

# Appeal to the IESG

march 10, 2010

J-F C. Morfin

Dear IESG Members,

You announced on January 11, 2010 that you had approved the IDNA2008 document set, as produced by the WG/IDNABIS, except for the Mapping document:

1. The IDNA2008 documents set introduces a fundamental clarification of the way the Internet architecture meets an external diversity, in this case the linguistic diversity. This significantly extends the Internet architectural paradigm, i.e. the commonly accepted basic set of assumptions that is documented in [RFC 1958](#), [RFC 3439](#), [RFC 3935](#), and IDNA2003. However, the IESG has not warned the Internet community and has not called for IAB documentation concerning:

- the set of new architectural opportunities that this new architectural paradigm opens,
- the possible confusions that it may introduce until these opportunities are properly governed.

2. In line with this new paradigm, IDNA2008 only documents the Internet side (the protocol) of IDNA. It does introduce it, but it does not document the powerful, yet complex, and possibly versatile, capacities that are allotted to the user's side in order to enable the Internet multiplication that the linguistic diversity requires. Furthermore, the IESG has not published a disclaimer indicating that this makes IDNA2008 too early to deploy, and useless to test, prior to research and coordinated documentary work being engaged, all of which is in a Usage Area that is still to be organized inside or outside the IETF.

## Contents

- [1 Not enough precautions being advised](#)
- [2 Why am I appealing as an Internet User?](#)
  - [2.1 Interest in the Internet](#)
  - [2.2 Importance of IDNA2008](#)
- [3 The points of appeal](#)
- [4 Context of the Appeal](#)
  - [4.1 Working Group Summary point of view](#)
  - [4.2 The WG/IDNABIS point of view](#)
  - [4.3 Point of view of an Application side designer](#)
  - [4.4 AD + WG/Chair point of view](#)
  - [4.5 ICANN published point of view](#)

- [4.6 The IAB point of view](#)
- [4.7 The point of view of an Internet and IETF deliverables user](#)
  - [4.7.1 I am a person](#)
  - [4.7.2 I am an Internet user](#)
  - [4.7.3 I am an Internet lead user](#)
    - [4.7.3.1 RFC 1958](#)
    - [4.7.3.2 RFC 3439](#)
    - [4.7.3.3 RFC 3935](#)
    - [4.7.3.4 IDNA2003](#)
  - [4.7.4 IDNA2008](#)
  - [4.7.5 However, there is a big "however"](#)
  - [4.7.6 Am I wrong if?...](#)
- [4.8 IDNA2010/IDNA2012 BCP engaged work](#)
- [5 Annexes](#)
  - [5.1 Annex 1. Application of the Precautionary Principle by the IETF](#)
  - [5.2 Annex 2. Is IDNA an ML-DNS to be?](#)
  - [5.3 Annex 3. A practical innovation and complex deployments strategy](#)
  - [5.4 Annex 4. Report to the WG/LC](#)
  - [5.5 Annex 5. WG/LC IDNABIS report](#)
  - [5.6 Annex 6. Punyplus](#)
  - [5.7 Annex 7. Projet.FRA report](#)
  - [5.8 Annex 8. Mail exchange with Patrik Falström](#)
  - [5.9 Annex 9. AD and Chair point of view](#)
  - [5.10 Annex 10. IAB recommendations and considerations](#)
  - [5.11 Annex 11. IDNApplication Medley](#)

## Not enough precautions being advised

I believe that in these two related occurrences the IESG has failed in its precautionary duty as it results from RFCs ([Annex 1. Application of the Precautionary Principle by the IETF](#)) and, moreover:

- At the very inception of the WG/IDNABIS, I asked the Chair about his IDNA2008 strategy regarding an ML-DNS ([Annex 2. Is IDNA an ML-DNS to be?](#)), which is loosely defined as what the users really need. Further to his response, I committed to the IDNA2008 support in order to obtain a consensual basis towards the further, fully consistent, documentation of such an ML-DNS. The Chair succeeded in obtaining an IDNA2008 consensus that I supported.
- Working with the france@large team on multilinguistics as the cybernetics of the linguistic diversity, and the **Intersem** as the Semiotic Internet that is necessary to support semantic exchanges (Internet of thoughts), I considered the ML-DNS issue through the requirements of the two projects as required by the standardization

path, in the foreseeable contexts ([Annex 3. A practical innovation and complex deployments strategy](#)), but also the immediate real world consequences of an IDNA2008 approval: Project.FRA of an open francophone ontology using its namespace as its semantic addressing taxonomy, and Multilinc of one naming zone per sociolinguistic relational space. I made sure that the WG/IDNABIS was informed enough of the comments from france@large ([Annex 4. Report to the WG/LC](#)), maintained a draft in turn recording the WG/LC issues and some of the responses obtained ([Annex 5. WG/LC IDNABIS report](#)), published a Draft to discuss the inability to support exceptions such as the French and Latin languages' majuscules ([Annex 6. Punyplus](#)), and engaged thought with the IUCG in the discussion of an "Interplus" architecture matching the concerns expressed by the Application Area Director on the user side (this architecture is just a possible IDNA2008 application and is, therefore, quoted in this appeal only as an example of the user reading of the IDNA2008 perspective of the Internet technology).

This is in the capacity of the Project.FRA designer that I reported and stated my position to the IESG, just as in how I had initially informed the WG/IDNABIS of the Multilinc project ([Annex 7. Projet.FRA report](#)).

I was told by the IESG Chair that: these comments stimulated a significant amount of discussion about the documents from the IDNabis WG. The IESG approved the documents of which I was most concerned. However, they took my report only as "suggestions" and as "research" on some actions that I had reported that needed to be immediately followed and that I was concerned about.

## Why am I appealing as an Internet User?

For decades, I have been a "lead user" of the world digital ecosystem (WDE) international network, or an **IETF Internet User** in ICANN lingo: an **@large**. We will see an evolution at the end of this memo: lead users want to emphasize their ability to be fully addressable without resorting to P2P interuser applications. Some translate this in saying that they are "**interusers**" - this is also related to the **Interplus** technology and Internet extension that they work on as a result of IDNA2008.

## Interest in the Internet

My interest in the Internet is in the basic, valued-added, and extended digital services that I may obtain from it in order to expend my own relational spaces with other people, communities, and machines alike at an intelligent layer. This is why I fully adhere to the WSIS unanimous declaration that the world information society, **MUST** be "**people centered, a caractere humain, centrada en la persona**".

I call a basic service a service that concerns the signals carrying the content.  
I call a value added service a service that concerns the envelope of the content.  
I call an extended service, a service that concerns the semantic of the content.

Therefore, along with the current state of the art, I subsequently actively and innovatively kept enhancing the architecture, ability, and reach of my own international digital continuity as not being decentralized, but rather **distributed**, i.e.:

- **centered** on my own personal "HSS" (*homo sapiens sapiens*) **semantic processor** assisted, surrounded, and protected by all the resources of my own "cybship", and
- **multicentered** on the similar semantic core and digital agent processors of my fellow Internet users.

## Importance of IDNA2008

The importance of IDNA2008 results simultaneously from:

- its own content enlarges the Internet capacity from the satisfaction of the Unicode standards, to the full cultural and semantic needs of human users.
- the lack of documented consensual guidance on the way to securely and stably do it in dramatically enlarged namespaces
- the constraints imposed on users by ICANN in its limited namespaces (TLDs and IP addresses), while users will now understand how to evade.

The result:

- sets an IDNA stable basis for a multi-layered DNS that matches my local, global, lingual, application, etc. needs in terms of name resolution.
- illustrates how the Internet legacy DNS building block is to be used to support the Internet presentation layer.
- confirms, in terms of real life operations and usage, the **Internet concept of multiplicity**: "internet multiplicity is supported at the fringes in using end to end metadata exchanges (cf. [RFC 1958](#): "everything else should be done at the fringes"). **Multiplication is supported by fringe metaprocess instantiations.**
- introduces the Internet presentation layer,
- implies a unique virtual DNS root matrix,
- leads to global scope user addresses, etc.
- has excessively limited considerations of:
  - the impact on the user side,
  - on the zone manager side,
  - and on the transition of the millions of existing users.

This is why:

- from the very outset of the WG/IDNABIS, I asked the Chair if the way he intended to respect the Charter would lead to a user or only to a network side satisfactory

solution. The response in being the latter, led me to commit to the support of the WG's work and to compatibly extend it to support the needs of the user side, in particular through the Multi-Layer DNS pile.

Experience showed that the Internet side and the User side were not able to be debated in the same WG. Therefore, with the IETF Chair's help, I initiated the [iucg@ietf.org](mailto:iucg@ietf.org) non-WG mailing list and engaged in building and supporting the <http://iucg.org> site. This permitted the exploration of optional, transparent "**Intelligent User Interface**" (IUI) middleware as a response, on the user side, to:

- the IDNA2008 requirements,
- the Internet multiplicity needs,
- and to actively consider the support of the Intersem (Semiotic Internet) for which I am striving.

In conformance with the rules of appeal:

- during the Last Call period and further to the IDNA2008 approval, I have exchanged mails on this matter with:
  - Vint Cerf, Chair, WG/IDNABIS
  - Lisa Dussault, Director, Application Area
  - Russ Housley, Chair, IESG
  - and sent a position statement where I explained as to why I would need to appeal if the Internet community was not warned and helped to adapt to the changes, opportunities, and risks that would possibly be introduced by the change from IDNA2003 to IDNA2008 in the current context.
- I explained the context that we face due to ICANN's real life attitude, and lack of proper architectural information
- I announced this appeal as a way to permit a reflection period for the Internet Community
- I introduced the kind of non-community debated architectural solution that I would be forced to deploy and that everyone could imitate or enhance without proper order due to a lack of standardization structure, with possible important risks for the naming industry and the Industrial Property protection (this is the reason why I have not published a draft yet. It is, however, documented and actively read on the IUCG site).
- I reported that I positively tested its concepts as only being a simple user's reading of the Internet technology in light of IDNA2008, in turn calling for not a single change in the Internet deployed technology, and in a first step only for a punycode extension.
- I tried to make people aware that it might imply an unprecedented change in the Internet management, from decentralized to distributed.

Like every other protocol, IDNA2003 was designed to support the typographic complexity of scripts in being closely related to Unicode. Based upon field experience, IDNA2008 re-visited it in order to address the spiraling orthotypographic complexity of the human linguistic diversity. This first led to an impasse relating to the mapping of Unicode characters into other Unicode characters prior to the generation of a punycode equivalent string to produce an A-label. This impasse showed the upper limit of the Internet architectural paradigm.

- Its consensual resolution introduced a new Internet architectural paradigm where:
- an "unusual" but necessary "off the network" operation is to be applied to the user input in order to prepare that user input for use in an "on the network" protocol and to maintain interoperability between the network and usage (natural languages).
- this new paradigm expects that the Internet "core" deals with technical internal complexity, while external diversity is addressed at an Internet usage interface (IUI) peripheral level. This could be expressed, from the observation of the IDNA2008 discussion, as the principle that the Internet core is where the IETF can use MUSTs, the IUI is where it can only use SHOULDs, and external usage is where it can only use MAYs.
- since the full support of linguistic issues belongs to the "presentation" layer, the Internet "presentation" layer is subsequently evidenced (actually the "xn--" IDNA header means that IDNA uses the Internet "xn" presentation).

## The points of appeal

This is why this appeal does not concern the IDNA 2008 document set, as approved by the IESG, which is now of prime stable importance when considering the Internet architecture from a user perspective, but also concerning the way the IESG has approved this IDNA2008 document set, while:

- not obtaining and inserting a disclaimer from the IAB in turn warning the Internet users community about the architectural consequences of the "unusual" strategy that the IETF adopted in this document set "to insure interoperability" (cf. Mapping document).
- not classifying it as an IESG disclaimer warning for the Internet users community about the necessary incompleteness of the new introduced IDNA architecture, when compared to IDNA2003, due to its open nature on the user side.
- in spite of the Working Group Summary statement that: "There was an impasse relating to the mapping of Unicode characters into other Unicode characters prior to the generation of a punycode equivalent string to produce an A-label [please see the Definitions document]. This was resolved by introducing the non-normative

“mappings” document”, wherein that Mapping document was not simultaneously approved, but while it should at least have been acknowledged at the same level as the Rationale document.

- in so doing, in not having considered its precautionary duty enough as it results from the IETF mission, and introducing confusion that is to be urgently clarified before its consequences might endanger the entire Internet system architecture and the operational deployment unsuitability for the reasons detailed in this appeal.

## Context of the Appeal

To try to best set the context of this appeal, I will try to document it from various points of view.

- the Working Group Summary point of view.
- the WG/IDNABIS point of view
- the point of view of an application side designer.
- AD + WG/Chair point of view
- ICANN published point of view
- the IAB point of view
- my own point of view as an Internet/IETF user
- the "IDNA2010" BCP work

### Working Group Summary point of view

"The initial impetus for the revisiting of the IDNA2003 proposed standards emerged in written form in RFC4690. An informal technical team worked to develop a framework for consideration that was later discussed, edited, and ratified to create the IDNABIS working group in 2008. Readers will note that this is nearly 2010 but the new specifications bear the label IDNA2008 because the work was started in that year. The documents resulting from the IDNABIS Working Group effort have been extensively discussed over a two year period by the WG and by interested parties especially language experts in the Chinese, Japanese and Hangul spaces, speakers of Hebrew, Indic languages as well as a working group of expert Arabic speakers. The WG has had the participation of several Unicode consortium representatives. There was controversy during the development of these documents but a rough consensus has formed around the recommendations.

There was an impasse relating to mapping of Unicode characters into other Unicode characters prior to the generation of a punycode equivalent string to produce an A-label [please see the Definitions document]. This was resolved by introducing the non-normative “mappings” document observing ways in which this aspect of dealing with internationalized domain names might be approached. The principal rationale for re-visiting the IDNA2003 recommendations arose from field experience and a recommendation from the IAB [RFC4690]. A major objective was to re-specify the standard in such a way as to make it

independent of changes to the Unicode code set (something that evolves more or less continuously). The method for achieving this is to use the Unicode properties feature (per code-point/character) to determine whether the code point should be allowed, disallowed or used only under certain conditions in domain name labels. There remains the problem of deployment in a world of web servers, browsers and other applications using domain names that may have already implemented the IDNA2003 recommendations.

