

RFC 7669 : Assigning Digital Object Identifiers to RFCs

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 25 octobre 2015

Date de publication du RFC : Octobre 2015

<http://www.bortzmeyer.org/7669.html>

Depuis mai 2015, tous les RFC ont un DOI. C'est quoi et à quoi ça sert ? C'est ce qu'explique ce RFC avec lequel, je vous préviens, je ne suis pas du tout d'accord.

Les DOI sont un mécanisme d'identification formel de documents numériques. Par exemple, le RFC 7626¹ a désormais le DOI 10.17487/RFC7626. Les DOI sont désormais attribués à tous les RFC (et c'est rétroactif : le RFC 1 a le DOI 10.17487/RFC0001). Ils étaient prévus à l'origine pour être utilisé avec le système de résolution Handle (RFC 3650). Celui-ci s'étant complètement cassé la figure (et à juste titre), les DOI sont aujourd'hui résolus via un URI (par exemple, pour le RFC 7676, l'URI est).

Ce RFC explique que les DOI sont largement utilisés (y compris, désormais, par les RFC) mais oublie de dire que c'est souvent par obligation, par exemple parce que certaines bibliothèques numériques anglo-saxonnes refusent les documents qui n'ont pas de DOI.

Les DOI sont techniquement décrits par une norme ISO, la ISO 26324 :2012 <http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=43506>. Comme presque toutes les normes du dinosaure ISO, elle n'est pas en accès libre. Si vous tenez à la lire, il vous en coûtera 88 francs suisses, et vous n'aurez pas le droit de la distribuer ensuite aux copains. (On note l'incohérence de l'IETF, qui avait signé une grandiose déclaration <<https://www.ietf.org/media/2012-08-29-open-stand.html>> en faveur des SDO ouvertes.) Les DOI sont ensuite attribués de manière hiérarchique, l'IDF accréditant des agences qui attribuent ensuite des DOI.

Pourquoi des DOI et pas un autre type d'identificateurs ? La section 1 du RFC cite un passage du « *Report on best practices for citability of data and on evolving roles in scholarly communication* » <<http://>

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc7626.txt>

www.stm-assoc.org/2012_07_10_STM_Research_Data_Group_Data_Citation_and_Evolving_Roles_ODE_Report.pdf » qui tape sur les URI en montrant une grande incompréhension du sujet. Le document en question reproche aux URI leur manque de permanence, en oubliant la règle numéro 1 de la permanence des identificateurs : **la permanence est une fonction de l'organisation, pas de la technique**. Autrement dit, un URI peut être stable, s'il est géré par une organisation sérieuse. C'est exactement pareil pour un DOI. Si l'IDF, organisation purement privée, disparaît, que deviendront les DOI ?

Notez aussi une particularité des RFC, qui peut être très intéressante pour leur trouver des identificateurs : les RFC sont immuables. Une fois publiés, ils ne sont jamais changés, même pas d'une virgule (le seul moyen de faire un changement est de publier un nouveau RFC). Résultat, on pourrait parfaitement désigner les RFC par un identificateur très stable, leur condensat, comme avec les NI du RFC 6920 (au passage, ce RFC 7669 a le NI `ni:///sha256;NAuLJduQ-wP7LBI5nYQw0-Ubxh8DsQdVVaxooj4oYB4`).

Un autre argument tout aussi erroné figure dans la section 1 : que les DOI permettent de trouver plus facilement la version gratuite des RFC en ligne. Il est vrai que des gens peu honnêtes ont vendu des versions payantes des RFC (ce qui est légal, les RFC sont libres d'utilisation) sans prévenir leurs acheteurs qu'ils pouvaient les obtenir gratuitement en ligne (ce qui est assez inéthique). Mais croire que les gens ont besoin des DOI pour trouver un RFC en ligne est assez fort de café ! (Cet argument est pourtant repris dans la section 3.)

