

RFC 8615 : Well-Known Uniform Resource Identifiers (URIs)

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 3 septembre 2019

Date de publication du RFC : Mai 2019

<https://www.bortzmeyer.org/8615.html>

Plusieurs normes du Web s'appuient sur l'existence d'un fichier à un endroit bien connu d'un site. Les deux exemples les plus connus sont `robots.txt` et `favicon.ico`. Autrefois, ces endroits « bien connus » étaient alloués sans schéma central. Depuis le RFC 5785¹, c'est mieux organisé, avec tous ces fichiers « sous » `/.well-known/`. Notre RFC remplace le RFC 5785 (et le RFC 8307), avec peu de changements significatifs.

Prenons l'exemple le plus connu, `robots.txt`, fichier stockant la politique d'autorisation des robots qui fouillent le Web. Si un robot examine le site Web `http://www.example.org/`, il va tenter de trouver ledit fichier en `http://www.example.org/robots.txt`. Même chose pour, par exemple, `sitemap.xml` ou P3P (section 1 du RFC). Ce système avait plusieurs inconvénients, notamment le risque de collision entre deux noms (puisque'il n'y avait pas de registre de ces noms) et, pire, le risque de collision entre un de ces noms et une ressource normale du site. D'où l'importance d'un « rangement » de ces ressources bien connues. Elles doivent dorénavant être préfixées de `/.well-known/`. Ainsi, si le protocole d'autorisation des robots était normalisé aujourd'hui, on récupérerait la politique d'autorisation en `http://www.example.org/.well-known/robots.txt`.

À noter que le RFC spécifie uniquement un **préfixe** pour le chemin de la ressource, `/.well-known/` n'est pas forcément un **répertoire** sur le disque du serveur (même si c'est une mise en œuvre possible).

Le RFC 8615 note aussi qu'il existe déjà des mécanismes de récupération de métadonnées par ressource (comme les en-têtes de HTTP ou les propriétés de WebDAV) mais que ces mécanismes sont perçus comme trop lourds pour remplacer la ressource unique située en un endroit bien connu.

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc5785.txt>

Le nom `.well-known` avait été choisi (cf. annexe A de notre RFC) car il avait peu de chances de rentrer en conflit avec un nom existant (traditionnellement, sur Unix, système d'exploitation le plus utilisé sur les serveurs Web, les fichiers dont le nom commencent par un point ne sont pas affichés).

Bref, passons à la section 3 qui donne les détails syntaxiques. Le préfixe est donc `/.well-known/`, les noms en « dessous » doivent être enregistrés (cf. section 5.1), et ils doivent se conformer à la production `segment-nz` du RFC 3986 (en clair, cela veut dire qu'ils doivent être une suite de caractères ASCII imprimables, avec quelques exclusions comme la barre oblique). Du point de vue sémantique, ils doivent être précis, pour éviter l'appropriation de termes génériques (par exemple, l'application Toto qui veut stocker ses métadonnées devrait utiliser `toto-metadata` et pas juste `metadata`.) À noter que l'effet d'une requête `GET /.well-known/` (tout court, sans nom de ressource après), est indéfini (sur mon blog, cela donne ça `</.well-known/>`; devrais-je le configurer pour renvoyer autre chose? Sur Mastodon, ça donne 404.)

Quelques conseils de sécurité pour le webmestre (section 4) : ces ressources « bien connues » s'appliquent à une origine (un « site Web ») entière, donc attention à contrôler qui peut les créer ou les modifier, et d'autre part, dans le contexte d'un navigateur Web, elles peuvent être modifiées par du contenu, par exemple JavaScript.

La section 5 décrit les conditions d'enregistrement des noms bien connus à l'IANA. Le registre `<https://www.iana.org/assignments/well-known-uris/well-known-uris.xml>` contient par exemple les métadonnées du RFC 6415. Y mettre des noms supplémentaires nécessite un examen par un expert et une description publiée (pas forcément un RFC). Dans les termes du RFC 8126, ce sera Spécification Nécessaire et Examen par un Expert. Il y a un mini-formulaire à remplir (section 3.1 du RFC) et hop, le nom bien connu sera enregistré. Plusieurs existent désormais `</search?pattern=rfc%26well%26known%26uri&default_connector=AND>`.

Notez qu'il est très difficile de savoir combien de sites ont des ressources `/.well-known`. Bien sûr, Google le sait, mais ne donne pas accès `<https://support.google.com/websearch/bin/answer.py?hl=en&answer=136861>` à cette information (une requête `inurl:/.well-known` ou `inurl:"/.well-known"` ignore hélas le point initial et trouve donc surtout des faux positifs). Si on n'a pas accès à la base de Google, il faudrait donc faire soi-même une mesure active avec un client HTTP qui aille visiter de nombreux sites.

Les changements depuis le RFC 5785 sont résumés dans l'annexe B du RFC :

- Les plans d'URI pour WebSocket, du RFC 8307, ont été intégrés,
- D'ailleurs, le préfixe `/.well-known/` n'est plus réservé au Web, il peut être utilisé pour d'autres plans d'URI, ce qui a modifié le registre des plans `<https://www.iana.org/assignments/uri-schemes/uri-schemes.xml#uri-schemes-1>` pour y ajouter une colonne indiquant s'ils permettent ce préfixe (section 5.2),
- Les instructions pour l'enregistrement d'un nouveau nom ont été légèrement assouplies,
- La section sur la sécurité est nettement plus détaillée.