Implementation tactics for dealing with the old and new specifications may vary. Perfect backward compatibility with IDNA2003 was not possible (without simply replicating it, negating the rationale for the redefinition effort of IDNA2008). This is particularly true of the special characters “esszet” or “sharp-S” used in German and the “final sigma” used in Greek. These were mapped into other valid Unicode characters under IDNA2003 but allowed in IDNA2008 because the Unicode code points for these were introduced in the interim between IDNA2003 and IDNA2008 standardization efforts. Registries have several implementation choices including bundled registration of previously mapped and newly allowed characters or rejection of conflicting new registrations. The treatment of languages that are expressed in Right-to-Left form (see “Bi-Di” document) has been revised relative to IDNA2003 and it is believed that the revision is clearer and more precise in its form and limitations on the use of numeric characters, for example."

## **The WG/IDNABIS point of view**

The WG/IDNABIS has twice approved the Mapping document by rough consensus declared by the Chair. It has resolved its inner impasse (cf. Working Group Summary) on this document. This document states what the WG thinks, and what the approved document set is based upon.

The WG/IDNABIS and the authors of the Mapping document did not want it to be approved as a standard track because "it does not specify the behavior of a protocol that appears 'on the wire'". However, it illuminatingly explains that IDNA2008 "describes an operation that is to be applied to user input in order to prepare that user input for use in an 'on the network' protocol. As unusual as this may be for an IETF protocol document, it is a necessary operation to maintain interoperability".

It makes clear the necessity of a work to be engaged to maintain and serve interoperability on the user side. Without this work, IDNA2008 is not interoperable with the user side.

## **Point of view of an Application side designer**

I asked Patrik Falström, author of one of the documents, about his understanding of the development work to be carried out on the User side, of which he is an acknowledged international expert and a professional designer, as results from IDNA2008 ([Annex 8. Mail exchange with Patrik Falström](#)). This shows the opened possibilities on the user side, but also



the lack of warranty for identical results as well as simple guidance. This emphasizes the need for an IDNA related BCP by users and user application developers.

## AD + WG/Chair point of view

At the end of the LC, the AD posed several pertinent questions that have not been answered yet ([Annex 9. AD and Chair point of view](#)), but of which france@large has commented on. The Chair suggested ICANN to be called upon as the representative of the users. I wholly object to this for several reasons:

- ICANN only takes care of the administration of addressing and class IN naming tables. It has nothing to do with the edge Internet. A technical disagreement with a linguistic community might lead to a split of the Internet.
- ICANN has no legitimacy to represent users. As an example, france@atlarge is the eldest ALS. It was incorporated by French candidates to the ICANN BoD in 2000 to serve the French @large community. However, france@large was denied ALAC membership. INTLNET is a non-profit that I created with Tymnet back in 1978: it was denied to join the non-commercial constituency. For years now, I have been denied the possibility to join the various ICANN IDN WG: I do not object to their right to do this, however, since I do not share their strategic doctrine. However, this implies that they cannot be considered as an open, neutral forum to represent users, ccTLD, IDNccTLD, and IDNgTLDs.
- I thank the AD for the questions that she posed. I first objected because what was important to me was to get IDNA2008 adopted after its impasse having been consensually addressed through the still not-adopted Mapping document. I also thank the Chair for insisting that there is a WG consensus supporting that respective Mapping document. I object to the IESG delay in adopting it and in having adopted the rest of the IDNA2008 document set: this leaves pending the question of where the user side of IDNA is to be studied, debated, and documented before one can start testing IDNA2008. This gives the false feeling to ICANN (cf. infra) that they can usefully proceed with their FAST TRACK project, which will probably be matched by the "MY TRACK" counter project by dismayed IDNgTLDs who will have prematurely used the unique virtual root matrix.

## ICANN published point of view

As the Internet User Rod Beckstrom, President and CEO of ICANN, puts it, "**One of the most significant innovations in the Internet since its inception is the introduction of top-level Internationalized Domain Names (IDN TLDs)**". This is certainly true. However, from the context of his declarations, the Internet User Rod Beckstrom seems to believe that:

- this results from the IESG approval of the IDNA2008 document set as a "revised version of the IDNA technical standard"
- the transition of which could be:
  - tested through the sole ICANN Fast Track project, which only involves a few non-Roman IDNccTLDs and neither Roman IDNccTLDs nor innovative IDNgTLDs,
  - and further on accompanied by the sole ICANN entity, while many TLDs outside of the ICANN zone area may deploy through a unique virtual root matrix.

This is erroneous.

However, this dangerous confusion seems to be shared by several other key stakeholders of the Internet Governance due to the lack of IESG and/or IAB proper guidance. They ignore the fact that the IESG approved IDNA2008 document set:

- documents the way the **Internet** should behave when accessed using IDNs,
- **does not document** the way Internet Users should transition safely and securely from an environment of millions of IDNA2003 IDNs that are already in use today in the unique class IN, to the **millions of TLDs** in the multiplicity of classes and presentations of the **unique virtual root matrix**, in which this already engaged transition in turn leads to.

Would Internet Users at large stay under the false impression that IDNA2008 is a comprehensive Internet protocol and proceed anyway as if it was? Could this not have catastrophic consequences for the stability and credibility of the Internet as well as for the entire Internet economy? They ignore the fact that the IDNA2008 document set is a first necessary step on the network side, which is just as necessary as a second step on the users and operators' side, and possibly a third step in between the network and its users in order to organize their related administration.

## The IAB point of view

IAB has explored what can be learned from the current difficulties in implementing Internationalized Domain Names (IDNs) without the intention for this document to influence any current working group charter.

This document spells out recommendations and security considerations ([Annex 10. IAB recommendations and considerations](#)) that should be considered by any globalization or multilingualization effort.

The IDNApplication concept that I support consists in respecting the IAD recommendations, and in having a single separate application, on the user side, dedicated to the ML-DNS (multilayer DNS). The problem raised by such an architecture is that it can easily support a unique ASCII/UTF-8 virtual root multiclass matrix.

# The point of view of an Internet and IETF deliverables user

This section was reviewed privately and shortly at the IUCG, out of the appeal context.

## I am a person

This means that I am centric to the information society. This also technically means that I am a subject operated by a natural semantic processor of the HSS brand (Homo Sapiens Sapiens) whose capacities interact with material and immaterial objects and semiotically relates with other subjects throughout various ecosystems. For example, my body is my interface with the physical ecosystem; it supports my semiophysical capabilities (utterances, gestures, etc.) that include my semidigital abilities (use of a computer, mobile phone, word processor, mouse, etc.).

## I am an Internet user

This means that one of these ecosystems that I belong to is the world digital ecosystem (WDE) of which the Internet is the current main content level communication commodity. My digital self is, therefore, embodied in my "cybship", i.e. the real and virtual elements that are necessary to my online presence and all the various online properties and systems of mine that I may use to navigate and operate in my digital life.

## I am an Internet lead user

This means that I am interested in improving my own digital semiotic capacities and the semiotic and behavioral feedback of those of whom I use to interact and relate. This implies:

- the best adminance of my, and of the common, investments,
- the best governance of my, and of the common, digital and general life,
- their most pertinent concertation (usa: concerted cooperation).

This memo only focuses on the adminance part, i.e. how to provide governance with the structural tools that it needs for all of us to best use them.

The adminance of my, and of the common, structural tools calls for stability and the enhancement of what exists:

- **Stability** can be ensured by way of the maintenance of equipment and the governance of its rules of use.
- **Enhancement** results from structural investment innovation.

Innovation can be inventive, incremental, disruptive, or architectural.

- Inventive innovation is usually academic.
- Incremental innovation in the signal and content communication areas primarily belongs to the ITU and IETF scopes.
- Disruptive innovation usually spurs industrial ventures.
- Architectural innovation is often a synonym for complexity simplification, something that often results from looking at an established solution set within a new user perspective.

If I consider the current state of my communication capabilities, I can clearly see that I am "half-way through":

- **Signal and passive content** bandwidth and communications are available with half the world being connected (Internet, mobile phones) under stable terms, which might be endangered by some political and commercial interests.
- **ambient and active content and semantic** bandwidth and communications are not available yet.

I can actually perceive an intrinsic architectural difference between what I already have and what I am missing.

- What I have supports me, as in any other **end user**, among the growing billions of accessing interchangeable other people or machines, while only being identified through a stable or transient IP address. This unique **neutrally shared** system works well, but it has a growth problem that is leading to concerns about its most precious feature: **neutrality**. This is a system matching end to end diversity by **growth**.
- What I am missing will have to successfully support me as a particular **interuser**, who is calling and being called because he or she is unique throughout the diversity of the billions of interusers. The question is then: could such a multiplicity work? This would be a system addressing diversity by **multiplication**.

Up to now, Internet technology seems to have supported multiplication (charsets, IP addresses, number of accesses, traffic, root servers, etc. ) by growth. This has been documented in four occurrences, in particular: [RFC 1958](#), [RFC 3439](#), [RFC 3935](#) and US and French laws, IDNA2003. IDNA2008 now supports growth by a multiplications ... I have, with the rest of the multitude, to provide.

### [RFC 1958](#)

[RFC 1958](#) is based upon the permanent change principle ("Principles that seem sacred today will be deprecated tomorrow. The principle of constant change is perhaps the only principle of the Internet that should survive indefinitely"), and on the guidelines that should, therefore, be revisited upon experience but that seem to remain rather pertinent:

- The goal is connectivity, the tool is the Internet Protocol, and the intelligence is end to end rather than hidden in the network.
- Connectivity is its own reward.
- In an ideal situation there should be one, and only one, protocol at the Internet level.
- End-to-end functions can best be realized by end-to-end protocols.
- Everything else should be done at the fringes.
- Nobody owns the Internet, there is no centralized control.
- Heterogeneity is inevitable and must be supported by design.
- If there are several ways of doing the same thing, choose one.
- All designs must scale readily to very many nodes per site and to many millions of sites.
- Performance and cost must be considered as well as functionality.
- Keep it simple. When in doubt during design, choose the simplest solution.
- Modularity is good. If you can keep things separate, do so.
- In many cases it is better to adopt an almost complete solution now, rather than to wait until a perfect solution can be found.
- Avoid options and parameters whenever possible. Any options and parameters should be configured or negotiated dynamically rather than manually.
- Be strict when sending and tolerant when receiving.
- Be parsimonious with unsolicited packets, especially multicasts and broadcasts.
- Circular dependencies must be avoided.
- Objects should be self describing (include type and size), within reasonable limits.
- All specifications should use the same terminology and notation, and the same bit- and byte-order convention
- Nothing gets standardized until there are multiple instances of running code.
- Avoid any design that requires addresses to be hard coded.
- A single naming structure should be used.
- Public (i.e. widely visible) names should be in case-independent ASCII. Specifically, this refers to DNS names.
- Addresses must be unambiguous (unique within any scope where they may appear)
- Upper layer protocols must be able to identify end-points unambiguously.
- Prefer unpatented technology.
- Any implementation which does not include all of the required components cannot claim conformance with the standard.
- Designs should be fully international, with support for localisation (adaptation to local character sets).
- All designs must fit into the IP security architecture.
- Confidentiality and authentication are the responsibility of end users and must be implemented in the protocols used by the end users.
- Wherever an algorithm is called for in a protocol, the protocol should be designed to permit alternative algorithms to be used and the specific algorithm employed in a particular implementation should be explicitly labeled.
- Preference should be given to algorithms which have stood the test of time and which are not unnecessarily inefficient.
- To ensure interoperation between endpoints making use of security services, one algorithm (or suite of algorithms) should be mandated to ensure the ability to negotiate a secure context between implementations.

## [RFC 3439](#)

[RFC 3439](#) documents the Internet as a very **large system** calling for an absolute respect of the Simplicity Principle, which states that complexity is the primary mechanism that impedes efficient scaling. This directly matches our lead user architectural primary concern: simplifying complexity rather than incrementally addressing it or disruptively increasing it.

However, if [RFC 3439](#):

- does explain that "optimization can also be considered harmful. [as it] introduces complexity, [as it introduces ] introducing tighter coupling between components and layers" because "much of the non-linearity observed [in] large systems is largely due to coupling."
- it does not propose any simplification doctrine based on uncoupling. However, it seems to introduce one when it notes that: "Adding a new Internet service is just a matter of distributing an application to a few consenting desktops who wish to use it."

## [RFC 3935](#)

[RFC 3935](#) perceives the Internet of Engineers as:

*a large, heterogeneous collection of interconnected core and edge systems that can be used for communication of many different types between any interested parties connected to it. The Internet is a truly global network, reaching into just about every country in the world.*

which can be matched with the legal operators perception of:

*a combination of computer facilities and electromagnetic transmission media, and related equipment and software, comprising the interconnected worldwide network of computer networks that employ the Transmission Control Protocol/Internet Protocol or any successor protocol to transmit information.*

and the user perception of "online" services based, by French law, on trust understood, at network level, as stability, predictability, neutrality

A convergent open definition could then be:

***all the packet-switched interconnected processors to support trusted public and private online communications and relational services.***

You forget about adjacencies. This will make your text less readable for Asians. Moreover you quote "bandwidth and communications" several times.

## **IDNA2003**

The widest diversity that the Internet has had to meet up to now is the linguistic diversity (there are around 25,000 sociolinguistic units and languages and IP evolve all the time). Such diversity is addressed in network architectures by the presentation layer. IETF, so far, has documented no presentation layer support for the Internet. The linguistic diversity has, therefore, been met by increasing the character bandwidth (ISO 10646 - Unicode, which now supports nearly 100,000 signs), this is usually called i18n, or internationalization.

A user can be offered five main levels of support for linguistic diversity:

- **none**

Linguistic issues are not considered by technology. This may result from the use of signs, numbers, keys, or nothing like a physical voice.

- **lingualization**

A given natural or coded language is at some stage embedded in the technology. Plurilingualization is possible by using polynoms (synonyms in different languages).

The Internet is lingualized ([RFC 3935](#): The "IETF uses the English language for its work [] because of its utility for working in a global context.").

- **globalization**

A lingualized technology is extended in order to be able to transparently quote any other language.

