

PURL et URN : localisation et identification pérennes

Nous poursuivons notre découverte des moyens d'accès aux ressources du web via des URL. Dans ce troisième article, nous aborderons un dispositif, le PURL ou *Permanent URL*, qui tend à répondre au problème de l'absence de permanence des URL

Dès 1994 [1], Tim Berners-Lee précisait la relation entre l'URI (*Universal Resource Identifier*, devenu depuis *Uniform Resource Identifiers* ou identificateurs de ressource uniformes) et deux types particuliers de ceux-ci : les URL et les URN [2]. Une URN diffère d'une URL en proposant avant tout un nom univoque et permanent pour une ressource¹ ; ce qui signifie un identifiant indépendant du nom de domaine et donc de sa localisation. Mais, si cette URN est permanente, elle ne donne pas un accès immédiat à la ressource comme le font les URL via le protocole http. Un résolveur de noms est alors nécessaire². En dix ans, les URL, malgré leurs problèmes d'obsolescence, se sont fortement développées ; plus délicats à mettre en œuvre, les URN se mettent lentement en place.

Principe fonctionnel des PURL

En 1996, pour répondre au problème de « liens cassés » sur le web, l'OCLC propose un système simple, non propriétaire et gratuit : les PURL (*Permanent URL*) ou URL permanentes³ [3]. Ce système, fonctionnant en architecture distribuée, permet d'associer un nom de type PURL à l'URL d'une ressource existante et de retourner cette URL au client. Il s'agit de la redirection d'une URL existante et donc modifiable vers une URL virtuelle construite par le système et qui se veut durable. Ce système s'appuie sur un résolveur de liens, qui gère une table de correspondance entre des URL et des PURL.

La structure de cette URL virtuelle suit le schéma classique d'une URI : un protocole (ici http),

un espace de nom⁴ (ici le nom de domaine du résolveur proposé par l'OCLC : purl.org) et un segment attribué par l'utilisateur à la ressource. Exemple : `http://purl.org/net/francio = http://esperanto-panorama.net/francio`

Le nom donné au segment permet donc de distinguer localisation et identification des ressources. L'adresse PURL peut ainsi être très différente de son URL associée.

Même si la ressource est physiquement déplacée sur le réseau, elle conserve ainsi son identification. Il faudra toutefois veiller à modifier la table de correspondance si vous êtes administrateur ou à notifier le résolveur de liens de ce changement, la mise à jour n'étant pas automatique. Dans ce dispositif, l'administration de la table de correspondance revêt une importance capitale. En revanche, la PURL elle-même ne peut être modifiée. Elle est préservée et historisée. Ainsi est-il toujours possible de consulter l'historique des modifications d'une PURL.

L'OCLC propose un service en ligne. Il vous suffit de créer un compte pour pouvoir produire très rapidement une PURL. Dans cette configuration, vous bénéficiez d'un nom de domaine mutualisé /Net/, avec la possibilité de créer des sous-domaines personnalisés. Mais il est également possible de demander un domaine dit supérieur à la place du domaine NET, voire d'exploiter les applications mises à la disposition d'OCLC pour mettre en place vous-même un résolveur de liens⁵.

Des PURL pour quelles ressources ?

Un juste milieu doit être trouvé entre le coût d'administration de

cette table de correspondance de liens et l'intérêt de fournir un accès durable quelle que soit la vie de la ressource. Des ressources durables et fréquemment citées comme des périodiques électroniques, des articles, des contributions, des rapports et, bien sûr, des documents de référence sont des bons candidats à une PURL. Par exemple, l'espace de nom du Dublin Core possède une PURL qui pourrait être plus fréquemment utilisée⁶. La page principale d'un portail de ressources documentaires ou, de façon générale, les ressources au sommet d'une hiérarchie, sont également de bons candidats ; les objets de niveau inférieur n'ont pas besoin de PURL si leur organisation repose sur des bases pérennes. Mais le choix des ressources pour lesquelles cette mécanique sera mise en œuvre reste à étudier dans votre contexte.

