

# Nommer les choses : les identificateurs

Stéphane Bortzmeyer, AFNIC

Collège International, 31 mai 2006

# Qu'est-ce qu'un identificateur ?

Une série de caractères qui désigne une ressource

Exemples :

1. URL : `http://www.bortzmeyer.org/wikipedia.html`
2. Numéro de téléphone : +33 1 39 30 83 00
3. ISBN : 2-86746-414-5
4. DCC (classification Dewey des catégories) : 641.57
5. ...

# Caractéristiques des identificateurs

- ▶ Syntaxe (un ISBN peut s'écrire de plusieurs façons),
- ▶ Unique ou non,
- ▶ Hiérarchique ou plat,
- ▶ Résolvable ou pas,
- ▶ Lié à une localisation ou pas,
- ▶ Alloué de manière distribuée, hiérarchique ou centralisée,
- ▶ Attribué de manière « ouverte » ou pas,
- ▶ Permanent ou pas.

# Unicité

## Une bonne chose ?

L'identificateur peut être unique ou pas. L'unicité est souhaitable mais elle a un coût.

## Les conséquences

- ▶ S'il n'y a pas d'unicité, un identificateur devient ambigu : qui va recevoir un message envoyé à `president@afnic.fr` ?
- ▶ L'unicité **impose** un registre unique ou, si l'identificateur est hiérarchique, une **racine**.

Quasi-unicité : le tirage au sort dans un espace très vaste peut permettre des identificateurs quasi-uniqes sans registre, de manière distribuée.

# Identificateurs non uniques

Pouvant donc se passer de **registre** :

- ▶ Les clés Freenet (<http://www.freenetproject.org/>), un réseau réellement à 100 % *peer to peer*,
- ▶ Les mots-clés qu'on tape dans un moteur de recherche,
- ▶ Les *tags* de [del.icio.us](http://del.icio.us),
- ▶ Les adresses IPv6 locales du RFC 4193<sup>1</sup> comme la “mienne”, `fd14:6941:e887::/48`. Elles sont en fait quasi-uniqes.

---

<sup>1</sup>Pour voir le RFC de numéro NNN,

<http://www.ietf.org/rfc/rfcNNN.txt>, par exemple

<http://www.ietf.org/rfc/rfc4193.txt>

# Identificateurs uniques

- ▶ Les noms de domaine,
- ▶ Les URL,
- ▶ Les numéros de téléphone,
- ▶ Les ISBN,
- ▶ ...

# Hiérarchie

Il s'agit ici de la hiérarchie de l'identificateur, pas de celle de son allocation.

En pratique, les deux sont souvent liées.

Assurer l'unicité dans un espace plat est très difficile (pas de passage à l'échelle). D'où le succès des identificateurs hiérarchiques : noms de domaine, adresses IP, ...

Certains identificateurs peuvent se **résoudre** en d'autres informations (souvent un identificateur de plus bas niveau, lié à la localisation, et permettant le **routage**).

Exemples :

- ▶ Noms de domaine, grâce au DNS,
- ▶ DOI (Digital Object Identifier) grâce au DNS (Handle n'est plus utilisé) puis à un CGI,
- ▶ Numéros de téléphone, grâce à SS7 ou bien à ENUM,
- ▶ Adresses IP grâce à ARP ou ND,
- ▶ UCC (identificateurs utilisés par RFID) grâce à ONS.

# Exemples d'identificateurs non résolubles

- ▶ Les *tags* du RFC 4151,
- ▶ Les numéros INSEE des personnes en France,
- ▶ Les ISBN,
- ▶ Les UCC, tant que ONS n'est pas déployé.

On peut toujours transformer un identificateur non résolvable en identificateur résolvable, grâce à un **moteur de recherche**. S'il y a un registre existant, c'est même trivial.

# Localisation

L'identificateur peut être complètement abstrait ou bien il peut dépendre de la localisation de la ressource.

Par exemple, les numéros de téléphone étaient autrefois très liés à une localisation.

Autre exemple, les adresses IP sont très liées à un opérateur. Ces identificateurs sont rigides et ne permettent pas la portabilité.

Mais ils sont efficaces et permettent le routage !

# Attribution

Qui attribue les identificateurs ? Est-ce centralisé, hiérarchique ou bien distribué ?

Un organisme « ouvert » ou pas ? À quelles conditions légales ou financières ? Et si le registre disparaît ?

# Permanence

Il est très souhaitable d'avoir des identificateurs stables dans le temps.