This calls for at least three features:

- support of the **typography** that is used by every language. This is **internationalization**. Every existing character should be supported end to end by the medium.
- support of the cultural **orthotypography** (or typographical syntax: what defines the meaning and rightful usage of typographic signs, formats, etc.). This is "**localization**". Computer localization is documented in "locale" files.
- **content tagging** in order to indicate the **orthotypographic conventions** of the source. This is the [RFC 4645](#), 5646, and 4647 set.

- **multilingualization**

This means that the technology in turn supports the human orthotypography. This is necessary for any semantic oriented services.

- **universalization**

At this level, sememes (semantic units) are used. They are externalized through the mecasemiotic or mecalanguage (sememes driven, machine semiotic or natural language versions) of each end.

IDNA2003 in being tightly coupled to Unicode and its machine level orthotypography (CLDR project) was rigidly bound to an end to end globalization. Together with the IANA langtags registry, this provided users with a consistent experience of the globalized Internet system, fully supporting the various diversities (speed, protocols, addresses, traffic types, machines typography and orthotypography, languages, etc.) through single corresponding bandwidths.

## **IDNA2008**

IDNA2008 does not change that Internet. To the contrary, because it rightly acknowledges that my semantically driven orthotypographic needs are too complex for a single end to end solution, it proposes to use that globalized Internet as a rock solid, end to end, LDH DNS, intelligence on the fringes, digitally globalized, etc. core tool-kit, together with its [RFC 1958/3439](#) checklists, in turn allowing us, interusers, to build our own Internet usage Interface (IUI) and our own IUI to IUI conventions.

Multiplication no longer has to be supported by bandwidth growth (in this case typographic and orthotypographic bandwidth), but is distributed among the myriad of our user processors. It is up to us to support our human semantic inter-relational protocols, that is to say natural languages.

This acknowledges the [RFC 3439](#) Law of Diminishing Returns, which states that if one factor of production is increased while the others remain constant (such as some mapping at the protocol level, i.e. within the Internet core), the overall returns will relatively decrease after a certain point (IDNA general usability). IDNA2008, in uncoupling IDNA from Unicode versions, acknowledges that "trying to squeeze out" Unicode efficiency in core Internet protocols "past" the ASCII DNS "only adds complexity, and hence leads to less reliable systems".

In addition, brain to brain interintelligibility proceeds from a non-binary simplicity. It became too complex to support it in using global protocols past the passive content layer. The elegance of IDNA2008 is to use the Unicode based punycode algorithm to build an hourglass solution where the entire multilingual name space is robustly and flexibly supported by multiplication at the IUI to IUI layer, over protected end to end legacy ASCII DNS.

## **However, there is a big "however"**

This "however" is that a user guide is missing. As an **interuser** having to support my language, I need a user guide to inform me as to how to best use the [RFC 1958/3439](#) checklist, at the IUI layer, when having to properly and securely interface the Internet IDNA2008 protocol.

As an **interuser**, I know that there is now an "Internet tool-kit" permitting me to engage in my second half-way journey towards the brain to brain communications and facilitation services that I strive for:



- first in supporting extended network services on my computer side (we understand):
  - **basic** services, as transport related,
  - **value-added** services, as envelope related,
  - and **extended** ones, as content related.

I expect that these extended services will:

- not only serve a robust WIRIWIG (what I register is what I get) ML-DNS domain name pile (i.e. all the UTF to ASCII **various presentations of the same domain name**) depending on the used algorithms and resolution systems.
- but also support ambient content (for example, local speed limit) and **active** content (what you get is what I wanted you to receive, not what I sent).

This is the field of the Interplus proposition - an IUI architecture that has the unfortunate capacity to easily support the Internet presentation layer as a fall-out of IDNA, the French orthotypography that IDNA2008 does not support, the non-administered yet unique DNS virtual root matrix with billions of non-necessarily unique TLDs in all the classes, and a few other possibilities.

- then, I will be in a position to enter the semiodigital area of the semantic bandwidth and communication, which we call the **Intersem**.

This is why some concerted guidance and community assistance would be of great help.

- Maybe with an Internet Adminance enhanced cooperation that I could trust,
- certainly with an IDNA usage oriented best practice document and forum,
- and an IUI architectural research on the best IDNA and Internet Multiplication capacity matching configuration that I should adopt.

### **Am I wrong if?...**

Am I not better off by having a single Internet Multiplication Management System on my PC rather than every application using its own solution brand?

For example, I would prefer one single IDNApplication supporting IDNs rather than different custom modules in my various browsers, mail agent, etc.

If I feel that I should reconsider my entire "cybship" architecture so that it may adapt from sailing in its enlarging river environment in order to navigate the varied weather of the multiple internationalized high seas.

I would like to first adapt my system terminology in order to be relevant:

- every system belongs to that system intricacy that the universe is.
- the part of that universe that may relate with a system is its exotem.
- the interface between a system and its exotem is its peritem.
- it is subjective as to how one considers a system from its inside.
- perijective from its peritem.
- circumjjective from its neighborhood.
- objective from its exotem.

This would mean that:

- the Internet exotem is every system that may immediately, or in the future, relate with the Internet.
- the Internet peritem should, therefore, be designed so as to be interoperable with all of its exotem.
- by its open definition (communication continuity of the universe digital eco-system), the Internet has an unlimited size (every processor made interoperable by packet switching networks) from nanos to galaxies. In this, it is another "universal", such as the cosmos, life, time, etc.
- henceforth, its exotem necessarily includes the "internat(ure)", the natural ecosystem communication continuity.
- since we cannot change the natural simplicity, interoperability with the "internat" calls for the Internet simplicity that is demanded by [RFC 3439](#) to be interoperable with the simplicity of the "natural" diversity. This is what the "Mapping" document alludes to when it refers to the "unusual" need for the IETF to document exotemic particulars in the case of the natural linguistic diversity.

At this stage, IAB guidance is needed. [RFC 3439](#) states that the Internet, as a very large system, is to be based upon the Simplicity Principle, which states that complexity is the primary mechanism that impedes efficient scaling. IDNA2008 "unusually" accepts to

- uncouple the end to end globally unique core network system
- from the user linguistic diversity semantically controlled exotem
- introducing a linguistic Internet peritem that is to be built by the myriad of users' own linguistic peritem,
- as a **simplification of complexity**.

This leads one to think that the "Mapping" document, which was necessary to address the IDNA2008 impasse but that the IESG has not approved yet, seems to clearly identify:

- the Internet **system** to where the IETF can use **MUSTs**
- the Internet **peritem** to where the IETF can use **SHOULDs**
- the Internet **exotem** to where the IETF can only use **MAYs**

Guidance is, therefore, needed from the IAB in at least two areas:

- Is the Internet Simplicity Principle's simplicity the same as Nature's simplicity (that can be summarized by Maupertuis' universally accepted principle). How do they differ? How do they interoperate? What are the implications?
- Does the **Internet peritem** belong to the **IETF scope**?

Other guidance that I will need, as an interuser, is to know about the reproducibility of the IDNA2008 architectural simplification, because there are other areas where "the probability of achieving a stable implementation of an architecture is inversely proportional to its number of components" ([RFC 3439](#)). Moreover, complexity is often "software-based, not hardware driven, and therefore had cost curves worse than Moore's Law.", which may lead to situations such as ROAP.

In particular, I need to know if the Amplification Principle constrains or frees the various IDNA2008 deployment strategies. It states that there are non-linearities that occur on a large scale that do not occur on a small to medium scale. IDNA2008, as I see it, still extends the size of the Internet, and multiplies the number of small things. Is this multiplication acting as a catalyst or, to the contrary, as an uncoupling inhibitor of their possibly huge impact amplification? Furthermore, [RFC 3439](#) states that the minimization of interaction contributes to simplicity. It adds: "As a result, consideration of "core vs. end-point" state interaction is crucial when analyzing protocols such as Network Address Translation (NAT), which reduce the transparency between network and hosts.". Will IUIs have a NAT like effect or will the strict polynymy (synonymy in different languages) of the ML-DNS pile (bijective conversion between the various versions of a domain name through its different layers: A-label/U-label) protect users from this?

## **IDNA2010/IDNA2012 BCP engaged work**

After the closure by the IETF WG/IDNABIS of the IDNA2008 [work](#), we now have to validate this new version of IDNA after six years of operations under IDNA2003. Subsequent to full-scale field tests being carried out, users' contributions, ICANN comments, submissions of new gTLDs, and our operational practices, we will need to reach a final consensus in order to operationally consolidate this major evolution of the Internet of users.

It is now up to us to make **IDNA2010** an Internet community convergence point.

The plan is to use the real-life **Projet.FRA** and **Multilinc** experiments as test beds. These projects have already documented the fact that prior to engaging in commercial services, e.g. to support the French majuscules, some glyphs, and special bidirectional usages, they need an **IDNA2010 user side document that has been tested by the community and approved by the IETF**.

This is why an **IDNA2010 special interest group** of the IUCG is currently being created with the target of continuing the work on a finalized IDNA operational and culturally adaptive version.

Because ccNSO started referring to "IDNA2010" as the next version of IDNA2008, I created:

- <http://idna2010.org> as a project of BCP, documenting the implementation of IDNA2008 on the user side.
- <http://idna2012.org> as a project of BCP, documenting the administration and maintenance of the joint operations of IDNA2008 on the Internet side and IDNA2010 on the user side.

In order to not confuse issues, it was agreed with the WG/IDNABIS Chair that the [workon@idna2010.org](mailto:workon@idna2010.org) would be frozen until it published a first draft. This appeal is considered as a first Draft.

To permit IESG Members to address the kind of issues involved in the permanent SIG IDNA group (and in some cases the emotional weight involved that belongs to linguistic orthotypography but does not belong to RFCs), here is a documented list of topics for such an "IDNA Internet tool kit" support and information group. ([Annex 11. IDNApplication Medley](#)).

# Annex 1.

## Application of the Precautionary Principle by the IETF

This Memo documents how the IETF pursues its mission in adherence to the Precautionary cardinal principle and may benefit from its application in its goal to make the Internet work better and in addressing some of its documented problems.

### Contents

- [1 Introduction](#)
- [2 Better safe than sorry](#)
  - [2.1 Definition](#)
  - [2.2 Legal status](#)
  - [2.3 Areas of applications](#)
  - [2.4 Innovation ethic](#)
- [3 Enacting the Precautionary principle](#)
  - [3.1 When does the principle apply?](#)
  - [3.2 Investigating a possible case of application](#)
  - [3.3 Attitude to be adopted](#)
- [4 IETF mission, cardinal principles, and the standard process](#)
  - [4.1 Cardinal principles](#)
  - [4.2 The standard process](#)
    - [4.2.1 Criteria for selecting new work](#)
  - [4.3 Charter](#)
- [5 A way to address identified IETF difficulties](#)
  - [5.1 The Balance Between Research, Invention, and Adoption](#)
  - [5.2 The Reach of the Internet](#)
  - [5.3 Funding the Internet R&D](#)
    - [5.3.1 Commercial funding](#)
    - [5.3.2 Precautionary alternative to commercial funding](#)
  - [5.4 Difficulties of the IETF in accomplishing its Mission](#)
  - [5.5 The IETF has Difficulty Handling Large and/or Complex Problems](#)
    - [5.5.1 Engineering in the large](#)
    - [5.5.2 Contribution of the lead users community](#)
- [6 Precautionary considerations section](#)
  - [6.1 Approaches](#)
  - [6.2 Practical organization](#)
  - [6.3 Conceptual terminology](#)
- [7 Proposed Precautionary oriented questionnaire](#)
  - [7.1 Evolution of this questionnaire](#)
- [8 Security considerations](#)
- [9 IANA considerations](#)

## Introduction

The IETF Precautionary duty consists in responsibly considering the risks that may be created, on the Internet architecture and its deployment, international relations, global economy, cultural, societal, natural environment, people's future, etc. by the influence of each of its engineering documents, independently from the high quality and technical relevance that they must achieve.

The IETF adherence to this Precautionary cardinal principle is implied as a requirement or as a need in the RFCs that define the mission of the IETF [[RFC 3935](#)], as well as in the way it organizes itself [[RFC 2418](#)] and processes Internet standardization [[RFC 2026](#)], its research and development objectives [[RFC 3869](#)], and its professional ethic in approving its documents [PROTO project] and considering the needs for improvements that it itself has identified [[RFC 3774](#)].

There is a general increase in users' concerns, from civil society, the regalian domain, private sector, and international organizations throughout the world, about technological precaution. Such a concern necessarily extends to the Internet, which has become a key structure for humankind. This Memo seeks to identify what precaution means exactly when documenting the Internet as the IETF does, and to summarize how the IETF addresses and progresses on this issue. It also seeks to call on the contribution of all those who may help the IETF in new ways towards a better Internet for every person on earth.

## Better safe than sorry

Everyone knows about precaution as a responsible duty for every individual. However, the recent technological development calls for a rethinking of this notion within the collective standardization process, global operations, and world governance.

## Definition

The Principle of precaution is an emerging societal, ethical, cultural, political, technical, and economic principle that states that if an action or strategy might cause severe or irreversible harm, in the absence of a scientific and implementation governance responsible consensus that harm would not ensue, the burden of proof falls on those who would advocate taking the respective action or following the respective strategy.

## Legal status

It is not a fully juridical principle, as it still can hardly provide regulations that are sanctioned by laws, or describe what actions to take. In most cases, it seeks to trigger reactions in advance, before any irreversible damage occurs. However, in some legal systems and international agreements, the precautionary principle is a general and compulsory principle. The principle of Precaution appears as a new mode of collective or global action or governance, which is often associated with subsidiarity and proportionality principles.

In addition, many countries choose to consider consumer points of view, and media reporting, to create a new space for debate, where politicians, experts, and journalists are answerable to other actors (e.g. consumer associations, juridical authorities, etc.).

### **Areas of applications**

The precautionary principle reflects the belief that when there is scientific or a practical deployment, uncertainty concerning a project, an action, or a standard and their consequences, preventive measures may be justified.

This principle is often invoked when the consequences are considered sufficiently great to require expensive amelioration, even when the risks are considered low. In the Internet case as well as in Internet direct or indirect usage areas this will be invoked where scientific evidence is insufficient, inconclusive, or uncertain and preliminary scientific evaluation indicates that there are reasonable grounds for concern that the potential dangerous effect on the network, human, economic, cultural, and operational environment may be inconsistent with the level of quality and protection that is chosen by users or their country.

