



Институт развития  
информационного общества

Institute of the  
Information Society

## **Net4D**

### **A coherent multi-lingual DNS system and network**

IIS Workshop, 16 June 2008  
Moscow , Russia

Francis MUGUET  
ENSTA ( France )  
Focal Point of the Linguistic Diversity Coalition  
at the Internet Governance Forum

*Legal Notice:* The points of views expressed in this presentation, as usual in any scholarly presentation, do not represent of the official view of any institution, coalition or entity.

### **INTRODUCTION**

Multilinguism is one of the key aspect of the new Information Society.

People have a right to be able to express oneself in their mother tongue.

Linguistic diversity is the key to peace, because it teaches understanding of other cultures.

Linguistic diversity is the key to creativity, because one thinks in a language, and different languages are leading to a richness of concepts.



## PUNYshment for Domain Names

For the [Internationalized domain names](#) (IDN), the [solution proposed](#) by ICANN is based on a [Puny Code](#) :

Punycode transforms a Unicode chain ( in general [UTF-8](#)) into an ASCII chain in a unique and reversible way. ASCII characters stay unchanged, and non-ASCII characters are represented by ASCII characters.

For example `académie-française.org` gives `xn--acadmie-franaise-npbla.org`.

`http://русский.idn.icann.org` gives `http://xn--h1acbxfam.idn.icann.org`

This approach appears as a patch. However often, patches as quick and easy fixes to a specific problem ( IDN for a web browser ), are often ending up into overly complicated and untractable developments, unable to provide general solutions ( Mail, file transfer, etc... ).

Unexpected problems such the [Funy Code](#) are now appearing.



## and PUNYshed from Mail ?

In order to send a mail to `secrétaire@académie-française.org`, or to `иван@русский.рф`, the problem becomes more complicated and is still not solved, because `secrétaire` and `иван` are encoded in [UTF-8](#) et `académie-française` and `русский.рф` are encoded in [Puny Code](#).

The protocol to try to solve this issue has not been yet [finalized](#) by ICANN. ICANN current approach is that the SMTP server has to be modified so that it is the SMTP server that carries the IDN query and the Unicode to [Puny Code](#) transformation.

It is uncertain at the present stage, within the ICANN scheme, if a user could use the full array of Unicode to express his/her linguistic difference in a specific IDN. For example, a person might not be able to use a Chinese name in a Cyrillic IDN, therefore severely limiting the rights of linguistic minorities.

One may therefore ask the simple question : Should it be possible to conceive an

homogeneous and coherent [UTF-8](#) system ?

In order to give a practical answer, one must analyze the current [DNS](#) software, instead of being blinded by political arm twisting over the control of the root databases.



## BIND : the Key Software

- The spotlight has been only the control of DNS root databases
- Left in the dark : The software tools to access the DNS databases
- The actual subnetwork of DNS servers is neither owned nor under contract with ICANN, the DNS servers are voluntarily maintained by users ( mostly ISPs, web hosting companies, some registrars,... ).
- almost all machines in this subnetwork are running the free software ( FreeBSD license ) called rather aptly [BIND](#) which is maintained by the [Internet Systems Consortium](#). ( ISC).
- BIND 9 is striving for a strict compliance with IETF standards, ie, with the [Request for Comments \(RFCs\)](#) established by the [Internet Engineering Task Force \(IETF\)](#) , but this is not yet fully achieved.
- There are few other available DNS server software ( see a [Comparison](#) on Wikipedia ) but most often they follow BIND features

## BIND the Key Software (II)

The ISC T-shirt is rather amusing :



## BIND, as a PUBLIC RESSOURCE

In fact, it is very fortunate that BIND allows to carry different resolving services related to different classes of network.

2.1.3 Resource Records : *The data associated with domain names are contained in resource records, or RRs. Records are divided into classes, each of which pertains to a type of network or software. Currently, there are classes for internets (any TCP/IP-based internet), networks based on the Chaosnet protocols, and networks that use Hesiod software. (Chaosnet is an old network of largely historic significance.) The internet class is by far the most popular. (We're not really sure if anyone still uses the Chaosnet class, and use of the Hesiod class is mostly confined to MIT.)*

This possibility has been mostly ignored except for the proposal made by John C Klensin for a new class that is not limited to ASCII from its initial definitions. This would have allowed to a cleaner Internationalized Domain Name system, instead of relying on the patch that constitutes Punycode. However, the seamless implementation of such a two class system, where records of a new class are used as remedies to the shortcomings of the class "IN" would have created technical difficulties. These problems should not occur when starting with only one class, conceived from the onset for internationalization.



