

RFC 5785 : Defining Well-Known URIs

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 7 avril 2010. Dernière mise à jour le 7 octobre 2014

Date de publication du RFC : Avril 2010

<https://www.bortzmeyer.org/5785.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`. Jusqu'à présent, ces endroits « bien connus » étaient alloués sans schéma central. Notre RFC propose de mettre tous ces fichiers « sous » `/.well-known/` pour construire des URI. (Il a depuis été remplacé par un document plus récent, le RFC 8615¹.)

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 a plusieurs inconvénients, notamment le risque de collision entre deux noms (puisque'il n'y a pas de registre de ces noms) et, pire, le risque de collision entre un de ces noms et une ressource normale du site. Cela fait donc des années que plusieurs personnes réclamaient un « rangement » de ces ressources bien connues. C'est désormais fait : les ressources bien connues 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 5785 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/rfc8615.txt>

La section 1.1, par contre, met en garde contre un usage tous azimuts de `.well-known`. Comptenu de l'architecture du Web <http://www.w3.org/TR/2004/REC-webarch-20041215>, il ne s'agit pas d'un mécanisme général de récupération d'information, juste d'une optimisation quand, par exemple, il faut évaluer une politique **avant** d'accéder à une ressource.

Le nom `.well-known` a été choisi 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). Une recherche sur Google avait confirmé que ce nom était largement libre.

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). 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.)

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.xhtml> était vide au début mais a été ensuite rempli par exemple par les métadonnées du RFC 6415. Y mettre des noms supplémentaires nécessite un examen par un expert et une norme publiée (pas forcément un RFC). Dans les termes du RFC 5226, ce sera "*Specification Required*" et "*Expert Review*". Il y aura un mini-formulaire à remplir et hop, le nom bien connu sera enregistré. Plusieurs existent désormais `</search?pattern=rfc%26well%26known%26uri>`.

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.