### **Innovation ethic**

The concept includes an implicit ethical responsibility that is assumed by engineers, governance, lead users, and every informed person, entity, or process towards maintaining the integrity of the common technical and natural systems and acknowledges the fallibility of human understanding. It is very technically familiar to the Internet operators who are, for example, constantly confronted with the DNS system pollution and mismanagement risks and the real-time precaution duty that this requires.

As such, the application of the principle modifies the status of innovation and risk assessment: it is not the risk that must be avoided or amended, like when considering security, but rather a potential risk that must be prevented. Thus, in the case of the operations of scientific and engineering research and development, there is a third party, beyond the technical scientist and engineer side, and the regulator and operator governance: the consumer, or the user in the Internet case.

### **Enacting the Precautionary principle**

The precautionary principle is not an absolute rule. It should be first accepted as a cultural concept helping to clarify arguments, in particular to determine where the burden of proof lies between the designers of a new protocol or architectural proposition, the operators, and the users.

### **When does the principle apply?**

It may apply everywhere, as a perfect protocol does not necessarily deploy perfectly. Two conditions should lead to precautionary considerations:

- a threat or a user's fear of serious or irreversible damage, pollution, or confusion.

- a scientific and operational uncertainty as to the extent of that respective possible damage.

Once these two conditions have been addressed, precautionary measures can still be taken or advised, but they should be proportionate.

The principle should not be used to try to avoid all the risks or to block a competitive project.

### **Investigating a possible case of application**

Five points should be considered:

- whether there are scientifically or operationally rational grounds for the concern.
- the extent of the threat: local to an application, network wide, application level, cultural or economic impact, etc. and human lives.
- the perceived value of what is put at risk, the future being more important than the past.
- if the impact can be managed or not in an acceptable manner.
- the level of public and user concerns.

### **Attitude to be adopted**

Evaluation of the technical, scientific, and operational level of uncertainty should consider what would constitute sufficient evidence; the level and kind of uncertainty, the potential to reduce uncertainty; the possibility of locally or time contained experimentation; previous experience, etc.

If the principle applies, it shifts the burden of proof and legitimacy.

- The decision maker must assume that the threat is real and that in making it accepted by the concerned parties that it is negligible or sufficiently compensated reverts back to him or her.
- The concerned parties are permitted to take preventive measures without having to wait until the reality of the seriousness of the threat becomes fully known. The more significant and uncertain the threat is, the more the precaution should be required.

Decisions applying to the precautionary principle must be open, informed, and democratic and must include the affected parties.

### **IETF mission, cardinal principles, and the standard process**

The precautionary principle is implied at the core of the cardinal principles that the IETF is in adherence to when it carries out its mission. This mission is to produce, along the Internet standard process [[RFC 2026](#)], high quality, relevant technical and engineering documents that influence the way people design, use, and manage the Internet in such a way to make the Internet work better [[RFC 3935](#)] in every part of it, as a whole, and for everyone.

### **Cardinal principles**



These documents describe protocols and practices for which secure and scalable implementations of a certain impact, consistency, and coherence are expected to have wide deployment and interoperation on the Internet in order to form a part of the infrastructure of the Internet, or to interoperate with technologies that are designed outside the IETF.

- **Open process:** any interested person of the IETF compatible language and culture can directly participate in the work, know what is being decided on, and make his or her voice heard on the issue.
- **Technical competence:** the issues on which the IETF produces its documents are issues where the IETF has the competence that is needed to speak of them, and the IETF is willing to listen to technically competent input from any source. Where it does not have, and cannot gather, the competence needed to make technically sound standards, it should not attempt to take on the leadership role.
- **Volunteer Core:** the IETF participants and leadership are people who come to the IETF because they want to do work that furthers the IETF's goal of "making the Internet work better" in every part of it, as a whole, and for everyone.
- **Rough consensus and running code:** IETF makes standards based on the combined engineering judgment of its participants and their real-world experience in implementing and deploying its specifications.
- **Protocol ownership:** when the IETF takes ownership of a protocol or function, it accepts the responsibility for all the aspects of the protocol, even though some aspects may rarely or never be seen on the Internet.

### **The standard process**

The IETF standard process has been optimized over time to permit the production of high quality, relevant, technical, and engineering documents concerning the design, use, and management of the Internet in such a way as to make it work better: it is now also necessary to make it interoperate better with other systems, as a whole, and for everyone. This is a broad part that is obtained through, and due to, the quality of the Internet standard process.

This process is characterized by at least three main aspects: the criteria used when selecting new work, the charter defining it, and the progression of the work from WG establishment to the final approval by the IESG. Different precautionary aspects are embedded or are still to be ascertained.

#### **Criteria for selecting new work**

When determining whether it is appropriate to engage in new work, promoters, Area Director(s), and the IESG consider the following points that are precautionary in terms of selecting a good topic, but also that is the best for managing the IETF resources and voluntary participants' time and efforts.

- Are the considered issues clear and relevant to the Internet community, achievable within a reasonable timeframe, and are there risks and urgency for the work?
- Is there an overlap with other currently engaged work, and sufficient and broad enough available interest within the IETF and expertise in the topic?
- Does a base of interested consumers (end-users) appear to exist for the planned work? This is where a precautionary attitude will attract the interest of lead users and of consumer organizations?
- Is the IETF the best suited place to take the lead in the "real world"?
- Are all the known intellectual property rights that are relevant to the proposed work issues well understood?
- Is the proposed work an open effort or an attempt to "bless" non-IETF technology? Is there at all (inside or outside the IETF) a good understanding of what is technically and operationally relevant to the proposed topic: it will be a starting point for any precaution related debate.

## Charter

The formation of an IETF working group requires a charter to be approved by the promoters and the IESG with advice from the IAB. A charter is a contract between a working group and the IETF to perform a set of tasks. This notion of a contract will be of the essence in further precautionary work and considerations in order to make sure that everyone has the same understanding of the expected influence of the document being worked on/out.

To facilitate the evaluation of the intended work and to provide ongoing guidance to the working group, [RFC 2026](#) states that the charter must describe the problem being solved and should discuss the objectives and expected impact with respect to: architecture, operations, security, network management, scaling, and transition (where applicable). This list should also include concerns about the network, international, cultural, economic, strategic, technological, etc. environment.

The IETF acknowledges that the proposed working groups often comprise technically competent participants who are not familiar with the history of Internet architecture or IETF processes. This can, unfortunately, lead to good working group consensus about a bad design. To prevent this, Area Directors may assign a Consultant from among the ranks of senior IETF participants. This practice should extend to User senior experts in the non-Internet areas that may be impacted. An example is the direct participation of the ICANN Staff to the WG-IDNABIS.

## A way to address identified IETF difficulties

The precautionary principle deals with the problem of introducing innovation in any environment. In documenting network elements, the IETF constantly faces that kind of situation. This is why it turns out that this principle may help the IETF to better approach, and sometimes seamlessly address, difficulties that it has identified over the years but where its participants have not yet perceived the roots.

## The Balance Between Research, Invention, and Adoption

The IETF has traditionally been a community for experimentation with things that are not fully understood, the standardization of protocols for which some understanding has been reached, and the publication of (and refinement of) protocols that were originally specified outside the IETF process.

The first precaution is to decide whether or not these activities should be performed within the IETF. Then, one should not chiefly look at the type of activity, but rather the potential benefit to the Internet system and the Internet community - an experiment that yields information about the fact that an approach is not viable could be of real benefit.

### **The Reach of the Internet**

The people interested in the Internet's evolution are from every culture under the sun and from all walks of life. The IETF puts its emphasis on technical competence, rough consensus, live testing and operations, and individual participation. It, therefore, needs to be open to competent input from any source whatever the language. The IETF uses the English language for its final work and documents. This may raise users concerns, provoke a lack of trust, and lead to conflicting alternative usage developments. The IETF must, therefore, as a precaution try out imaginative multilinguistics propositions (multilinguistics is understood here as the methodology to address the diversity of the phenomena of linguistic diversity). This is why it should consider how to best benefit, as in ISO, from the multilinguistic adage: "semantics add, pragmatics subtract".

### **Funding the Internet R&D**

There are three observed sources of network R&D:

- the academic effort
- the commercial funding sources that more likely fund the research that leads to a direct competitive advantage.
- the Internet lead users that are partly visible through external innovation the IETF uses further on [\[RFC 3935\]](#)

This is because no single organization (e.g., no single government, software company, equipment vendor, or network operator) has a sense of ownership of the global Internet infrastructure, research on the general issues of the Internet infrastructure are often not adequately funded.

### **Commercial funding**

If there was no commercial funding for Internet research, then few research projects would be taken to completion with implementations, deployment, and follow-up evaluation. However, the IAB has documented [\[RFC 3869\]](#) that if commercial funding is the main source of funding for future Internet research, the future of the Internet infrastructure could be in trouble. In addition to issues about which projects are funded, the funding source can also affect the content of the research, for example, towards or against the development of open standards, or taking varying degrees of care about the effect of the developed protocols on the other traffic on the Internet.

## **Precautionary alternative to commercial funding**

This is why self-precaution calls for the IETF to consider all of what can be done together with the academic and lead user communities to make them more focused on research concerning the basic, applied, and usage architectures of the Internet and be more productively listened to. This might also attract governments and non-commercial sponsors, as they themselves are specialized innovative users, so that they might sustain the necessary FLOSS funding and join forces with the lead users, who are more interested in people/SNHN (Small Network/Home Network - Standalone Host) centered innovation.

## **Difficulties of the IETF in accomplishing its Mission**

The IETF has now a clearly defined and commonly understood Mission [[RFC 3935](#)]. Its Mission Statement does not document the Precautionary principle, but it does imply it. This has many positive consequences. However, some identified problems [[RFC 3774](#)] remain that a broader consideration of the Precautionary principle might help to address.

- Working Groups can potentially be hijacked by sectional interests to the detriment of the IETF's mission. Embedding Precautionary considerations in the Internet process (WG work) and documents (Drafts, PROTO evaluation, RFC) can certainly help to reduce that risk, in which the responsibility of the non-hijacking falls on the hijackers themselves.
- It would be desirable to have road maps and architectural views for portions of work that extend beyond a single working group: it may also be the case that it is no longer possible to fit the whole Internet within a single architecture. This is something the external point of view of lead users and other SDOs should help to appreciate through the expression of their own Precautionary concerns, most probably offering cross-fertilized views of the "good of the Internet" and how to make the Internet useful to the greatest number of stakeholders for the long-term. The resulting dialog can only help a general debate where the IETF itself may in turn benefit from the application of the Precaution principle by other SDOs and users.
- One of the requirements of the Precautionary principle is for the IETF to issue Precautionary considerations concerning its own engineering choices and practices, covering both the techniques that are used to derive and verify the technical solutions needed, and the management and organizational strategies that are expected to help with the engineering process. This includes points such as:
  - timely, high-quality, predictable outputs, respecting the WG Charters and timelines.
  - methodology to ensure that there is a uniform view in the WG of the scope of the WG activity, especially over the intended purpose of the solutions wherein the Precautionary considerations may have to document the state-of-the-art scientific consensus of negligible risk.
  - the WG reflection road-map that leads to that conclusion.

- the identification and articulation of the engineering trade-offs that have been adopted without inappropriately reducing the 'fitness for purpose' for the intended users.
- the conclusion of adequacy with no further possible refinement in order to meet the requirements that are placed on it by the intended purpose that is documented in the charter.

This may call for more detailed and respected charters, WG commonly agreed upon road-maps, detailed plans, timelines to be followed, audits on users' explicit criteria for the concerned standard development, early reviews and dialogs with users directly or indirectly interested in the immediate focus of the documents, and questions about the problem areas that might arise due to interactions with other popular IETF protocols.

This should give a clear and simple incentive that will help to transparently address the problems listed in [RFC 3774](#). The users' dialog should provide metrics to measure the achievement of the desired quality and performance of both WGs and the whole IETF: the shorter the users' wish list, the fewer users who would have to post comments and the higher their trust, along with the document's probable quality, and the better the Internet should work.

Direct cooperation with lead users, should also enable the parallel testing of running prototype code to verify the suitability of the documented propositions. This should also ensure improved dialog of the WG and of the interested lead users, and probably - through them - with other WGs.

### **The IETF has Difficulty Handling Large and/or Complex Problems**

As [RFC 3774](#) states it, the IETF has historically been most successful when dealing with tightly focused problems that have few interactions with other parts of the total problem solution. Given that the Internet has become more complex, such tightly focused problems are becoming the exception. IETF standardization procedures are optimized for tightly constrained working groups and are generally less effective if 'engineering in the large' is needed to reach a satisfactory solution.

The IETF does not always seem to be aware of the interactions between the protocols that are bound to be thrown off by deployment in more complex situations and thereby the IETF fails to minimize the chances of unwelcome consequences arising when a new protocol is deployed. This is precisely where the Precautionary principle has a key role to play, since the duty of precaution consists in making sure that such unwelcome situations do not arise.

#### **Engineering in the large**

As we have learned more about the ways in which systems on and outside the Internet interact, the design of components needs to take into account increasingly external constraints. The understanding of these constraints tends to require more engineering in the large.

Engineering in the large can encompass many aspects of system design including:

- Architecture
- Frameworks
- Security
- Internationalization
- General systemic modeling, issues, and precaution.

It also appears that handling large and/or complex systems calls for a deeper focus on a narrower range of issues (for practical reasons) with a wider understanding of the system and the context in which they will be used or in the way that they will be supported.

There are two main ways to practice this engineering in the large:

- by way of theory, specialized studies, and inter-SDO relations: this is the role of the IAB, which is suited to represent the deep interests of the technology's core.
- by way of experimentation, usage, and in the field adaptation: this is the increasing role of the "lead users" who are directly confronted with the demands of the real world's diversity and to its permanent operational evolution.

#### **Contribution of the lead users community**

A lead-user is a user that adapts what he or she purchases to the real needs at hand rather than adapting his or her needs to what is available on the market. This adaptation may sometimes amount to an entire redesign or brand new development. The FLOSS phenomenon is a good example. Lead users usually represent between 0.1 and 2 percent of a customer population, depending on the technological level of a sector and the cost of the engaged upgrading. They become, in some areas, a heavy industrial trend, where the "market" actually sub-contracts its solutions to industry.