Now it is interesting to mention the RFC 2929 :

**CLASS is a two octet unsigned integer containing one of the RR CLASS codes. See section 3.2.**

**DNS CLASSes have been little used but constitute another dimension of the DNS distributed database. In particular, there is no necessary relationship between the name space or root servers for one CLASS and those for another CLASS. The same name can have completely different meanings in different CLASSes although the label types are the same and the null label is usable only as root in every CLASS. However, as global networking and DNS have evolved, the IN, or Internet, CLASS has dominated DNS use.**

**There are two subcategories of DNS CLASSes: normal data containing classes and QCLASSes that are only meaningful in queries or updates.**

**The current CLASS assignments and considerations for future assignments are as follows:**

### **Decimal Hexadecimal**

- |          |   |
|----------|---|
| <b>0</b> | <b>0x0000</b> - assignment requires an IETF Standards Action.               |
| <b>1</b> | <b>0x0001</b> - Internet (IN).  |
| <b>2</b> | <b>0x0002</b> - available for assignment by IETF Consensus as a data CLASS. |

3        0x0003 - Chaos (CH) [Moon 1981].  
 4        0x0004 - Hesiod (HS) [Dyer 1987].  
  
 5 - 127    0x0005 - 0x007F - available for assignment by IETF Consensus as data  
           CLASes only.  
  
 128 - 253  0x0080 - 0x00FD - available for assignment by IETF Consensus as  
           QCLASes only.  
  
 254  0x00FE - QCLASS None [RFC 2136].  
 255  0x00FF - QCLASS Any [RFC 1035].  
 256 - 32767    0x0100 - 0x7FFF - assigned by IETF Consensus.  
  
 32768 - 65280    0x8000 - 0xFEFF - assigned based on Specification Required as defini  
           in [RFC 2434].  
  
 65280 - 65534    0xFF00 - 0xFFFE - Private Use.  
 65535  0xFFFF - can only be assigned by an IETF Standards Action.

This leaves the possibility of  $2^{16} = 65536 - 5$  ( taken by the IN, CH, HS, None, Any classes ) = 65531 classes ( among which 255 for private use ) that could be used to carry other DNS services, using BIND.



ICANN cannot, in good faith, object to the use of yet another class, since ICANN recommended in May 2001 this [approach](#) :

*Moreover, it should be noted that the original design of the DNS provides a facility for future extensions that accommodates the possibility of safely deploying multiple roots on the public Internet for experimental and other purposes. As noted in RFC 1034, the DNS includes a "class" tag on each resource record, which allows resource records of different classes to be distinguished even though they are commingled on the public Internet. For resource records within the standard root-server system, this class tag is set to "IN"; other values have been standardized for particular uses, including 255 possible values designated for "private use" that are particularly suited to experimentation.*

*As described in a recent proposal within the IETF, this "class" facility allows an alternative DNS namespace to be operated from different root servers in a manner that does not interfere with the stable operation of the existing authoritative root-server system. Those that have deployed alternative roots have not used a different class designation, however, choosing instead to have their resource records masquerade as emanating from the standard root, and creating the potential for disruption of other's operations.*

Another view it is that the actual subnetwork of DNS servers ( in fact a P2P network, before the term was coined ) should be able carry several DNS systems, in other words to "degrouper" the "lines" of this "common carrier" to introduce "competition".



## TECHNICAL IMPLEMENTATION ASPECTS

At the present time, only DNS look-up utilities ( such as [dig](#), [host](#) ) associated with [BIND](#) allow to query a DNS server with the field *class*. Current browsers do not allow that.

With [dig](#) one may query a DNS server while specifying the *class* field. The default query class (IN for internet) is overridden by the -c option. *class* is any valid class, such as HS for Hesiod records or CH for Chaosnet records.

From the point of view of users : what is of utmost importance, and what differentiates this proposal from the "alternate root servers" of ill repute, is that users do not have to specify a DNS server different from the one provided by his/her ISP.

1. her/his browser may query the DNS server with the field *class*.
2. the ISP DNS server is updated to a future version of BIND that may carry different *classes*.

In order to simplify the query by a browser and other programs ( mail, file transfer, etc.. ) the definition of an [URL](#) ( [URI](#) ) must be generalized and updated. For example : the domain wikipedia in the new gTLD .open with the *class* net4d could become

`http://4d%fr.wikipedia.open`

Another example, the cyrillic domain Москва in the cyrillic gTLD po in the cyrillic *class* Сеть , could be written as `http://СТ%Москва.po` .



## TECHNICAL IMPLEMENTATION ASPECTS (II)

One may notice that the problem of homographs and ensuing phishing and litigations are avoided. For example, it could be possible to ensure that in the the Cyrillic *class* Сеть, domain names are only written with Cyrillic characters. Mixing of Latin and Cyrillic is simply not allowed when registering a domain name in the Cyrillic *class* Сеть.

A browser may be configured by the user, so that his/her browser use by default, a certain *class*. corresponding to a certain language to avoid any confusion.

In order that the browser ( such as [Mozilla](#) ) may query the DNS server with the field *class*, the modification to the browser is relatively minor, and less complicated than the implementation of the Puny code that is TLD dependent ( [Mozilla IDN-enabled TLDs](#)). In its query the browser must include the field *class*. After this modification, there is no need of *plug in* or a specific client.