Quels usages des PURL aujourd'hui ?

L'étude des chiffres fournis sur le site de l'OCLC en 2008 comparés à ceux de 2002 [4] montrent une évolution relativement faible des PURL avec 29% d'augmentation entre ces deux dates, soit 162 000 PURL nouvelles en six ans, ce qui paraît dérisoire face au volume du web dans son ensemble ! Par contre, le nombre de « Client Unique » (+1 461%) et surtout celui des PURL traduites (+ 524%) a considérablement augmenté. Peu de PURL, donc, mais des PURL bien exploitées.

On peut également citer le service POI⁷ (PURL-based Object Identifier), qui exploite le mécanisme des PURL conjointement avec le protocole OAI-PMH. En effet ce dernier a son propre système d'identifiant, non exploitable directement sur le web puisqu'il n'intègre pas de protocole normalisé connu par celui-ci. Le service POI permet d'exploiter l'identifiant d'une ressource d'un entrepôt OAI en l'intégrant dans une adresse Purl [5].

Depuis dix ans, le dispositif des PURL nous a permis de nous familiariser avec le principe des identi-

¹ « A URN differs from a URL in that it's primary purpose is persistent labeling of a resource with an identifier », 1997.

² Nous avons déjà vu le principe des résolveurs et des serveurs de noms avec l'OpenURL : cf. notre numéro 2/2008, p. 12-13.

³ À ne pas confondre avec *Personalized URL* ou PURL qui à une page personnalisée associe une URL elle-même personnalisée. Voir : www.abc-netmarketing.com/Les-URL-personnalisees-ou-PURL.html

⁴ Un espace de noms est un réservoir abstrait apportant un contexte précis aux éléments qu'il contient et désambiguïsant ainsi des éléments possédant le même identifiant.

fiants pérennes et des résolveurs de liens. Les promoteurs de ce système avaient bien indiqué à l'époque, tout en travaillant au sein du dispositif des URN, que celui des PURL se mettait en place en attendant que le fonctionnement des URN soit rodé.

Déployer des URN

Nous pouvons constater au jourd'hui l'existence de schémas d'URI (a) ou de sous-espaces au sein de l'espace des URN (b)⁸. Ceux-ci ne résolvent pas la question de l'accès à la ressource, mais celle de la pérennité des identifiants. Nous ne citerons que deux exemples.

(a) Un schéma d'adresses appelé *URI info* est proposé depuis 2006 afin d'identifier des ressources d'information qui possèdent des identificateurs dans des systèmes d'enregistrement publics existants. Par exemple, une adresse URI pour les numéros de contrôle de la bibliothèque du Congrès LCCN : info:lccn/2002022641, ou pour la classification décimale Dewey : info:ddc/22/eng/004.678 [6].

(b) Parlant d'identifiant pérenne, on songe immédiatement aux numéros ISBN, ISAN ou autres. L'URN possède effectivement des espaces de noms spécifiques pour l'ISBN ou l'ISSN depuis 2001 ou pour l'ISAN depuis 2006 : urn:isbn:0-395-36341-1 ; urn:issn:0012-4508 ; urn:isan:0123-1230-3210-2310-1

Il est possible d'initier d'autres espaces de nom et ainsi de démarquer l'attribution d'URN à des ressources, comme l'a fait la Bibliothèque nationale suisse (urn:nbn:ch) en utilisant la recommandation 3187 du RFC⁹ et en se dotant de politiques d'attribution clairement énoncées¹⁰.