In the Internet viral environment, they represent a strong capacity for innovation and contribution. The IETF has not been able so far to take full advantage from them. However, the IETF was initially lead-users engineering group. One of the reasons of the disconnect is that, for a time, lead-users were mostly small entrepreneurs with market priorities. This time is over as there is more and more motivation for people to drive their own "cybercraft" and become "home lead users". Another reason is their more user centric distributed vision of the network, which prevails now in the WSIS framework, while the IETF's vision is still more network centric. There also are some remnants of the "alt-root" / "open-root" issue about the management of the virtual DNS root. The multilaterality of a systemic understanding of the Internet helps to merge these visions: precautionary concerns demand their dialog since it actually consists in the IETF convincing users that its solutions do not harm them, which could turn out to be an excellent quality test for the IETF deliverables.

The difficulty is a practical interface problem because visions, languages, priorities, time-to-usage, practices, and operational needs all differ. There is, therefore, a need to build, provide, and maintain such interfaces. The IUCG@ietf.org mailing list and its attached site and resources have been created to experimentally explore that possibility in two connate directions: the community of the users of the IETF documents, and the community of the lead users of the Internet that documents influence the design, use, and management.

Based on experience, the target should be to replace the technical/process appeal procedure. When an individual user thinks that the Internet standard process has produced a result that is harmful to the Internet or thinks that the IETF processes have not been adhered to, there should be a process to aid that individual in seeking to change that result, or to possibly even prevent the outcome in permitting the IETF community to understand him or her better and possibly benefit from users' conceptions and experience.

## **Precautionary considerations section**

In order to prevent this kind of conflict without destabilizing their own work, WGs or authors can choose to include a Precautionary considerations section in their IETF document.

## **Approaches**

This may resort from two general approaches:

- to explain that the IETF consensus, as verified by the operators' community, considers uncertainty and risks to be negligible or adequately compensated compared to the expected return.
- or to detail the risks and uncertainties that remain to be solved or that may have to be met under certain circumstances – such as, for example, security violations documented in the Security consideration section, technological transitions or interfacing, governance policies. or usage trends.

## **Practical organization**

WG Chairs and WG Members may be confronted with two kinds of equally embarrassing minority situations:

- either there is an opposition to the consequences of the WG choices but no alternative proposition.
- or there is one or several alternative propositions to the WG choices.

In both cases, one or several "design teams" should be created. In the first case to document a list of points calling for clarification or a consensual IETF precautionary position, comforted by the Operators' community. In the second case, to try to reach a multi-consensus where each position should be made interoperable et published with precautionary comments by its opponents.

## **Conceptual terminology**

Usually an IETF document treats an individual system, concept, protocol, black box, etc. Precautionary considerations necessarily belong to a systemic global vision of interpenetrated systems. In order to help differentiate the parts of the involved systems in a multiple system network context, the following concepts may be used:

- endotem: is the system internality. This is what the document is usually about. Its coherence and stability have been discussed at length by the WG, designers, and reviewers.
- exotem: is the rest of the world system, which is external to the considered system and that may be polluted or irreversibly changed or harmed by such a considered system.
- peritem: is the interface between the endotem and exotem. Security considers its break-ins. It is to be noted that in an interpenetrated system (such as a virtual network, an inter-governance mechanism, a distributed architecture, etc.) the peritem may have top level discontinuities. In such cases it should be extended down to the first continuous layer.

## **Proposed Precautionary oriented questionnaire**

For consistency purposes, this section builds on the "RFC-to-be draft-ietf-proto-wgchair-doc-shepherding", current template. Its aim is to help every IESG and IAB Member, IETF Participant, and Internet user to share the same precautionary logic based on a common questionnaire, which can be maintained online.

- (1.a) What are the positions and who are the Shepherd (IESG) or Mentor(s)(Users) for this document? Do they believe that this version of the document is ready for for final review and publication?
- (1.b) Has the document been adequately reviewed from key WG members and from key non-WG members, including Users? Is there any concern about the depth or breadth of the reviews that have been performed? Have such concerns been expressed along the regular open Internet standard process (other oppositions should not be considered as not prepared enough).
- (1.c) Are there concerns that the document needs more review from a particular or broader perspective, e.g., architecture, security, operational complexity, someone familiar with AAA, multilingualism and internationalization, system theory, operational deployment, usage governance, XML, and ISO standards?
- Does the document respect the Charter and the resulting WG's road map? If not, why?
- (1.d) Are there any specific concerns or issues with this document that one should be aware of? For example, uncomfotability with certain parts of the document, or concerns over whether there is really a need for some. In any event, if the WG has discussed those issues and has indicated that it still wishes to advance the document, this should be detailed and answered. Has an IPR disclosure that is related to this document been filed? If so, the reference to the disclosure should be noted and summarized in the WG discussion and conclusion of this issue.



- (1.e) How solid and large is the WG consensus behind this document? How construed, shared, and deep are the concerns raised by this document? Are there existing or investigated alternative approaches?
- (1.f) Has anyone threatened an appeal or otherwise indicated extreme discontent? Is the position documented? Are the alternative or complementary propositions documented?
- (1.g) Is the document presented clearly enough for Users to understand it, as well as for the RFC Editor to publish it?
- (1.h) Has the document split its references into normative and informative? Are there normative references to the documents that are not ready for advancement or are otherwise in an unclear state? Are these normative references from the IETF, or by other SDO(s)? If such normative references exist, what is the strategy for their completion and expected publication date? Are there normative references that are downward references, as described in [RFC3967]?
- (1.i) Does the document have an IANA consideration section and is it consistent with the body of the document? If the document specifies protocol extensions, are reservations requested in the appropriate IANA registries? Are the IANA registries clearly identified? If the document creates a new registry, does it define the proposed initial contents of the registry and an allocation procedure for future registrations? Is the IANA registry created to be interoperable with external or Users' registries that are being (or will be) used on the Internet. What are the procedures to protect clear interoperability?
- Are there usage oriented tables or information to be gathered, supported, and maintained? Are their metadata (and possible additional attributes) documented online?
- (1.j) Do the sections of the document that are written in a formal language, such as XML code, BNF rules, MIB definitions, etc., validate correctly in an automated checker? If new terminology is being used in the document, is it consistent with other (multilingual) terminology(ies) in use in the international and Internet standards?
- (1.k) The IESG approval announcement implies some bureaucracy . Are there concerns related to it?
- (2.a) Which Internet Modeling perspective does the document use? Is it adequate? Is it documented enough? Does it benefit from reasonable support?
- (2.b) Does the document include a "Precautionary consideration section"? Is it coherent with the "Security consideration section"?
- (2.c) If the Precautionary considerations section recommends precautions, can they be deemed as complete and sufficiently documented?

- (2.d) If the Precautionary considerations section concludes that possible risks, uncertainty, or harm that are created by the influence of the document are negligible: is it convincing? Does the IESG and IAB formally support its conclusions? Would an appeal against them succeed or fail?
- (2.e) What would be, if any, the alternatives to the document proposition, their impact, delay, and degree of reliability? To which extent do they imply or require an architectural model change?
- (2.f) What is the timed deployment strategy? Is the proposition worth its cost and delays? Does it call on external cooperation? Were the providers of that cooperation involved in the document's preparation?

### **Evolution of this questionnaire**

This questionnaire should evolve with the Internet architecture and with the digital convergence in order to consider the technological and usage interface of the datacommunications strata the Internet technology belongs to with the telecommunications (hardware, signal) and the metacommunications (semiotic, meaning and semantic coherence) strata.

### **Security considerations**

The Precautionary domain is an environmental extension of security. Failing to consider it would be a major security violation.

### **IANA considerations**

The main Internet issue, until it transitions to an MDRS (Multilingual Distributed Referential Systemic) permitting the multi-dissemination of the user-centric metastructure necessary to support the Multilingual and Semantic Internet, is the precaution of protecting the IANA independence from political, strategic, commercial interests.

This Memo gives some hints on the way to proceed, but did not intent to discuss the non-technical Internet Governance issues.

## Annex 2.

### Is IDNA an ML-DNS to be?

These mails belong to one of the first exchanges of the WG/IDNABIS.

#### At 03:29 17/05/2008, JFC Morfin wrote:

Dear Vint,

As I explained it, prior to sending this mail, I circulated its draft to many people round the world and took their remarks into consideration, specially @large members who like those of france@large and Multilinc translate "@large" as "an Internet co-owner", and MAAYA Members.

The IDNA issue is a key priority for the continuation of the IETF technology as the network technology of the world digital ecosystem (WDE). If the IETF cannot match the world's expectations in that area it must say so now, so others can consider alternative solutions before we see different uncoordinated local solutions developed and deployed.

*>At 22:37 09/05/2008, Kenneth Whistler wrote: What matters, I think, is what contemporary communities \*are\* using or might reasonably be inferred to want to use if available, for domain names.*

Maybe, what really matters are:

- (1) the political world consensus for a multilingual Internet for a people centered multilingual society of information,
- (2) the universal resentment at being told what one is supposed to want. People accept being told by others as to what they can only do, if they understand exactly why.

The world expects a Multilingual DNS that works for every language and every script the way the DNS works for English and ASCII. Let us call this the ML-DNS specification. It is very simple, terse, and clear.

**Question (A):** does this IETF WG-IDNABIS seek to document an IDNA based ML-DNS in order to be ready for testing by Dec. 2008 (Y/N)?

**Question (B):** If A is "N", what are the clearly defined and committed detailed specifications of the Nov. 2008 IETF deliverable?

Among the points to be clarified in these specifications are:

1. - will it be mainly focussed towards Mobiles, Browsers, Applications, or the three of them?

2. - will it be phishing proof at every DN level?
3. - which scripts or charset and languages will be supported? or will it be transparent to scripts choices?
4. - will it be IDN2003 compatible?
5. - will it strive to be future ML-DNS interoperable?
6. - why was the IDNA option chosen as the best way to support ML-DNS vs. other possibilities?
7. - will Microsoft, Google, and Firefox fully and identically support it? Will they also permit the support of any other ML-DNS proposition?
8. - will it support easily additional symbols such as logos?
9. - will it stay ISO 3166 conformant?

Thank you for your committed answer.

### **At 04:17 18/05/2008, Vint Cerf wrote:**

Jefsey,

you forget perhaps that the internet is based on voluntary adoption of standards. IETF does not and cannot enforce; and ICANN has limited means to do so through SOME contracts with SOME parties but not at all levels of DNS nor resellers nor some other operators.

so we just have to write the best technical specifications we can and make them available.

vint

### **At 01:47 19/05/2008, JFC Morfin wrote:**

Dear Vint and all,

It seems that we are all 100% in agreement, except for perhaps James Seng who wants to read things that no one ever wrote. I apologise for coming back to this. I hope that it will finally be the last time.

**At 01:04 18/05/2008, Vint Cerf wrote:** Let me try to be very clear.  
*I do not think that we should conflate the IDNA work as chartered with the broader question of ML-DNS as it is becoming apparent that ML-DNS (what ever might be meant by that) does not seem to fit into the present charter.*

This is exactly what we wanted to see clearly spelled out. Because, some of us - per my suggestion - thought from reading the Charter and John Klensin's documents that a succesful IDNA could be targeted as a good ML-DNS and that we did not want to disrupt its respective process.

*IDNA-bis is intended to refine the basic IDNA mechanisms based on experience to date with the earlier IDNA2003 formulation. IDNA-bis is not chartered, e.g., to invent new classes of domain names.*

We are in full agreement. And no one ever suggested this. Our two year ICP-3 community test bed (dot-root) plus our three years of exploration on semantic national protection (AFRAC) and a parallel two years on a multilinguistics investigation have shown us that the ML-DNS does not need to use classes (as John Klensin suggested once). This could help, as would also a built-in presentation layer, but it is not necessary because classes are not unlimited.

*It seems to me that we are on a productive vector to finalize the four documents with which the working group began.*

Absolutely. And we fully agree with this. James' misunderstanding is, however, very positive as it enables this to be made absolutely clear. We fully support the WG-IDNABIS for what it is and certainly we want to contribute to its success as a good, adequate, compact, clear document set. Either it is good enough to satisfy most of our needs, or it will be easier to encapsulate within any ML-DNS strategy (whatever it may turn out to be). We do not have any solution as of yet. We only have needs currently.

*I have not heard any argument (that I was able to understand) that rejected the essential framework of IDNAbis. I AM hearing that some participants would like to consider alternatives to what we seem to be calling IDNA2008 or IDNA200X now. I am not hearing any kind of consensus that we should abandon IDNA2008; rather, I am hearing a desire to finalize these documents. Much of the discussion surrounds the means by which sets of Unicoded characters are either accepted as PVALID or rejected as unacceptable for use in domain names.*

Full agreement.

After I explained to them the reasons for James' misunderstanding (he explains very well in his own mail) Louis, other French lurkers, and I believe that proposing to abandon/discuss IDNA2008 would be highly disruptive. Our worry, due to other past confusions and to the possible economic/architectural possible impact, was that our own possible investigation concerning an ML-DNS would not be confused as an alternative but as a more complete solution in a continuity that is totally out of the scope of the WG-IDNABIS.

I think that this mail of yours fully clarifies this for everyone.

1. IDNAbis must be a success.
2. It is not intended to be an ML-DNS, so working separately in parallel to the preparation of a possible ML-DNS BOF is encouraged.
3. There is no reason as to why ML-DNS should be discussed any further here. Answering Vint's suggestion, I have created an ML-DNS exploratory mailing list to that end.

*I would urge the participants in the IDNAbis Working Group to focus on refining and finalizing the four documents that have been submitted as the basis for the Working Group effort.*

We obviously fully concur.

1. We will continue to document the <http://wikidna.org> with information on/for the IDNA development, deployment and support. *\_Everyone\_* is welcome to come and give a hand.
2. The strategy of this site should be to support complete compatibility at the linguistic IDN level between IDNA and ML-DNS.

**At 01:27 18/05/2008, James Seng wrote:***Oh I understand the questions fine. And the subtle implication of the lack of usefulness of this WG.*

This may be your opinion. However, it is not ours.