For old browsers, users may go through a portal, specific to each *class*.

From the side of a web server such as [Apache](#), in order to enable virtual hosting, modifications are also relatively minor.



## TECHNICAL IMPLEMENTATION ASPECTS (III)

Features available for the *class* IN, should be available in principle, in agreement with the RFCs. but RFCs compliance is not yet fully achieved in current version 9 of BIND. For all practical purposes, the difficulty would depend on the way how the BIND 9 code is written. While "parametrization" of existing subroutines should be preferred, a brute force method is to replicate for a new class, obfuscated subroutines that are "hard coded" with the *class* IN. ISC is in the process of rewriting BIND from scratch for version 10, in a modular way, that would greatly simplify the task. Modifications required in the free software [Mozilla](#) et [Apache](#) can be written quickly, and most importantly, included easily in the next official release. For proprietary software, this is going to depend on the good will of manufacturers, however considering the domination of [Apache](#) among web servers, and the importance of [Mozilla](#) among web browsers, one should expect that they are not going to afford to lag behind.

Last but not least, each *class* may use of its own distinct network of root servers.



## Net4D, networks to empower the second generation of the Web: the Semantic Web

- Net4D are another classes of Network, like Hesiod and Chaosnet, ICANN has no jurisdiction on this network, only on the class "IN".
- There could be other classes in competition with the ICANN IN class and the NET4D classes. Fair and ethical competition is welcome.
- Net4D classes are not designed to provide similar minimal services as ICANN, it has in mind to provide value added services, in view to empower the Semantic Web.
- Net4D domain holders should abide by a specific ontology, as a contractual requirement to the effect of :
- Establishing pollution free zone concerning metadata, and providing pathway for the interoperability of metadata concerning specific activities following the Semantic Web approach.
- Providing a Open Digital Resource Identifier system that is clearly needed for future evolution of the Web and to authenticate metadata
- Providing a Open Digital Resource Identifier (ODRI) that is P2P friendly, that is facilitating a maximal flow via P2P, therefore allowing sites with little bandwidth to exchange vast amount of data.



## Empowering the Semantic Web

Net4D are classes of Next Generation Domain Services that are empowering the Semantic Web.

Two main networks/services are for the moment being considered :

- [Web4D](#): The Network of People
- [Epc4D](#) : The Network of Things

Other possible SW gTLDs:

- equitable commerce global market place ( operated by UNCTAD )
- trademarks ( operated by WIPO )



## Semantic Web and Linguistic Dialog

As an example of a SW gTLDs : the [Linguistic SWgTLDs](#) or LSWgTLDs. An extension shall be assigned to each language so that sites or sites' versions written in specific languages can be easily found and identified. It would facilitate greatly the task of search engines and would foster linguistic diversity.

The main points of the breakthrough are the following :

1. Facilitate exchange between sites in different languages that share the same structure of meta-data, the same meta-language.
2. Help automatic translation.

Automatic translation would be much improved if automatic tools could work with several human certified translation of the same text. For example, if the same document has been available in English and in French by the authors on the same site, and translated by human users in Russian and Korean on other sites, it would be tremendous advantage for automatic translation tools to have access and make use of all existing versions in different languages of the same document. For example "Société Civile" would be translated in yet other languages such as Spanish as " Sociedad Civil" ( meaning Civil Society not Civil Company or Business ! ), with the help the correct English version. Of course, it is required that the translation tools could retrieve and identify the various versions at different locations, therefore the need follows for an identifier, as well as standardized metadata. SWgTLDs could be the keys to practical not just elite multilingualism on the Web.





## GOVERNANCE

- If the current DNS monopolistic situation no longer prevails. If a fair competition is introduced over all *classes*, then it is possible to let [ICANN](#) evolve to its own destiny, with its uncertain legal status, under to its historic preferential governmental parentage. This would avoid international political tensions.
- Concerning Net4D and other *classes* governance, it is suggested to consider a transparent, inclusive, multi-stakeholder partnership, including intergovernmental and governmental organizations, technical operators, businesses, academia, civil society, fully recognized within an international public law context, according to the [UNMSP](#) proposal.
- The role of the [W3C](#) that researches and develops, for the public good, open (non-proprietary) standards, protocols and languages for the [Semantic Web](#) should be recognized, and a substantial part of financial revenues, originating from the sales of WEB4D and EPC4D domains, should be allocated to support [W3C](#) activities.
- The Net4D *classes* should be open and interoperable with others resolving schemes ( eg [Handle.net](#) ). for example through the use of the [NAPTR field](#).



## CONCLUSIONS

- DNS 1.0 --> Monopoly : ICANN, Web 1.0 HTML, US parentage, English only.
- DNS 2.0 --> Open competition including *inter alia* Net4D , Semantic Web, XML, Web4D - EPC4D fully international and multilingual
- An open, coherent and secure approach to linguistic diversity, not just a patch.