.org/assignments/uri-schemes.html>>, qui permet aux développeurs de ces applications de trouver tous les plans en un endroit, et de limiter le risque de collision. La politique d'enregistrement est délibérément assez libérale pour éviter la prolifération de plans non-enregistrés.

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc4395.txt>

Petite note d'internationalisation au passage : les plans sont les mêmes pour les URI (qui doivent s'écrire uniquement en ASCII) et les IRI du RFC 3987, qui peuvent utiliser Unicode (ce n'était pas clair dans le RFC précédent, le RFC 4395).

En section 3 de notre RFC, les règles à suivre. D'abord, la syntaxe générale des URI doit être respectée (ce qui est facile, elle est très générale). Par exemple, l'identificateur de fragment (ce qui suit le croisillon, cf. RFC 3986, section 3.5), garde la même sémantique (notamment le fait qu'il n'est utilisé que par le client, pas par le serveur Web), un nouveau plan d'URI ne peut pas utiliser cette syntaxe pour exprimer autre chose.

Est-ce que le nouveau plan d'URI est une bonne idée? Pour que son enregistrement soit accepté, le plan doit présenter une utilité à long terme, répond la section 3.1. 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. Et le Web contient autre chose que des clients et serveurs, il y a les relais, les caches, etc. (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. La définition d'un nouveau plan doit décrire la syntaxe de l'URI (rappelez-vous que la syntaxe générale des URI ne couvre qu'une petite partie des règles). La section 3.2 impose que le nouveau plan respecte les règles syntaxiques existantes. Par exemple le // a une signification bien précise dans un URI, il précède 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). D'une manière générale, notre RFC déconseille l'utilisation de la barre oblique, sauf si on accepte des URI relatifs, pour éviter qu'un logiciel trop zélé n'ajoute en prime un traitement spécial pour `.` ou `..` (souvent interprétés pour dire « ce niveau » ou « le niveau supérieur »).

Il faut bien sûr que le plan soit correctement et complètement défini (sections 3.3 à 3.5). 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:`, où le modèle « accès à une ressource » ne s'applique pas)? S'il existe une opération par défaut (« je lance mon navigateur Web sur `example:foo-bar-42`, que fait-il? », elle doit être sûre, au sens où elle ne doit pas avoir d'effets de bord. Enfin, le cas des URI qui servent juste à faire correspondre à un identificateur non-URI est normalement plus simple, avec une définition assez évidente. C'est le cas par exemple des `mid:` du RFC 2392, qui transforment un `Message-ID:` du courrier électronique en URI.

On a parlé plus haut des IRI. La section 3.6 nous rappelle les principes d'internationalisation des URI, et donne des bons conseils (faire bien attention à ne pas autoriser plusieurs représentations d'un même caractère, cf. RFC 3986, section 2.5). Le RFC donne aussi de mauvais conseils comme de prétendre (sans donner un seul exemple) qu'il faut restreindre le plus possible le jeu de caractères autorisé, certains caractères étant « dangereux ».

Pendant qu'on parle de danger, la section 3.7 rappelle la nécessité de bien documenter les questions de sécurité et de vie privée liées au nouveau plan. C'est une des nouveautés par rapport à l'ancien RFC 4395 que cette mention de la vie privée. Cela reflète l'importance croissante accordée à ce problème à l'IETF : de nombreux RFC mentionnent désormais la vie privée et pas seulement la sécurité informatique traditionnelle.

Enfin, le plan doit recevoir un nom (section 3.8), qui soit à la fois assez court pour être pratique et assez long pour être descriptif (et ce nom doit suivre la syntaxe de la section 3.1 du RFC 3986 qui limite notamment à ASCII, même pour des IRI). Pire, comme ces plans sont visibles (des URI sont souvent affichés dans les 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* ».

Pour les plans privés, spécifiques à une organisation et non enregistrés, notre RFC recommande d'utiliser un nom de domaine inversé comme préfixe, afin de limiter les risques de collisions (sections 3.8 et 6). Si la société Example, titulaire du domaine `example.com`, veut un plan pour son système Foofoo, elle utilisera donc `com.example.foofoo:` comme plan d'URI. Cela évite toute collision avec une autre société qui aurait un Foofoo. (Le RFC 4395 recommandait un tiret au lieu d'un point dans ce nom inversé.)