*But I dont see a need to have the discussion since ML-DNS and questions are clearly out of scope. As Vint indicated, you or anyone is feel to pursue another working group or mailing list for it.*

There was not any reason to do it, if we did not need to. Now, you clarified that there was a technical need, and Vint said that we were not disruptive at having a try (we feared your misunderstanding).

This is precisely what we will do, as per the Internet standard process. Exploring through tentative Drafts, and aiming at a BOF.

We will also consider as to how to prepare an adequate legal IDNA acceptance by registrants and registry managers, because we are registrants and registry managers and we need to be fully interoperable with the best IDNA, which we want to help coproduce in this WG.

*I also didnt miss the suggestions that some groups may break off from IDNA without ML-DNS (whatever that is) and with reference to Olympic suggesting that Jefsey is somehow speaking on the behalf of the Chinese, however cleverly crafted that it imply he might also not be.*

???

I made it clear that I am speaking on behalf of some French and francophone @large and @large international community informed members, defining them as "Internet lead users that consider themselves as Internet co-owners".

*I have spend plenty of time recently in China as well as catching up with old friends from CNNIC. But there are CNNIC folks on this list and I am pretty sure they can speak for themselves if they have any objections without someone claiming to represent "@large" or myself even to speak for them.*

I thank you for clearly spelling out your belief that I am pretending to speak on behalf of CNNIC.

This clearly shows the exact origin of your misunderstanding: I am not speaking on behalf of CNNIC. Moreover, I am not pretending to speak on behalf of CNNIC, and I do not see what CNNIC has to do with this at all. I do not see why I could be interested in speaking on behalf of the CNNIC??? They already have an IDNA based solution that, therefore, is not an ML-DNS solution (according to you who is the expert).

I also fail to see what objection we could have. We ask a question?

***At 04:17 18/05/2008, Vint Cerf wrote:***

*Jefsey, you forget p'rhaps that the internet is based on voluntary adoption of standards. IETF does not and cannot enforce; and ICANN has limited means to do so through SOME contracts with SOME parties but not at all levels of DNS nor resellers nor some other operators. so we just have to write the best technical specifications we can and make them available.*

I did not forget. I am just pragmatic and I have to convince people that are not familiar with the IETF that this the way it works. I thank you for spelling it out: they certainly felt relieved. James Seng's reaction has shown that only asking "if IDNA should or should not be considered as an ML-DNS solution" was a sensible conflicting issue for some. This is what had to be clarified first, both at this WG and among our own @larges.

No one wants confusion or conflict. As Louis explained it, @large people are not really happy about the response that IDNA cannot be considered as an ML-DNS and that they must have a try at it. Furthermore, they did not want to be confused with opponents who they consider as their back-up solution, in which they certainly want to contribute if they can.

I hope this definitely clarifies the issue.

## Annex 3.

### A practical innovation and complex deployments strategy

The current work of WG-IDNABIS leads one to think that IDNA might exemplify a practical IETF innovation strategy towards a stable and faster deployment in spite of the complexity of the Internet system.

#### Contents

- [1 Introduction](#)
- [2 Rough model introduction](#)
- [3 The model gap](#)
  - [3.1 The three ways to fill the gap](#)
  - [3.2 The IETF options](#)
    - [3.2.1 There is plenty of architectural room for innovation](#)
    - [3.2.2 There is not so much room](#)
    - [3.2.3 There is insufficient room](#)
- [4 Taking advantage from the difficulty](#)
  - [4.1 Indirect enforcement](#)
  - [4.2 How to obtain it](#)
- [5 The example of the WG-IDNABIS case](#)
  - [5.1 Users' ML-DNS concept contribution](#)
  - [5.2 The practical risks](#)
  - [5.3 The so-called impossible multilingualization of the Internet](#)
  - [5.4 Other areas for a similar strategy](#)
- [6 Security Considerations](#)
- [7 IANA considerations](#)

#### Introduction

The Extended Relational Space Model (XRSM) that the IUCG (Internet Users Contributing Group) is working on, has a distributed, systemic, and user centric network architecture approach that could extend the [RFC 3935](#) core values. A first draft has been submitted on the precautionary principle. and the general introduction work has been engaged, which should help better understand the positions that will be used in its framework, and the resulting characteristics of its "Internet PLUS" model.

However, this work requires time and reflection. The current WG-IDNABIS debates show that some earlier considerations could be of use in order to explore how the IDNA situation



might lead to, and exemplify, a practical strategy to obtain a stable, innovative, and faster deployment of the IETF propositions in spite of the Internet system complexity.

## **Rough model introduction**

Among the other fundamentals, the XRSMModel considers three main strata in e-semiotic (usage of signs to utter semantic meaning) communications:

- physical, plug to plug interconnectivity - the telecom strata.
- logical, end to end interoperability - the datacom strata.
- nootical, mind to mind interintelligibility - the metacom strata.

The "decentralized, end to end, dumb network, protocol on the wire" Internet makes up most but not all the logical strata. The DNS includes some semantic aspects that belong to the nootical strata. Jon Postel identified it when he forbade namedroppers from discussing and confusing semantic and protocol issues (at that time they discussed the English meaning of the TLDs tags).

## **The model gap**

Along this model, there is a gap to fill between the top of the Internet and the bottom of the Intersem (Multilingual and Semantic Internet) metacom layers. This gap minimally concerns:

- the missing Internet session and presentation layers.
- extended services that the OSI model did not consider and that were included in the first architecture of the international network.
- new requirements that we (will) learn from the Intersem analysis and testing.

## **The three ways to fill the gap**

There are at least three ways to fill that technology gap.

- The usual way of the IETF, which is in charge of the entire existing billion systems Internet legacy. It is detailed below.
- the various proprietary ways, integrated services providers or alliances (to differentiate from basic access service providers), will naturally push forwards through business competition.
- the way lead users call "Internet PLUS", standing for "parallel layered usage systemic". It consists, for each layer, in proposing those users with one or several alternative modules that extend the IETF architecture parts the user needs. This layer (alternative) extended building block approach will probably be more agile because it is not bound to constraints that the user may not be interested in, such as backward compatibility or interoperability that he never needs. Internet PLUS building blocks will, otherwise, have to be completely interoperable with the IETF architecture and solutions. Yet, they will add extra features, the total pile of which may result in the missing services.

## **The IETF options**

The IETF architecture is usually confronted with one of the following three cases:

### **There is plenty of architectural room for innovation**

This can result in a network technology enhancement if the three (IETF, ISP's, and User's) approaches described above can agree upon a solution.

### **There is not so much room**

Therefore, a new deployment is usually constrained on a short-sighted basis (cf. [RFC 3968](#) rationale by IAB): large commercial ISPs are then the probable way to finance and deploy. This will be in the hope that the situation will unlock further on, one way or another. This may result in long term detrimental impacts and major network unbalance, or even balkanization. The users, being less represented, cannot oppose this easily. This is even more so the case if the propositions are confuse and call for long debates. When propositions are confuse, and may bias the network context, the debate may become detrimentally controlled by some commercial interests.

### **There is insufficient room**

There is insufficient room within the IETF end to end decentralized network territory. Therefore, the IETF must go beyond the network border, in turn trespassing and venturing in the user application area, where it has no enforcement capacity and can only engage the credibility of its image, while its documents may be disrespected by some.

## **Taking advantage from the difficulty**

This last situation may turn out to be very interesting if it is adequately and carefully handled. The reason for this is that in such a case, the IETF usage application layer recommendations become legitimately enforceable and, taken as such, as the default specifications, the Internet PLUS applied answer will respect and impose via its usage. The enforcement capacity of the Internet PLUS is that by nature it remains within, but it also enlarges the network ends.

### **Indirect enforcement**

As a result, every user who chooses to adopt an Internet PLUS approach de facto enforces a mandatory respect of the IETF positions to those wanting to interoperate with his machines.

This does not prevent special usages in special usage cases by informed users (by essence, Internet PLUS users should be more informed or be specifically assisted users).

This additionally assists a progressive adhesion, transition, and technology enhancement to the point where ISPs will probably find their best interest in joining the bandwagon, in turn adding credibility to the whole proposition.

If the IETF proposition is inadequate, the Internet PLUS users will provide an earlier warning and probably more documented information.

### **How to obtain it**

The only constraint for the IETF is to make sure that it keeps a specification balance. This means that it should not

- make its specifications overly specialized in favor of any class of ISP interests, Internet PLUS developers, or users. Therefore, the specifications do not become overly restrictive for the other classes.
- over delay them, precluding Internet PLUS developers to develop their extensions as interoperable ones.

In the same way, IETF documents should not be overly general or become confuse. This would lead one of the categories to take the leadership in imposing de facto its own solution, all to the disadvantage of the others. This phenomenon is common and is known as "internationalization" vs. "multilateralisation" (multilingualization in the language area), as the unilateralization of one single standard instead of a multilateral normalization. This may help in the short term and survive the medium term but is likely to lead to earlier obsolescence and costly redesigns.

Actually, IETF WGs should consider publishing an explanatory rationale. This rationale could include advice for application and Internet PLUS module developers as well as for a best practice for users. This would prevent architectural rigidity, standardization tensions and governance conflicts.

### **The example of the WG-IDNABIS case**

Coming from the existing IDNA RFCs, the WG-IDNABIS discusses IDNA updates "for the Internet to work better", whatever the Internet governance may decide.

It is in this way that it trespasses the network ends and encapsulates the Internet into a sort of architectural "advised practice layer" that Zone managers and people can disregard, in turn risking the trust that users have in the IETF.

### **Users' ML-DNS concept contribution**

This is why the various ML-DNS projects may turn out to be the best way to implement and enforce IDNA and progressively turn the Internet into a truly Multilingual network. However, an ML-DNS architecture should be considered as a part of an Internet PLUS framework in order to be of full and quick use.

In this WG, the group of Internet Lead Users initially asked the Chair, on behalf of many others users from various language oriented mailing lists, about IDNA. They asked whether he intended to make it the "ML-DNS" (whatever it may be and that is to be discovered) that the WSIS expects, or to have a progressive strategy towards it.

## **The practical risks**

IDNA exemplifies the various practical difficulties that an IETF decided upon innovation may meet.

- a charter not taking into account ML-DNSs (whatever it may be that the users want) as an opportunity to adequately use it.
- delay in producing a stable document (a few months to possible several years of documentation instability)
- a set of complex documents.

## **The so-called impossible multilingualization of the Internet**

This was because one cannot bundle the Legacy Internet with multilingualization: Jon Postel understood it, [RFC 3536](#) does not even want to use the word, and Unicode is not ready for it.

This is why a user-oriented IDNA uncoupled yet interoperable multilingual ML-DNS (whatever the users want it to be) is of interest. In using it, the Internet user may immediately, progressively, optionally, support the Intersem. In turn, the Intersem will probably be an IDNA killing client.

## **Other areas for a similar strategy**

The same approach could most probably be used for v6 and DNS security.

## **Security Considerations**

The Internet PLUS is not an alternative technology. It is not even a technology evolution. It is a concept whereby when a functional extension is possible and further on needed in the pile, it can be implemented upon the users' decision, even if it locally reduces some global constraints, as long as its real usage (may be not its full capabilities) remains fully interoperable.

This concept may obviously raise security concerns and lead to a discussion about new security related solutions. However, such concerns and extensions should be discussed for each of them, one by one, prior to implementing each of them. Furthermore, its investigation includes the documentation of an "Interest" charter for the secure and responsible use of the Internet as its own test-bed.

## **IANA considerations**

The concept of Internet PLUS is based upon a distributed vision of the Internet systemic diversity. This also concerns the IANA that is to be used as part of a Multilingual Distributed Referential System. However, Internet/Internet PLUS interoperability calls for a strict respecting of the IANA tables and parameters.

## Annex 5.

### WG/LC IDNABIS report

The IDNA is a key issue for the IUCG as a paradigm for the future of the Internet. There is therefore a need to make sure its description document set reflects a complete IETF and users consensus. To help this, this memo keeps track of the WG-IDNABIS/LC requested and received answers.

### Contents

- [1 Introduction](#)
- [2 General appreciation](#)
- [3 IDN - IAB](#)
- [4 IDNA Definitions](#)
- [5 IDNA Rationale](#)
- [6 IDNA Mapping](#)
- [7 IDNA Protocol](#)
- [8 IDNA BIDI](#)
- [9 IDNA Tables](#)

### Introduction

The IDNA is a key issue for the IUCG as a paradigm for the future of the Internet. There is therefore a need to make sure its description document set reflects a complete IETF and users consensus. To help this, this memo keeps track of the WG-IDNABIS/LC requested and received answers. The IAB Draft on IDN has been added because some remarks have been made which are important to the background and to the future of the Multilingual Internet.

The author is quoted if the comment is not from IUCG. This compilation is updated up to the answer of Harald Alvestrand on 08/29.

### General appreciation

- The document repartition seems adequate. However, even if the Mapping memo was not a part of the IDNA (why?) document set, it is more than logical and enlightening to have it read prior to the Protocol parts.
- The documents are rather confusing because it is impossible to decide whether:

- they consider IDNA as a part or not as a part of the DNS (we may also be influenced by the ML-DNS pile we work on).
- they differentiate (which) between characters and codepoints.
- they use NFKC or NFC, and what are their differences, intrinsically and from an IDNA point of view
- they want to be a complete standards, or a partial suggestions, set. This results from:
  - the non-normative forms are being used in places that one would deem normative
  - the constant discussion of Registries' capacities/obligations and the lack of documentation on the tools for executing them and managing the related registration/coding metadata and rules.
- (Martin Dürst): - Use only one name for talking about the document collection. Currently:
  - 'collection' (Abstract, 1.1)
  - 'series' (1.1)
  - 'set' (1.3)
  - 'and the associated ones' (1.1.1)
  - 'these documents' (2.1; very unclear when reading whether that phrase indeed refers to the document collection or to Unicode documents or what, similar again in 2.2)

This variability is confusing.