C'est même impératif pour certaines applications (*e-government*), par exemple.

## Ne pas se tromper de technique

La permanence dépend plus de considérations sociales que techniques (registres fiables, ne pouvant pas faire faillite, etc).

Par exemple, beaucoup d'URL ne sont pas stables parce que le webmestre est négligent et/ou incompetent. Mais ce n'est pas une propriété des URL, juste une conséquence d'une mauvaise gestion.

Exemples d'identificateurs peu stables :

- ▶ Adresses IP.

# Exemples d'identificateurs plus stables

Si on les gère correctement !

- ▶ Numéros de téléphone (si la « portabilité » marche bien),
- ▶ Noms de domaine (cela peut dépendre du registre, par exemple un “fr” peut vous être retiré - article 8 de la charte, un “com” peut être *reverse-cybersquatté* lors d'une UDRP).

“La vie, la santé, l'amour sont précaires, pourquoi les identificateurs échapperaient-ils à cette loi ?” (d'après Laurence Parisot).

Exemples d'identificateurs sans doute très stables :

- ▶ URI de type *tag* (RFC 4151),
- ▶ URN avec un NID attribué à un registre stable (cas des URN néo-zélandais du RFC 4350).

# La grande famille des URI

- ▶ URI *Uniform Resource Identifier* : toute la famille, donc inclus les URI, les URN, etc,
- ▶ URL *Uniform Resource Locator* : de loin l'identificateur le plus répandu sur Internet, dépend du nom de domaine. Pour les cas où la résolvabilité est essentielle,
- ▶ URN *Uniform Resource Name* : typiquement non résolvable. Très à la mode en *e-government*.

# Quelques identificateurs célèbres

Nous allons maintenant passer en revue des familles d'identificateurs célèbres, avec leurs caractéristiques :

- ▶ Unicité,
- ▶ Résolvabilité,
- ▶ Routabilité,
- ▶ Permanence,
- ▶ Allocation distribuée ou pas.

Les identificateurs utilisés dans les code-barres et dans RFID, attribués par le registre GS1.

- ▶ Unique (allocation hiérarchique).
- ▶ Résolvable uniquement si on utilise ONS (*Object Naming System*, une technique bâtie sur le DNS qui transforme un GTIN en nom de domaine).

Comme `www.example.org`.

- ▶ Résolvable grâce au DNS,
- ▶ Semi-permanent (mais c'est très dépendant de la politique du registre, de l'honnêteté et de la compétence du *registrar*, d'une bonne gestion par le titulaire. . .)
- ▶ Unique (grâce au RFC 2826, *IAB Technical Comment on the Unique DNS Root*).

# URI *tag*

Définis dans le RFC 4151.

Recommandés par exemple par la norme de syndication Atom.

Exemple (une entrée d'un *blog*) : `tag:`

`bortzmeyer.org,2006-02:Blog/dns-vulnerabilites.`

L'ajout de la date les rend stable par rapport aux variations du DNS mais empêche la résolvabilité.

- ▶ Non résolvables
- ▶ Permanents,
- ▶ Sans registre central. Dépend indirectement des noms de domaine mais pas du DNS.

# URN du gouvernement néo-zélandais

Définis dans le RFC 4350.

Par exemple `urn:nzl:govt:registering:  
recreational_fishing:form:1-0`.

- ▶ Non résolubles.
- ▶ Aussi permanents que l'État néo-zélandais.

*Host Identity Protocol* Défini dans le RFC 4423. Mais la spécification est loin d'être complète et aucun déploiement n'est en vue.

Vise à remplacer les adresses IP ou plutôt à les cantonner à un rôle très technique et largement invisible.

Une adresse IP sert à deux choses : désigner une machine et indiquer comment la joindre (routabilité).

Avec HIP, l'adresse IP ne serait plus qu'un identifiant "technique", ne servant qu'à joindre la machine. Son identité serait son *Host Identifier* (HI).

# Caractéristiques du HI

Le HI sera probablement une clé publique cryptographique.

- ▶ Résolvable,
- ▶ Permanent,
- ▶ Allocation hiérarchique par PKI ou distribuée par tirage au sort.

# En guise de conclusion

Il faut bien choisir son type d'identificateur

Pour cela, il faut formaliser son cahier des charges :  
permanence, routabilité, prix ?

Le contrôle des identificateurs est important

Qui contrôle le registre contrôle le monde... C'est un sujet politique et qui n'a rien de "purement technique".