RFC 7485 : Inventory and Analysis of WHOIS Registration Objects

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 9 avril 2015

Date de publication du RFC : Mars 2015

https://www.bortzmeyer.org/7485.html

Dans le cadre de la mise au point du protocole RDAP <https://www.bortzmeyer.org/weirds-rdap.html> d'accès aux informations sur des objets enregistrés, le groupe de travail à l'IETF s'était livré à une intéressante étude sur les objets enregistrés et accessibles via l'ancien protocole whois. Cette étude a servi à mieux définir ce qu'il fallait comme services dans RDAP, et à spécifier le format de sortie de RDAP, normalisé dans le RFC 7483[1]. Elle a montré, sans surprises, une grande variabilité : si beaucoup d'éléments d'information sont communs à un grand nombre de registres, c'est souvent sous des noms différents. Cette variabilité est, selon les goûts, un des charmes ou un des principaux défauts des registres de l'Internet.

L'étude qui a servi de base à ce RFC date de 2012 (section 3 de notre RFC, pour la méthodologie). Les données ont été récoltées via whois mais aussi via les interfaces Web d'interrogation des registres (à la fois registres de noms de domaine et registres d'adresses IP). Pour les TLD, le nom `nic.$TLD` a été utilisé, 106 ccTLD ont permis de récupérer de l'information sur ce nom, ainsi que 18 gTLD. Parfois, l'information a été vérifiée en essayant d'autres noms que `nic.$TLD`. Les capacités des serveurs whois interrogés étaient très variées (cf. section 5.1). Par exemple, si tous permettent évidemment des requêtes au sujet d'un domaine donné, certains permettent aussi des requêtes pour un contact précis ou pour un BE particulier. Pour les registres d'adresses IP, les cinq RIR ont été interrogés.

La section 4 présente les résultats pour les RIR. On y voit déjà la variété des étiquettes qui fait la beauté de whois. Ainsi, le nom de l'organisation titulaire d'un préfixe IP est étiqueté `organisation` à AfriNIC, `Owner` à LACNIC et `org-name` au RIPE-NCC. La date d'enregistrement d'un contact est `changed` à AfriNIC, `RegDate` à ARIN et `created` à LACNIC. Un préfixe IP se nomme `inetnum` à l'APNIC ou au RIPE-NCC, mais `NetRange` ou `CIDR` à l'ARIN (selon qu'il est IPv4 ou IPv6 !) Voici un exemple à l'APNIC :

1. Pour voir le RFC de numéro NNN, `https://www.ietf.org/rfc/rfcNNN.txt`, par exemple `https://www.ietf.org/rfc/rfc7483.txt`

```
inetnum:        1.2.2.0 - 1.2.2.255
netname:        KNET
descr:          KNET Techonlogy (BeiJing) Co.,Ltd.
descr:          4,South 4th treet, Zhongguancun,Haidian District,Beijing
admin-c:        ZX2975-AP
tech-c:         WL1818-AP
country:        CN
mnt-by:         MAINT-CNNIC-AP
mnt-irt:        IRT-CNNIC-CN
mnt-routes:     MAINT-CNNIC-AP
status:         ALLOCATED PORTABLE
changed:        ipas@cnnic.cn 20150107
```

Certains éléments ne sont pas présents dans tous les RIR et ceux qui le sont reçoivent, comme on vient de le voir, des noms différents.

Et pour les registres de noms de domaine ? C'est évidemment encore plus varié (au fait, les résultats bruts de la collecte faite pendant l'étude sont disponibles en ligne <https://docs.google.com/open?id=0B96TtoK8a--MTTRuVUt3UHZMdEk>). Au total, 68 éléments différents ont été identifiés, sous 550 étiquettes distinctes. Les auteurs du RFC les ont classé en éléments publics et autres éléments. Les éléments publics sont ceux qui sont mentionnés dans le document ICANN, « "*gTLD Applicant Guidebook*" <http://newgtlds.icann.org/en/applicants/agb/guidebook-full-04jun12-en.pdf> » de 2012, ou bien dans les RFC sur EPP, RFC 5730, RFC 5731, RFC 5732 ou RFC 5733 (un choix très contestable qui donne une sorte de privilège aux règles des TLD régulés par le gouvernement états-unien).