## IDN - IAB

- (Martin Dürst): Mentioning ISO-2022-JP for encoding Japanese domain names raises some suspicion. ISO-2022-JP may well be (or have been) used in the DNS or a similar system, but such use would be atypical, and should be documented by a reference. Based on the general "division of labor" of the three classical Japanese encodings (ISO-2022-JP, EUC-JP, Shift\_JIS), one would expect EUC-JP or Shift\_JIS rather than ISO-2022-JP in such a case. [Among the three, ISO-2022-JP makes it easiest to explain the "heuristic encoding detection" scenario described at the end of Section 1.1. But without a reference, it may look to some as if ISO-2022-JP was a made-up example.]
- 1. "An Internationalized Domain Name (IDN) is a name that contains one or more non-ASCII characters.". What is a "name" vs. a domain name?
- 1. "When an IDN is encoded with Punycode, it is prefixed with "xn--", ". Is an IDN a label?
- 1. "it is prefixed with "xn--", which assumes that ASCII names do not start with this prefix." Isn't the whole thing supposed to use "xn--" ASCII labels instead of non-ASCII entries?

- 1. "reversible Unicode-to-Punycode conversion .... reversible Punycode-to-Unicode conversion". Unicode is a table, Punycode is an algorithm. Punycode is not reversible, but its use can be restricted to codepoints in turn permitting it to perform reversibly.
- 1. "UTF-8 [RFC3629] is a mechanism for encoding a Unicode character in a variable number of 8-bit octets, where an ASCII character is preserved as-is." Characters belong to scripts that may or may not be supported by ASCII and Unicode encoding tables.
- 1.1. (Martin Dürst): For the bulleted list at the end of Section 1.1, it should be pointed out that UTF-8 can be detected, and distinguished from other 8-bit encodings, with much higher precision than just "a byte in the string has the 8th bit set". For details, please see <http://www.ifi.unizh.ch/mml/mduerst/papers/PDF/IUC11-UTF-8.pdf>.
- 1.1. (Martin Dürst): The heuristic for punycode that is given in Section 1.1 is "starts with xn--". However, on the level of getaddrinfo, we are dealing with domain names, not single labels, and something like www.xn--foo.jp should definitely be punycode even if it doesn't start with xn--.
- 1.1. "When used with a DNS resolver library, IDNA is inserted as a "shim" between the application and the resolver library": that shim should be located in Figure 2.
- 1.1. "Again this assumes that ASCII names never start with "xn--", and also that UTF-8 strings never contain an ESC character". It may be worth documenting whether "IDN", "names", "labels", and "strings" are or are not the same thing.
- 2. "Applications that convert an IDN to Punycode before calling getaddrinfo() will result in name resolution failures if the Punycode name is directly used in such protocols. " Why? Is that not actually due to the reason that was given in 3? "Applications that convert an IDN to Punycode before calling getaddrinfo() will result in name resolution failures if the name is actually registered in a private name space in some other encoding (e.g., UTF-8)."
- 3. "While implementations of the DNS protocol must not place any restrictions on the labels that can be used, applications that use the DNS are free to impose whatever restrictions they like, and many have." Wouldn't these two rules contradict the proposed WG-IDNABIS charter change? Wouldn't they permit the support of cases such as Tatweel, Tamil figures, and French majuscules?
- 3.4. "The DNS resolver will append suffixes in the suffix search list". Where is the "suffix search list" documented?
- 4. "even when DNS is used, the conversion to Punycode should be done only by an entity that knows which name space will be used." Fundamental. Yet, this is not considered by IDNA rationale or protocol.
- 4.1. "Indeed the choice of conversion within the resolver libraries is consistent with the quote from [RFC3490] section 6.2 stating that Punycode conversion "might be

performed inside these new versions of the resolver libraries". - "the recommendation is that a resolver library be more liberal in what it would accept from an application would mean that such a name would be accepted and re-encoded as needed". These recommended architectures (such as ML-DNS) are not considered in the IDNA rationale. Will IDNA be interoperable with these recommended architectures?

- 4. "encoding conversion between Punycode and UTF-8 is unambiguous". (?). This could lead to stabilization through punycode and A-labels, in turn making A-labels the DNS referent entry for UTF support?
- 5. Security considerations. This kind of consideration is already posing a problem for TM protection. This is the "Babel Names" case. This occurs when someone trademarks the U-label corresponding to a protected Roman script TM. When that U-label displays under its ASCII label form, it infringes on the Roman script TM rights. Ex.: "xn--coca-cola" or "xn--vint-cerf".
- (Martin Dürst): The solution that the document seems to be pushing most is heuristic detection, i.e. an API where strings in different encodings are fed in and the API sorts things out heuristically, converting if necessary. To some extent, this may be an unavoidable evil, but it would be good if the document were pushing more for clear encoding identification (for which I think GetAddrInfoW() (UTF-16) would be an example).
- (Martin Dürst): It may be a good idea to also look into the issue of escaped forms of domain names being fed into resolver APIs. One form of escaping is (UTF-8-based) %-encoding in URIs (and IRIs), which is allowed in URIs according to [RFC 3986](https://www.rfc-editor.org/rfc/3986), is the only way to encode non-ASCII in the host part of an URI where punycode isn't appropriate, and may be the result of a conversion from an IRI to an URI. For further background and discussion, please see <http://lists.w3.org/Archives/Public/public-iri/2009Aug/0012.html> and <http://lists.w3.org/Archives/Public/public-iri/2009Aug/0024.html> and the followup discussion.  
Another potential kind of escaping are HTML/XML numeric character references (of the form "& #xABCD ;"), although I expect them to be less of a problem because they are used higher up in the application and usually removed early on.

## IDNA Definitions

- Information on Unicode is scattered throughout the document. Wouldn't it be much better to describe a clear sequence?
  - what an IDNA is,
  - what IDNs are,
  - what IDNA labels are,
  - what they are made of,
  - how Unicode supports them, including NFC in the same 2.1. section,



- how a zone manager may impose profiling rules (description, enforcement).
- Most of the new terms are discussed before being defined. This starts with the confusing "looking them up" in part 1.1.1. (which means resolving, and not just asking about, validity or existence) as opposed to "registering"). IDNs are introduced 2.3.2.3. etc. This certainly reflects how difficult the work is in defining all these terms, but it is still quite confusing. For example, it is advisable to begin with part 4.4.
- The different classes of domain names that are discussed only seem to be related to IDNA without an exhaustive presentation of the DNS domain name context. The names are somewhat confusing. The drafts are certainly clear, but they do not reflect a progressive logic of discovery of the nature of a name/label that could be ported to programming functions.
- References to the lower/uppercase image can be understood by DNS old-timers, but is confusing to newcomers, as it does not reflect the same functionality and because U-label/A-label lower/uppercase treatment is not the same.
- Different keyboards and encoding are discussed, stressing that a DNS resolution calls for a U-label conversion, but nothing obliges local applications to transcode user entries to Unicode when they interoperate at a layer other than DNS. However, these applications may want to canonize these entries in their proper way. Interplus supports the idea that an application layer may use middle non-Unicode and non-ASCII coding. Among others, this facilitates interoperability with UTF-8 that Microsoft supports within private nets: the user interface may be common and the underlying machinery either IDNA or UTF-8.
- 1.1.1 (Martin Dürst): Audiences - "what names are permitted in DNS zone files," -> "what names are permitted in DNS zones," (whether these are files or whatever is implementation-dependent).  
This section is very important, and would be much more effective with less circumscription. Just use straightforward terms people/functions that everybody else names directly, such as 'registries', 'registrars', 'administrators creating subdomains', and so on, and then say that this list isn't exclusive. That will have the additional benefit of bringing the document up in more of the relevant searches.  
The second paragraph is also overly circumscriptive. Using "the one containing explanatory material" to refer to Rationale is a strong disservice to every reader, even if strictly speaking may be preferable to a forward reference. Please use [] style references, or labels such as "Rationale" with a short sentence pointing to 1.3, or move the "who should read what" info to 1.3 with a general pointer from 1.1.1. But please stop talking around stuff that can be easily expressed more directly (this general comment applies in many other places, too).
- 1.1.1 (Martin Dürst): should be 1.2, and 1.1.2 should be 1.3, and 1.3 should be 1.4, to simplify structure.
- 2.1 (Martin Dürst): Say that 0x means hexadecimal (first para)

- 2.3.1 (Martin Dürst): title: This looks as if this section defines one term, "LDH-Label". Change the title to something more general, such as "Definitions for ASCII-only Labels".
- 2.3.1. (Paul Hoffman): Anchor 10 above Figure 1 indicates that the figure might be shrunk. I propose instead that the four footnotes simply be moved to immediately after the figure, which makes the figure itself fit on one page.
- 2.3.1. (Paul Hoffman) In Figure 2, the terms "Binary Label (including high bit on)" and "Bit String Label" are not defined and are confusing without definition. Do we need this figure at all any more?
- 2.3.1. (Martin Dürst): general (but most urgently 3rd para): Make sure that the terms defined stick out, at least the same way as in 2.1 (one para per def, defined word is first word of para). Move clear and simple definition to front, and rationale, relationships,... to the end of the paragraph.
- 2.3.1. (Martin Dürst): Move normative text to Protocol ("those labels MUST NOT be processed as ordinary LDH-labels by IDNA-conforming programs and SHOULD NOT be mixed with IDNA-labels in the same zone")
- 2.3.1. (Martin Dürst): 3rd para: "but which otherwise conform to LDH-label rules" -> "but otherwise conform to LDH-label rules"
- 2.3.1. (Martin Dürst): 3rd para: "case-independent" -> "case-insensitive"
- 2.3.1. (Martin Dürst): 3rd para: "divided in" -> "divided into"
- 2.3.1. (Martin Dürst): "for future extensions that use extensions based on the same "prefix and encoding" model": a) 'extensions' is repeated; b) the IETF is great at not talking about future eventualities and describing general models that never may be used. In this and other sections, such stuff should also be cut out.
- 2.3.1. (Martin Dürst): anchor10: I do not understand why we need the (1)..(4) notes. Either the definitions are clear enough, or they should be fixed. Something like "NON-RESERVED LDH LABELS (NR-LDH-labels) NR-LDH LABELS" is also total overkill. The only thing that's necessary is "NR-LDH labels", with exactly the same capitalization and hyphenation as in the definition.
- 2.3.1. (Martin Dürst): Fig. 2: I'm somewhat confused here. Note (5) seems to suggest that U-labels have a fixed binary encoding (e.g. UTF-8) and are used directly in the DNS. Otherwise, the note doesn't make sense.
- 2.3.2.1. (Martin Dürst): "While that constraint may be tested in any of several ways, an A-label must be capable of being produced by conversion from a U-label and a U-label must be capable of being produced by conversion from an A-label.": This puts the cart before the horse. Change to "An A-label must be capable of being produced by conversion from a U-label and a U-label must be capable of being produced by

conversion from an A-label. There are several ways in which this constraint may be tested."

- 2.3.2.1. (Andrew Sullivan): Para. 2.3.2.1: An "A-label" is the ASCII-Compatible Encoding (ACE, see Section 2.3.2.5) form of an IDNA-valid string. It must be a complete label: IDNA is defined for labels, not for parts of them and not for complete domain names. This means, by definition, that every A-label will begin with the IDNA ACE prefix, "xn--" (see Section 2.3.2.5), followed by a string that is a valid output of the Punycode algorithm [RFC3492] and hence a maximum of 59 ASCII characters in length. The prefix and string together must conform to all requirements for a label that can be stored in the DNS including conformance to the rules for the preferred form described in [RFC 1034](#), [RFC 1035](#), and [RFC 1123](#). A string meeting the above requirements is still not an A-label unless it can be decoded into a U-label. So, to be less vague: the section is supposed to define certain terms, and that bullet ought to define "A-label". It does not. It tells us the necessary conditions for being an A-label, but not the sufficient. This could be remedied if the last sentence said instead, "If and only if a string meeting the above requirements can be decoded into a U-label, then it is an A-label." But I'm no longer sure that's true, given that we've lived with the I-D definition so long and yet not had it fully operationalized. Is there anything else? If there is, it needs to be added. These definitions, I say, must be completely operationalized (or else we have no excuse to call this document the definitions document). Since people have to write code on the basis of these definitions, they must be completely unambiguous.

**Author's current answer:** These changes, with Paul's suggested modifications, have been tentatively accepted and incorporated in the document. Anyone who objects should say so quickly.

- 2.3.2.1. (Martin Dürst): "Among other things, this implies that both U-labels and A-labels must be strings in Unicode NFC [Unicode-UAX15] normalized form.": A-labels are by definition in NFC, because they are ASCII-only. If you want to say that they must \*represent\* labels that are in NFC, that would be fine, but I think mentioning NFC here isn't really necessary.  
MAJOR!!!!
- 2.3.2.1. (Martin Dürst): says: "Any rules or conventions that apply to DNS labels in general, such as rules about lengths of strings, apply to whichever of the U-label or A-label would be more restrictive. For the U-label, constraints imposed by existing protocols and their presentation forms make the length restriction apply to the length in octets of the UTF-8 form of those labels (which will always be greater than or equal to the length in code points)."  
Now this is TOTALLY NEW to me. There sure is a restriction to 63 octets in the DNS itself, but because U-labels don't enter the DNS as such (neither as UTF-8 nor as UTF-16 or whatever), an arbitrary UTF-8-based length restriction seems totally unjustified. I'm not at all aware of such a restriction in IDNA2003.  
Indeed, punycode was explicitly designed, among else, to perform well for scripts with few characters. For small scripts that need 3 bytes per character in UTF-8 (all

Indic scripts, Georgian, Sinhala, Thai, Lao, Tibetan, Myanmar, Ethiopic, Cherokee, Unified Canadian Aboriginal Syllabics, Khmer,..., this restriction would mean a drastic reduction of the number of characters usable in a label. To give an example, when at W3C, I created some IRI tests (<http://www.w3.org/2001/08/iri-test/>).

The tests use Hiragana

<http://www.hontou.nagaiwake.wakara.naidomeinmei.oraberumadana.gakushinaitarinai.w3.mag.keio.ac.jp> and

[http://hontou.nagaiwake.wakara.naidomeinmei.oraberumadana.gakushinaitarinai.hontou.nagaiwake.wakara.naidomeinmei.oraberumadana.gakushinaitarinai.hontou.nagaiwake.wakara.naidomeinmei.oraberumadana.gakushinaitarinai.hontou.nagaiwake.wakara.naidomeinmei.oraberumadana.gakushinaitarinai.w3.mag.keio.ac.jp](http://hontou.nagaiwake.wakara.naidomeinmei.oraberumadana.gakushinaitarinai.hontou.nagaiwake.wakara.naidomeinmei.oraberumadana.gakushinaitarinai.hontou.nagaiwake.wakara.naidomeinmei.oraberumadana.gakushinaitarinai.w3.mag.keio.ac.jp)), which is

atypical in that Hiragana-only Japanese is rarely used except in children's books, but it is typical in that punycode is able to represent 41 Hiragana (123 octets in UTF-8) in 58 octets. Hiragana overall contains about 80 letters in a single block; punycode efficiency will vary with the size of the script (more efficient for smaller scripts, less efficient for larger scripts) as well as of course with every individual label.

