RFC 8141 : Uniform Resource Names (URNs)

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 30 avril 2017

Date de publication du RFC : Avril 2017

https://www.bortzmeyer.org/8141.html

Dans la grande famille des URI, il y a (entre autres) les URL, et les URN, comme urn:ietf:params:xml:ns:iodef-2 normalisés dans ce RFC, qui couvre aussi bien leur syntaxe (qui était auparavant dans le RFC 2141 [1]) que les procédures d'enregistrement (autrefois dans le RFC 3406). Il remplace donc ces deux anciens RFC.

Le RFC 3986, qui normalise la syntaxe générique des URI, délègue les détails des familles particulières d'URI à d'autres RFC comme celui-ci. Notre RFC 8141 précise la syntaxe générique pour le cas des URN, des URI dont les propriétés sont a priori la persistence et la non-résolvabilité (donc, plutôt des noms que des adresses, pour reprendre le vocabulaire des RFC 1737 et RFC 3305).

La section 2, "*URN Syntax*" décrit à quoi ressemblent les URN. Un URN est formé avec le plan urn (enregistré à l'IANA <https://www.iana.org/assignments/uri-schemes/uri-schemes.xml>), un NID ("*Namespace IDentifier*") qui indique l'organisme qui gère la fin de l'URN, puis le NSS ("*Namespace Specific String*"), tous séparés par des deux-points. Il existe de très nombreux NID déjà enregistrés, comme ceux du RFC 7207 (messagerie Eurosystem), RFC 7254 (IMEI), RFC 5165 (normes OGC), RFC 9562 (UUID)...Deux idées essentielles des URN sont que la création de NID est strictement gérée (il faut documenter précisément le pourquoi du nouvel NID) et que, dans chaque espace de noms créé par un NID, l'affectation des NSS est à son tour gérée, avec des règles rigoureuses, que l'on suit soigneusement. Les URN sont conçus pour du nommage « sérieux ». Il n'existe pas d'URN à enregistrement libre et, donc, le fait d'être correct syntaxiquement ne suffit pas pour qu'un URN soit un vrai URN : il doit en plus être passé par le processus d'enregistrement.

Par exemple, si on prend les URN néo-zélandais du RFC 4350, le NID est nzl et un URN ressemble donc à urn:nzl:govt:registering:recreational_fishing:form:1-0. Attention, le deux-points n'est un séparateur qu'au début, entre le plan et le NID, puis entre le NID et le NSS. Après,

1. Pour voir le RFC de numéro NNN, https://www.ietf.org/rfc/rfcNNN.txt, par exemple https://www.ietf.org/rfc/rfc2141.txt

c'est juste un caractère comme un autre. Donc, `govt:registering:recreational_fishing:form:1-0` n'est pas forcément du nommage arborescent. Notez que la syntaxe est moins restrictive qu'elle ne l'était dans le RFC 2141.

Comme avec les autres URI, les caractères considérés comme « spéciaux » doivent être protégés avec l'encodage pour cent (cela ne concerne pas le NID qui doit être purement en ASCII). On ne peut donc hélas pas faire de vrais URN internationaux. Ainsi, `urn:example:café` doit en fait s'écrire `urn:example:caf%E9`.

Notre RFC 8141 introduit un concept nouveau, qui n'était pas dans le RFC 2141, celui de composants. Ils sont de trois sortes, les "*r-components*", les "*q-components*" et les "*f-components*". Les premiers, les "*r-components*", servent à passer des paramètres à un éventuel système de résolution. Ils commencent par un point d'interrogation suivi d'un plus. Ainsi, dans `urn:example:bar-baz-qux?+CCResolve:cc=fr`, le "*r-component*" est `CCResolve:cc=fr`, indiquant probablement qu'on souhaite une réponse adaptée à la France ("*CCResolve*" = "*Country-Code Resolve*"). La syntaxe est définie (ainsi que le fait que ces composants doivent être ignorés pour la comparaison de deux URN) mais la sémantique est pour l'instant laissée dans le flou.

Contrairement au "*r-component*", prévu pour le système de résolution, le "*q-component*" est prévu pour la ressource à laquelle on accède. Il commence par un point d'interrogation suivi d'un égal. Un exemple pourrait être `urn:example:weather?=op=map&lat=39.56&lon=-104.85` (cherchez pas : c'est près de Denver).

