

RFC 4290 : Suggested Practices for Registration of Internationalized Domain Names (IDN)

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 10 février 2006. Dernière mise à jour le 17 mars 2006

Date de publication du RFC : Décembre 2005

<https://www.bortzmeyer.org/4290.html>

Les IDN ont toujours suscité des passions et il n'était donc pas évident de normaliser des politiques d'enregistrement, d'autant plus que chaque langue a ses particularités. L'IETF n'a donc pas essayé et le premier RFC non technique sur les IDN n'est donc pas une norme.

Le RFC 3490¹ spécifiait la norme IDN, permettant de représenter des noms de machines écrits en Unicode et pas seulement en US-ASCII. Ce RFC 3490, purement technique, ne spécifiait pas les politiques d'enregistrement des registres de noms. Ce choix est tout à fait justifié : au contraire des choix techniques, qui doivent être uniformes, il n'y a aucune raison de normaliser ces politiques, qui dépendent des choix de chaque registre.

Comme exemple de politiques possibles, dans le cas d'une population francophone, on peut imaginer que `pere-noel.fx` et `père-noël.fx` puissent être délégués à des entités différentes, malgré le risque de confusion ou bien au contraire que on définisse un **lot** ("*bundle*") qui est l'ensemble des noms IDN équivalents (par exemple, `pere-noel.fx` est équivalent à `père-noël.fx`). Tous les membres du lot seraient alors délégués au même titulaire.

Plusieurs efforts ont été tentés pour discuter de ces politiques et, à défaut de normes, spécifier un cadre commun pour les exprimer. Cela a été fait, par exemple, dans la liste `idn-reg-policy` <<http://www.imc.org/idn-reg-policy/>>, aujourd'hui inactive. Parallèlement, certains registres asiatiques ont écrit le RFC 3743 qui décrit leur cadre politique.

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc3490.txt>

Dans le cadre de la liste idn-reg-policy, Paul Hoffman, un expert en internationalisation avait proposé un langage pour décrire des tables de caractères avec leurs variantes (par exemple, `ÿ` peut être considéré comme une variante de `y` et donc `pierre-louÿs.fx` serait équivalent à `pierre-louys.fx`).

Notre RFC est proche des dernières versions des propositions d'Hoffman. Il ne présente pas **une** politique d'enregistrement mais il suggère des pistes, pointe des problèmes potentiels et propose une syntaxe pour exprimer les tables.

À titre d'information, voici la table des caractères utilisés en français, dans cette syntaxe (utilisant la notation Unicode donc `U+00E7` est le "c cédille"). Le point de départ de cette table avait été un excellent article de Jacques André <<http://www.gutenberg.eu.org/pub/GUTenberg/publicationsPDF/25-andre.pdf>> :

```
# Variant table for the French language

# A sentence with every composed character:
# Ils se sont réveillés à Nîmes, très tôt, dans la même Citroën
# niçoise sentant l'aïoli où ils avaient joué de cette flûte aiguë la
# nuit de Pâques, près de l'Haÿ-les-Roses.

# a-z
# a and its variants : à â
# á (a acute) does not seem to exist in French: for further study (town names?)
# a with ring above (U+00E5) does not exist in french but in picard
# http://www.picard.free.fr/lgpic/a-e.htm
U+0061|U+00E0:U+00E2
U+0062

# c
U+0063|U+00E7
U+0064

# e and its variants : è é ê ë
U+0065|U+00E8:U+00E9:U+00EA:U+00EB
U+0066
U+0067
U+0068

# i and its variants : î ï
U+0069|U+00EE:U+00EF
U+006A
U+006B
U+006C
U+006D

# ñ (U+00F1) does not exist in french but in breton
# http://www.ouest-france.fr/dossiershtm/cours-de-breton/lecture.htm
# or in basque.
U+006E

# o and its variants : ô. ö (U+00F6) exists in French? No word uses
# it.
# ò (U+00F2) does not exist in french but in occitan
# http://occitanet.free.fr/fr/index.html
U+006F|U+00F4
U+0070
U+0071
U+0072
U+0073
U+0074

# u and its variants : ù û ü
# ú (u acute) does not seem to exist in French: for further study
U+0075|U+00F9:U+00FB:U+00FC
U+0076
U+0077
U+0078

# y. ÿ only in names? (L'Haÿ-les-Roses)
U+0079|U+00FF
```

```
U+007A
# 0-9
U+0030
U+0031
U+0032
U+0033
U+0034
U+0035
U+0036
U+0037
U+0038
U+0039
# - (hyphen)
U+002D
# Ligature oe: the variant is a *string*
U+0153|U+006F-U+0065
# Ligature ae (æ). Only in names like Lætitia?
U+00E6|U+0061-U+0065
```

À noter que le RFC prend ses distances avec l'ICANN, qui a décidé d'établir une table des langages à l'IANA, table pour laquelle l'ICANN n'a ni légitimité, ni compétence et que la plupart des registres ont refusé de remplir.