Ces règles de la section 3 sont obligatoires pour les plans enregistrés de manière permanente. Mais le registre `<https://www.iana.org/assignments/uri-schemes.xml>` contient aussi des plans enregistrés à titre provisoire et les règles en question ne sont qu'indicatives pour eux (section 4). Les enregistrements provisoires sont soumis à une politique plus libérale, « premier arrivé, premier servi ». Cela permet de demander que les plans privés soient, s'ils ne sont pas fabriqués à partir d'un nom de domaine comme l'exemple précédent, enregistrés de manière provisoire. Parmi les plans provisoires, au moment de cet article, `bitcoin:`, `dtm:` (cf. RFC 5050), etc.

Notez que, dans le registre des plans d'URI `<https://www.iana.org/assignments/uri-schemes/uri-schemes.xml>`, outre les caractéristiques « permanent » et « provisoire », il y a aussi « historique » qui désigne les plans abandonnés (comme le `fax:` du RFC 2806).

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 7 (qui utilise les termes du RFC 5226 comme « examen par un expert » ou « premier arrivé, premier servi »). On remplit un formulaire (section 7.4 pour un formulaire vierge), il est examiné (dans la plupart des cas) sur une liste de diffusion comme `uri-review@ietf.org`, puis par un expert (section 7.2).

Une éventuelle mise à jour d'un enregistrement se fait par le même mécanisme (section 7.3).

Une des nouveautés de notre RFC, par rapport à son prédécesseur RFC 4395, est la création d'un plan d'URI qui sert à la documentation (dans l'esprit du RFC 5398 pour les numéros d'AS et RFC 2606 pour les noms de domaine). Si vous voyez un URI qui commence par `example:`, c'est... un exemple. Imaginons une base de données qui stocke des URI et qu'on montre un exemple d'export de cette base en JSON, on est sûr de ne pas entrer en conflit avec un vrai URI, et on n'a pas trop à se soucier de la syntaxe (le plan `example:` autorise tout), en utilisant ce plan :

```
{ "uris":
  { "date": "2014-09-02 09:58:23+00:00",
    "uri": "example:do-some-thing#test" },
  { "date": "2015-06-24 16:11:09+00:00",
    "uri": "example:foo.bar" },
  ...
}
```

Le formulaire d'enregistrement de ce plan figure dans la section 8 de notre RFC, si vous envisagez d'enregistrer un plan et que vous voulez un exemple. Si vous voulez un vrai exemple récent, vous pouvez regarder la section 7 du RFC 7565, qui enregistrait `acct:`. (Pour les enregistrements provisoires, la demande est indiquée à partir du registre IANA.)

L'annexe A de notre RFC liste les changements qui se sont produits depuis le RFC 4395. Les plus importants, à mon avis, sont :

- Fusion des registres permanent et provisoire en un seul registre `<https://www.iana.org/assignments/uri-schemes/uri-schemes.xml>`, avec une colonne "Status",
- Libéralisation de l'enregistrement des plans provisoires (cf. RFC 5226), de « Examen par un expert » à « Premier arrivé, premier servi » (les plans permanents restent à « Examen par un expert »),
- Ajout du plan `example:`,
- Changement des conventions pour les préfixes des plans privés non enregistrés (point au lieu du tiret entre les composants du nom de domaine),
- Diverses clarifications, par exemple sur le fait que tout plan d'URI s'applique aux IRI aussi bien qu'aux URI,
- Ajout de la vie privée aux considérations sur la sécurité.

Il y avait eu d'autres idées pendant le développement de ce RFC, mais toutes n'ont pas été retenues. Par exemple, à l'IETF 90 `<http://www.ietf.org/meeting/90/index.html>`, nous avons discuté de faire des plans avec préfixes (`coap+ws:`, `coap+sms:`, etc, au lieu du simple `coap:` du RFC 7252), par analogie avec les types MIME structurés du RFC 6839.