Quant au "*f-component*", introduit par un croisillon, il n'est destiné qu'au client (comme le classique identificateur de fragment des URI).

La section 3, consacrée à l'équivalence **lexicale** de deux URN, explique comment on peut déterminer si deux URN sont égaux ou pas, sans connaitre les règles particulières de l'organisme qui les enregistre. (Déterminer l'équivalence sert, par exemple, pour savoir si un URN a déjà été visité.) Ainsi, `urn:foo:BAR` et `urn:FOO:BAR` sont lexicalement équivalents (le NID est insensible à la casse, cf. section 2.1) mais `urn:foo:BAR` et `urn:foo:bar` ne le sont pas, le NSS étant, lui, sensible à la casse (les deux URN sont peut-être fonctionnellement équivalents mais cela dépend de la politique d'enregistrement de l'organisme désigné par `foo`). Il n'y a pas d'autre normalisation appliquée avant la comparaison, notamment sur les caractères encodés pour-cent.

Notez que la définition d'un espace de noms donné peut toujours rajouter des règles (par exemple que le NSS soit insensible à la casse) mais ces règles doivent uniquement créer de nouvelles équivalences, jamais séparer deux URN qui auraient été identiques dans une comparaison générique (par un logiciel qui ne connaissait pas les règles spécifiques de cet espace de noms, voir aussi la section 4.2).

Comme vu plus haut, il y a dans un URN, immédiatement après la chaîne de caractères `urn:`, l'espace de noms ("*namespace*" ou NID), une chaîne de caractères qui identifie le domaine d'une autorité d'enregistrement. Notre RFC 8141 explique les procédures de création d'un nouvel espace de noms dans le registre des espaces de noms `<https://www.iana.org/assignments/urn-namespaces>` que tient l'IANA. Si vous voulez juste faire des exemples, il existe un NID `example` specifié dans le RFC 6963 (oui, j'aurais dû l'utiliser au lieu de `foo`).

Comme expliqué dans la section 5, ce mécanisme d'espaces de noms suppose que, dans chaque espace, il existe une autorité d'enregistrement qui accepte (ou refuse) les enregistrements et que, d'autre part, il existe une autorité qui enregistre les espaces de noms (en l'occurrence l'IANA). La section 6 de notre RFC est consacrée aux procédures de cette dernière autorité et aux mécanismes pour enregistrer

un identificateur d'espace de noms (NID pour "*namespace identifier*"). (La **résolution** des URN en autres identificateurs n'est par contre pas couverte, mais on peut toujours regarder le RFC 2483.) Des exemples d'autorité d'enregistrement dans un espace de noms donné sont le gouvernement néo-zélandais (RFC 4350) ou l'OGC (RFC 5165).

Notez que certains URN sont créés en partant de rien, alors que d'autres sont juste une transcription en syntaxe URN d'identificateurs déjà enregistrés par ailleurs. C'est le cas des ISBN (RFC 8254) ou des RFC eux-mêmes (RFC 2648, avec le NID `ietf`, ce RFC est donc `urn:ietf:rfc:8141`).

Tiens, et où peut-on mettre des URN ? Syntaxiquement, un URN est un URI, donc on peut en mettre partout où on peut mettre des URI (section 4 du RFC). Par exemple, comme nom d'espace de noms XML. Mais ce n'est pas une bonne idée de mettre des URN partout. Par exemple, en argument d'un `<a href="...`, c'est même certainement une mauvaise idée (car le navigateur Web typique ne sait pas résoudre des URN).

Autre recommandation pratique de notre RFC, si un logiciel affiche un URN, bien penser à l'afficher en complet (contrairement à ce que font certains navigateurs Web qui, stupidement, tronquent l'URI, par exemple en omettant le plan `http://`). En effet, une abréviation peut ne pas avoir le résultat attendu. En effet, il existe un NID `urn:xmpp` (RFC 4854) et un plan d'URI `xmpp:` (RFC 5122). Si on n'affiche pas le plan `urn:`, il y a un gros risque de confusion.