Bonnes pratiques de production d'URL

Étude de la production documentaire autour des URI et des différentes recommandations ou notes rédigées par le W3C montre que celui-ci souhaite voir se déployer de bonnes pratiques autour des URI plutôt que des sys-

Références

- [1] Tim Berners Lee. *Universal Resource Identifiers in WWW*. June 1994. www.w3.org/Addressing/rfc1630.txt
- [2] W3C. *Naming and Addressing: URIs, URLs,...* 1997. www.w3.org/Addressing (Vue d'ensemble des matériaux du W3C sur les questions touchant à l'identification et à l'adressage)
- [3] Site de PURL : <http://purl.org>
- [4] Étienne Hustache. « Les URL permanentes : un moyen de pérenniser les adresses web ». *Bulletin des bibliothèques de France*, 2002, n° 2, p. 94-97. <http://bbf.enssib.fr>
- [5] Emmanuelle Bermès. *Les identifiants pour les objets numériques*. 20 septembre 2007. www.aristote.asso.fr/PIN/presentations/2007/PIN-identifiants.pdf
- [6] *Adresses URI sympas pour le Web sémantique*. Note de groupe d'intérêt du W3C du 31 mars 2008. www.yoyodesign.org/doc/w3c/cooluris. Trad. de JJ Solaris du document anglais du W3C : www.w3.org/TR/cooluris
- [7] Tim Berners Lee. *Les URLs sympas ne changent pas*. 1998. www.la-grange.net/w3c/Style/URI. Trad. de Karl Dubost de l'original en anglais : www.w3.org/Provider/Style/URI

tèmes de nommage particuliers qui, certes, remplissent un besoin mais qui apparaissent aussi comme des artifices et fragilisent d'autant l'architecture du web : le problème des liens cassés est reporté vers les (très nombreux) résolveurs de liens et systèmes d'enregistrement.

Pour revenir aux PURL, notons que celles-ci ne sont pas des URN (leur syntaxe ne suit pas la syntaxe normalisée des URN et leur finalité est bien la localisation), mais bien des URL. Elles ne sont donc pérennes que dans la mesure où le nom de domaine est maintenu et où l'information est administrée. Ce sont des URL accessibles indirectement via un résolveur d'URL.

Mais alors pourquoi ne pas écrire dès le départ une adresse pérenne ? C'est avec ce souci que le W3C fournit des recommandations de bonnes pratiques pour la préservation des URL dans le temps : ce sont les « *URL sympas* » de Tim Berners-Lee [7]. Parmi ces conseils : - ne pas intégrer des données reflétant le mode de gestion informatisée (GCI ou autre extension de programme). Il est même recommandé de supprimer les extensions ; - ne pas confondre l'URI et l'emplacement physique du fichier dans un répertoire sur un serveur, ces serveurs proposant de nombreux outils pour autoriser une plus grande souplesse dans cette

relation. N'oublions pas que l'espace URI est un espace abstrait ; - constituer l'organisation plutôt comme un index avec une structuration qui s'appuie, par exemple, sur l'année de production des ressources : cette année-là ne changera pas.

Des remarques de bon sens qui n'auront pas échappé aux professionnels de l'info-doc. En effet, un débat du même ordre ne se pose-t-il pas quand nous devons mettre en place des systèmes de classement ou de numérotation ? Si nous intégrons à notre modèle beaucoup d'informations codées : support, sujet ou éléments d'un organigramme, nombre d'entre elles risquent de s'avérer obsolètes dans un temps parfois très court.

En attendant la généralisation des bonnes pratiques qu'évoque Tim Berners-Lee concernant les URL et qui consistent à préserver la simplicité de désignation de vos URL, il est encore possible de s'appuyer sur le dispositif PURL, certes transitoire, mais simple et accessible. Et, pour répondre au principe de distinction entre le système de localisation de celui de la pérennisation des ressources, lançons-nous rapidement dans le monde des URN ! •

Sylvie Dalbin,
sylvieatd@aol.com
Odile Giraud,
ogiraud@docandco.fr

5 Voir le dispositif mis en place par le gouvernement australien : <http://purl.nla.gov.au>
6 <http://dublincore.org/2008/01/14/dcterms.rdf> = <http://purl.org/dc/terms>
7 www.ukoln.ac.uk/distributed-systems/poi
8 Site portail des Recommandations pour l'Internet : www.faqs.org/rfcs
9 *Using International Standard Book Numbers as Uniform Resource Names*.
10 Rubrique d'information sur les URN sur le site de la BNS : www.nb.admin.ch/sib/sib_professionnel/01693/01695/01706/index.html?lang=fr