Parmi les exemples d'éléments publics, on peut trouver évidemment le nom de domaine (l'étiquette la plus fréquente étant `domain name` mais cinq autres ont été trouvées) suivi de la date de création (en général `created` mais vingt-trois autres étiquettes sont possible, la deuxième plus grande variété de vocabulaire). 95 % des registres ont donc un élément « nom de domaine » (on peut se demander comment font les 5 % restants...), 85 % une date de création, 77 % un statut du domaine, etc. Par contre, « dernier transfert », le moins fréquent, n'est présent que chez 3 % des registres.

On retrouve la même variété pour les titulaires et les contacts. Aucun élément n'est présent dans la majorité des registres (même pas l'adresse de courrier électronique). Ceux qui le sont ont des noms très variés. Ainsi, pour le contact technique, le téléphone dudit contact est décrit par pas moins de dix étiquettes différentes.

Le cas des serveurs de noms est plus compliqué car, contrairement aux contacts, le serveur de noms n'est pas forcément un objet de première classe dans le modèle de données du registre : il peut être un simple attribut du domaine. C'est pour cela que, dans le monde EPP, le RFC 5732 n'est pas mis en œuvre partout. Ainsi, alors que 92 % des registres indiquent les serveurs de noms dans les réponses (sous 63 étiquettes distinctes, un record, en partie dû à des mécanismes de nommage comme `nameserver N` avec N indiquant le rang du serveur), tous ne permettent pas des requêtes whois directes sur un serveur de noms.

Enfin, il y a les éléments « divers », ceux qui ne sont pas spécifiés dans les règles ICANN ou dans les RFC sur EPP et qui incluent, par exemple, l'URL d'un site Web associé au domaine, un champ de commentaires (`remarks`), une date d'anniversaire (six TLD sont dans ce cas, probablement tous à l'AFNIC), un identificateur de marque déposée, etc. La section 6 de notre RFC suggère que des extensions au modèle de données de RDAP pourraient s'inspirer de ces éléments divers, par exemple le point de contact du NOC (pourquoi diable ne pas se contenter du contact technique ?), le fait que l'identité du titulaire soit masquée (ce qui est parfois noté, bien à tort, `anonymous`), etc.

Il n'y a pas que le modèle de données (et la terminologie associée) qui varie beaucoup d'un registre à l'autre, il y a aussi le format de présentation. Les deux les plus populaires sont clé :valeur et en bloc. Un exemple en clé :valeur est donné par `.org` :

```
Domain Name:TV5.ORG
Creation Date: 1995-09-29T04:00:00Z
Updated Date: 2013-09-30T14:24:25Z
Sponsoring Registrar:Gandi SAS (R42-LROR)
...
Registrant Name:Thomas Derobe
Registrant Organization:TV5 Monde
Registrant Street: 131 avenue de Wagram
```

Un exemple de bloc par le registre de `.za` :

```
Domain Name:
    sab.co.za

Registrant:
    The South African Breweries Limited

    Email: domains@sabmiller.com
    Tel: +27.118818414
    Fax: +27.118818136

Registrant's Address:
    P.O.Box 782178
    2196 Sandton Sandown
    ZA

Registrar:
    Ascio

Relevant Dates:
    Registration Date: 1997-05-06
    Renewal Date:      2015-05-06

Domain Status:
    Registered until renewal date
```

Mais il faut remarquer qu'on trouve de tout. Comment classer le résultat du whois de `.jp` ?

```
Domain Information:
a. [Domain Name]                YAMAHA.CO.JP
g. [Organization]               YAMAHA Corporation
l. [Organization Type]          Corporation
m. [Administrative Contact]     HN050JP
n. [Technical Contact]          TT9781JP
p. [Name Server]                ns1.dhs.jtidc.jp
p. [Name Server]                ns2.dhs.jtidc.jp
s. [Signing Key]
[State]                         Connected (2016/03/31)
[Registered Date]
[Connected Date]                2010/08/02
[Last Update]                   2015/04/01 01:12:19 (JST)
```

À noter également que onze registres envoient l'information autrement qu'en anglais dans l'alphabet latin (le RFC, comme beaucoup de documents écrits dans le milieu ICANN, confond langue et écriture et écrit « *"11 registries give local script responses. The WHOIS information of other registries are all represented in English."* »).

Bref, ce RFC illustre bien une des motivations principales du projet RDAP <`https://www.bortzmeyer.org/weirds-rdap.html`> : normaliser la sortie du serveur, de façon à faciliter les traitements automatiques.