La section 5 du RFC détaille ce qu'est un espace de noms (un ensemble d'identificateurs uniques géré, c'est-à-dire que tous les noms syntaxiquements corrects n'en font pas partie, uniquement ceux qui ont été enregistrés). Par exemple, les ISBN forment un tel espace (dont l'utilisation dans des URN a fait l'objet du RFC 8254). À l'intérieur d'un espace de noms, les règles d'enregistrement et le travail quotidien du registre ne sont pas gérés par l'IETF ou l'IANA mais par l'autorité d'enregistrement de cet espace.

La section 5 introduit les différents types d'espaces de noms. Il y a les espaces informels (section 5.2), dont le NID commence par `urn-` et est composé de chiffres et les espaces formels (section 5.1) dont le NID est composé de lettres et qui, contrairement aux informels, sont censés fournir un bénéfice aux utilisateurs de l'Internet (les espaces informels ont le droit d'être réservés à une communauté déconnectée). Contrairement encore aux informels, l'enregistrement des espaces formels fait l'objet d'un examen par un expert (cf. RFC 5226) et il est recommandé que cet enregistrement fasse l'objet d'une spécification écrite, par exemple un RFC. L'organisation qui gère un NID formel doit également démontrer sa stabilité et son sérieux sur le long terme.

Un des principes des URN est la **durabilité** : un URN devrait être stable dans le temps (et, logiquement, jamais réaffecté si jamais il est supprimé). Mais cette stabilité dépend essentiellement de facteurs non-techniques, comme la permanence dans le temps du registre (une organisation privée et fermée comme l'IDF est, par exemple, typiquement un mauvais choix pour assurer la permanence). Toutefois, si on ne peut pas garantir la stabilité d'un espace de noms, on connait en revanche des facteurs qui diminuent la probabilité de permanence et l'IETF peut donc analyser les spécifications à la recherche de tels facteurs (c'est une variante du problème très riche mais bien connu de la stabilité des identificateurs). Au passage, la plupart des grandes questions liées aux URN (permanence, compromis entre facilité d'enregistrement et désir de ne pas permettre « n'importe quoi ») sont des questions bien plus anciennes que les URN, et même plus anciennes que l'Internet, et ne feront probablement pas l'objet d'un consensus de si tôt (cf. section 1.2).

Enfin, le processus d'enregistrement lui-même. Il faut en effet un peu de bureaucratie pour s'assurer que le NID est bien enregistré et que le registre des NID `<https://www.iana.org/assignments/`

urn-namespaces> soit lui-même stable. Les procédures sont différentes selon le type d'espace de noms. Les informels, qui commencent par la chaîne de caractères urn- suivie d'un nombre, ont leur propre registre <https://www.iana.org/assignments/urn-namespaces/urn-namespaces.xml#urn-namespaces-2>, avec un processus d'enregistrement léger, mais très peu utilisé.

Le gros morceau est constitué des espaces de noms formels. Cette fois, le processus d'enregistrement est plus complexe, mais on obtient un « vrai » NID comme mpeg (RFC 3614), oasis (RFC 3121) ou 3gpp (RFC 5279).

Le formulaire d'enregistrement complet est disponible dans l'annexe A du RFC. Bon courage aux futurs enregistreurs. N'oubliez pas de lire tout le RFC. Notez par exemple qu'il faudra décrire les mécanismes par lesquels vous allouerez des noms (si vous gérez urn:example, et que je demande urn:example:boycott-de-mcdo, que répondrez-vous ?), et qu'il faudra une analyse de sécurité et aussi (c'est une nouveauté par rapport au RFC 2141) de vie privée.

Ce nouveau RFC remplace à la fois le RFC 2141 et le RFC 3406, et le processus fut long (six ans) et difficile. L'annexe B indique les principaux changements depuis le RFC 2141 : alignement de la syntaxe sur le RFC 3986 (en pratique, la nouvelle syntaxe est plus acceptante), introduction des composants (pour pouvoir faire des choses comme les fragments, qui marchaient avec tous les autres URI)...L'annexe C, elle, présente les changements par rapport au RFC 3406 : politique d'enregistrement d'un NID plus libérale (examen par un expert au lieu d'examen par l'IETF), suppression des NID expérimentaux (nom commençant par x-) en raison du RFC 6648...