La section 2 du RFC explique comment sont formés les DOI des RFC. Notez bien qu'un DOI est normalement opaque pour l'utilisateur : contrairement à un URI, il ne faut pas l'analyser pour en chercher des composantes. Néanmoins, si vous aimez savoir comment les choses sont faites, un DOI pour un RFC commence avec le préfixe `10.17487` (les DOI commencent toujours par 10, d'autres utilisations avaient été prévues avec d'autres nombres, mais n'ont jamais décollé), préfixe attribué aux RFC, et est suivi du texte `rfc` et du numéro du RFC. Le texte est apparemment insensible à la casse (en tout cas, les exemples du RFC sont tantôt en minuscules, tantôt en majuscules) mais, la norme n'étant pas librement accessible, je n'ai pas pu vérifier. Comme le DOI est opaque, il ne faut pas chercher à deviner un DOI pour un RFC donné. Par exemple, le numéro est actuellement formaté avec quatre chiffres mais cela pourra changer après le RFC 9999...

Comme indiqué plus haut, les vendeurs de DOI, après avoir répandu plein de FUD contre les URI, on finalement renoncé à leurs grandioses projets de déploiement de nouveaux protocoles et utilisent le même HTTP que tout le monde, avec les mêmes URI que tout le monde. Pour accéder à un DOI, on le préfixe par `https://dx.doi.org/`. On accède alors à une page Web fournie par l'entité qui a publié le document (et dont rien ne garantit donc la stabilité...) Cette page peut être le document recherché lui-même, une page de métadonnées avec un accès au document (c'est le cas des RFC) ou bien une page de métadonnées sans le document (cas des documents privés, non accessibles publiquement). Voir à ce sujet le « *DOI Handbook* » <<http://www.doi.org/hb.html>> »

Le processus d'attribution d'un DOI pour un RFC est décrit dans la section 4. Comme on l'a vu, il existe plusieurs agences d'attribution de DOI, dont plusieurs sont spécialisées dans un secteur très particulier. Le *"RFC Editor"* a choisi CrossRef, émanation des éditeurs. (Et qui n'est pas l'unique agence, contrairement à ce que dit le Wikipédia francophone.) Le coût d'adhésion à CrossRef est de 660 dollars et chaque DOI attribué à un document coûte un dollar de plus.

Le DOI du RFC est maintenant inclus dans les bases distribuées par le *"RFC Editor"*. Tous les RFC qui citent un autre RFC doivent maintenant inclure le DOI (une obligation imposée par l'IDF, cf. la section 4.4 du RFC.) C'est fait automatiquement quand on utilise les outils comme `xml2rfc` <<http://xml2rfc.ietf.org/>> (faut juste penser à vider le cache, en `/.cache/xml2rfc` d'abord, sinon certains RFC auront des DOI et pas d'autres).

L'attribution de DOI aux RFC parus précédemment a nécessité un peu de développement logiciel, utilisant l'API de Crossref pour solliciter les sept mille et quelques DOI des RFC existants.

Notez que tout cela s'est fait sans aucune concertation, ce qui est en contradiction avec les valeurs de l'IETF. Lorsque ce RFC a été soumis, la décision était déjà prise et mise en œuvre (l'annonce officielle date de mai 2015 <<https://www.ietf.org/mail-archive/web/ietf-announce/current/msg14125.html>>, avant, il fallait prêter attention à des appels d'offre du "*RFC Editor*" comme celui-ci <<http://iaoc.ietf.org/documents/Announce-RFC-DOI-SOW-3Sep14.pdf>>).

D'autres solutions auraient pourtant été possibles. D'abord, l'IETF elle-même a plein de mécanismes d'identificateurs, à commencer évidemment par les URI du RFC 3986. Au sein des URI, des plans particulier (comme NI, cité plus haut) étaient possibles. En dehors de l'IETF, des identificateurs plus ouverts comme ARK auraient été un choix mais il semble que les éditeurs scientifiques anglo-saxons ont décidé d'imposer le DOI partout.

Concevoir un système d'identificateurs est moins simple qu'il n'y paraît <<http://www.bortzmeyer.org/no-free-lunch.html>> et les DOI ont été conçus par des gens qui n'avaient pas de compétences particulières en ce domaine. Ici, le but était clairement de respectabilité plus que d'améliorer la vie des utilisateurs (comme cela avait été le cas pour l'ISSN <<http://www.bortzmeyer.org/identificateurs-normes-ie.html>>).