Currently, all (on Windows) of IE7, Mozilla Firefox, Safari, and Opera pass both length tests (single label and multiple labels). It would be very counterproductive if IDNA2008 required further artificial restrictions which essentially disfavor languages and cultures that haven't been lucky to get short encodings for their scripts in UTF-8. (I'd be fine if the Security section warns about the potential of some protocols or implementations not having appropriate space, but that's on a completely different level.)

- 2.3.2.1. (Andrew Sullivan): "A "U-label" is an IDNA-valid string of Unicode characters, in normalization form NFC and including at least one non-ASCII character, expressed in a standard Unicode Encoding Form (in an Internet transmission context this will normally be UTF-8)." The parenthetical remark, I think, encourages implementers not to recognise as U-labels strings that come in as (say) UTF-32, but that are otherwise perfectly valid. Who cares what is normal in an Internet transmission context, when we're defining terms? Why does that matter ?

**Author's current answer:** The comment was made because there is no requirement at all in IDNA (either 2003 or 2008) that UTF-8 be used; many applications on particular operating systems actually use something else (UTF-16 is most common). But I dropped the additional text. It now just says "(such as UTF-8)" as Paul suggested. Again, anyone who doesn't like this should speak up

- 2.3.2.1. (Andrew Sullivan): "To be valid, U-labels and A-labels must obey an important symmetry constraint. While that constraint may be tested in any of several ways, an A-label must be capable of being produced by conversion from a U-label and a U-label must be capable of being produced by conversion from an A-label. Among other things, this implies that both U-labels and A-labels must be strings in Unicode NFC [Unicode-UAX15] normalized form. These strings MUST contain only characters specified elsewhere in this document series, and only in the contexts indicated as appropriate."

This passage nowhere actually says that \_the\_ A-label produced by conversion from a

particular U-label must in turn produce, by the application of the algorithm, the same U-label. There is a symmetry (though not an obvious one) in U[1] being convertible to A[2] which is convertible to U[2] which is convertible to A[1], for instance. I have no idea whether such is possible, but there's no reason our formal definitions need to allow for it. This can be fixed so:

To be valid, U-labels and A-labels must obey an important symmetry constraint. While that constraint may be tested in any of several ways, an A-label A' must be capable of being produced by conversion from a U-label U', and that U-label U' must be capable of being produced by conversion from A-label A'. Among other things, this implies that both U-labels and A-labels must be strings in Unicode NFC [Unicode-UAX15] normalized form. These strings **MUST** contain only characters specified elsewhere in this document series, and only in the contexts indicated as appropriate. I don't care about the notation, as long as it is unambiguously clear that we're always talking about the "very same" label on both sides of the transformation. We could go so far as to define IDNA-equivalent A-labels and U-labels formally. I think this would do it:

A-label1 and U-label1 are equivalent if and only if all the following four conditions are true:

1. The encoding of A-label1 according to [RFC3492] results in U-label1.
2. The decoding of U-label2 according to [RFC3492] results in A-label2.
3. A-label1 is equivalent to A-label2 according to DNS matching rules for labels.
4. U-label1 is bistring equivalent to U-label2.

Some may reject the above as a bit of needless formalism, or want to reduce some steps. I argue that this is the most basic and therefore most clear (but admittedly inelegant) formulation. As usual, however, I'm utterly prepared to admit that I've actually got the rule incorrect. But if I have, that amounts to a hint of trouble with the document, because I've managed to misunderstand it (and though I be dim, I have been following this effort).

- 2.3.2.1 (Wil Tan): In protocol section 5.3, A-label Input section to add the lowercasing step prior to using the Punycode decoding algorithm. The section on symmetry constraint (-defs-10, section 2.3.2.1) should also have similar wordings.
- 2.3.1. Para5 (Wil Tan): Labels within the class of R-LDH labels that are not prefixed with "xn--" are also not valid IDNA-labels. To allow for future use of mechanisms similar to IDNA, those labels **MUST NOT** be processed as ordinary LDH-labels by IDNA-conforming programs and **SHOULD NOT** be mixed with IDNA-labels in the same zone.

**Author's current response:** Unless, in the moving around of text, we have slipped up, it is important to note that the restriction here applies only to IDNA-aware applications. That prevents it from being a restriction on the DNS generally. However, for IDNA-aware applications, it is a precaution against possible future prefix-altering changes as well as something of a mechanism for making it harder for bad guys to game future changes. If any non-IDNA arrangements come along that use "??--" label

encodings, they will of course have to be coordinated with each other and with IDNA; in the interim, this provision keeps IDNA out of their way (i.e., avoids preempting such approaches).

And, yes, the WG did discuss this at great length.

"I may have missed it, but don't recall any discussions about restricting the processing of other tagged domains. Is this the right draft to prescribe restrictions on how non-XN-Labels are processed?"

**Author's current response:** IMO, we are much too tied up in special definitions and confusing terminology already. Please let's not make it worse by introducing more unnecessary terminology in the form of "tagged domains". And it is the right place for defining how IDNA-aware applications handle R-LDH labels that are not valid A-labels, at least IMO and in the opinion of the mailing list the last two or three times we went through that topic.

- 2.3.2.2. (Martin Dürst): NR-LDH-label and Internationalized Label: The section doesn't say anything about "Internationalized Label"s, although this term appears in the title. (the definition is in 2.3.2.3)
- 2.3.2.3. (Martin Dürst): SVR record labels are not Internationalized labels, and therefore domain names used for SVR aren't IDNs. That's fine by me, but it should nevertheless be made clear (here or elsewhere) that IDNs can be used with SVR,... (this seems to be done at the end of 2.3.2.6, so this should be okay)
- 2.3.2.4. (Martin Dürst): This seems to say that there is no equivalence between an all-lowercase A-label and an otherwise equal label where some letters (maybe accidentally) have been upper-cased. I think the cause of the problem is (as often in this document) the lack of consistent language. Instead of "and then testing for an exact match between the A-labels", say "and then testing for equivalence between the A-labels [using normal DNS matching rules]". If that's not what's intended, then some more background may be appropriate.
- 2.3.2.5. (Martin Dürst): "a string of ASCII characters" -> "the string of ASCII characters"
- 2.3.3. (Martin Dürst): "Because IDN labels may contain characters that are read, and preferentially displayed, from right to left,": Remove 'preferentially'. This maybe refers to some hopelessly broken systems, or to the fact that Arabic Braille is LTR, or something else, but is totally irrelevant and potentially misleading in this context.
- 2.3.3. (Martin Dürst): Why doesn't this paragraph just refer to 'logical' representation, a term that people who know bidi are familiar with and that's widely used in Unicode.
- 2.3.3. (Andrew Sullivan): Also, as an aside, it would be helpful if defs pointed out more explicitly in its Para. 2.3.3 that there are BIDI-only terms defined in the BIDI document and not in defs.

**Author's current answer:** I moved it to the beginning. I don't think there's anything bidi-specific in -defs.

- 2.3.4. (Martin Dürst): "There has been some confusion about whether a "Punycode string" does or does not include the ACE prefix and about whether it is required that such strings could have been the output of the ToASCII operation":
  - a) The combination of 'required' and 'could' doesn't make ANY sense for me.
  - b) Is it unclear what "such strings" refers to (with ACE prefix? without ACE prefix?)
- 2.3.4. (Martin Dürst): "much more clear" -> "much clearer"
- 4. (Martin Dürst): There should be a very short paragraph saying that this section provides an overview and pointers into the security sections of the other documents. (or whatever else exactly the relationships are)
- 4.1. (Martin Dürst): "In addition to characters that are permitted by IDNA2003 and its mapping conventions": Does this mean "In addition to characters that are permitted by (IDNA2003 and its mapping conventions)" or "In addition to characters that are permitted by IDNA2003 and [in addition to] its mapping conventions"? Please clarify.
- 4.1. (Martin Dürst): "problems that might raise" -> "problems that might arise"
- 4.1. "Security on the Internet partly relies on the DNS. Thus, any change to the characteristics of the DNS can change the security of much of the Internet." This sentence seems extremely confusing, as IDNA does not affect (change characteristics) the DNS but is rather built on the fact that they will not be changed.
- 4.1. The same : "The security of the Internet is compromised if a user entering a single internationalized name is connected to different servers based on different interpretations of the internationalized domain name." The security of the Internet is not compromised, however, trust in the IDNA proposition might be.
- 4.2. (Martin Dürst): "these specifications"? The IDNA2008 collection of specifications? Or the specifications for the local character sets?
- 4.2. (Martin Dürst): "(or different versions of one application)" -> "(or different versions or parts of one application)" (yes, this can and does happen)
- 4.4. (Martin Dürst): "comparisons be done properly, as specified in the Requirements section of [IDNA2008-Protocol]": If comparisons are dealt with in Protocol, what's the purpose of 2.3.2.4? And what's the purpose of trying to explain it all again just after the quoted sentence?
- 4.5. (Martin Dürst): "Despite that prohibition, there are a significant number of files and databases on the Internet in which domain name strings appear in native-character form;": This makes it appear as if such files and databases are in violation of some spec. But they may simply contain IRIs instead of URIs. I would simply start the subsection with something like "As long as IDNA2003 labels have been kept in A-

label form, the only differences in interpretation arise for characters whose ..." and then, in a new paragraph, continue "For IDNA2003 labels that have been kept in native encoding,..."

- 4.7. The Summary might be considered adventurous? Corporations such as Nominum propose services that are supposed to protect the DNS. One of the purposes of ML-DNS is precisely to permit an architectural protection.
- Acknowledgments (Andrew Sullivan): "As is usual with IETF specifications, while the document represents rough consensus, it should not be assumed that all participants and contributors agree with all provisions," I don't feel comfortable with starting to make the Acknowledgements section a platform for disclaimers about WG consensus. I object pretty strongly to this addition. I don't think we're served well by trying to state in any document how rough the rough consensus is: the document either has to stand through the IETF process, or not. Besides, this evaluation is a prerogative of the Chair, the ADs, and the IESG. If this sort of disclaimer is needed, it ought to be added by the IESG (and even then I would object). I would like the sentence to be removed.

## IDNA Rationale

- 1.1 (Andrew Sullivan): bugs me: "Traditionally, DNS labels are matched case-insensitively [RFC1034][RFC1035]. That convention was preserved in IDNA2003 by a case-folding operation that generally maps capital letters into lower-case ones. However, if case rules are enforced from one language, another language sometimes loses the ability to treat two characters separately. Case-sensitivity is treated slightly differently in IDNA2008." It makes it sound as though there's just a kooky tradition in the DNS, and we could fix that up. But that's not true: the matching rules are `_defined_` to be case insensitive, so changing that would be a protocol change to DNS. Also, the text slides a little to easily between what are different contexts of "label" here, and I think it could be a source of confusion. I suggest this instead: The DNS matching rules for DNS labels are case-insensitive [RFC1034][RFC1035]. That convention was preserved for internationalized labels in IDNA2003 by a case-folding operation that generally maps capital letters into lower-case ones. However, if case rules are enforced from one language, another language sometimes loses the ability to treat two characters separately. Case-sensitivity is treated slightly differently in IDNA2008.
- 1.3.1. DNS "Name" Terminology. "" would be better read as "orthotypographic" as an orthographic error that can be a way to lose some special semantics differences due to orthotypographic conventions.
- 1.3.2. "IDNA-landr" typo?
- 1.4. "Reduce the dependency on mapping, in order that the pre-mapped forms (which are not valid IDNA labels) tend to appear less often in various contexts, in



favor of valid A-labels." calls for the Charter to be revised. Alternatively, it could say , remove dependence on mapping as per a mapping document, in which this document would include a section on the various ways to ensure DNS security and the barring of some U+codes in some presentations.

- 1.5. "This model has served the existing applications well, but it requires, with or without internationalized domain names, that users know the exact spelling of the domain names that are to be typed into applications such as web browsers and mail user agents. The introduction of the larger repertoire of characters potentially makes the set of misspellings larger, especially given that in some cases the same appearance, for example on a business card, might visually match several Unicode code points or several sequences of code points." may be read as if the users of these languages were more prone to errors than ASCII language.
- "If an application wants to use non-ASCII characters in public DNS domain names, IDNA is the only currently-defined option." IDNA is not a DNS option. It is an application way to transcode Unicode domain names in LDH domain names for the convenience of ASCII oriented international managers. The idea is to attain the adherence of local users and managers to IDNA and not to impose ASCII on them. DNS is UTF-8 compatible.
- "IDNA2008 divides all possible Unicode code-points into four categories: PROTOCOL-VALID, CONTEXTUAL RULE REQUIRED, DISALLOWED and UNASSIGNED.  
3.1.1. PROTOCOL-VALID  
Characters identified as "PROTOCOL-VALID" (often abbreviated "PVALID") are permitted in IDNs." Are we talking of code-points or of characters?
- 3.1.2.1 Not in the TOC
- 3.1.3 Disallowed "various HEART symbols" - is U+38FA also disallowed? or U+3966?
- 3.1.3. This is the first time anyone has spoken of NFKC. In IDNA Defs and other cases, it is NFC. Shouldn't both of them be documented? Shouldn't someone explain in which specific case one is used?
- "The character is an upper-case form or some other form that is mapped to another character by Unicode casefolding." this seems to create a very large mapping scheme that depends on a non-documented Unicode system needing correction (at least when it does not specifically support majuscules). Moreover, are we dealing with characters (that are orthogonal to Unicode) or with codepoints that represent characters and that are subject to Unicode casefolding. The proposition is to: (1) clarify the character/codepoint issue, (2) explain what Unicode case folding is and its limitations, (3) move them to CONTEXTO when these codepoints are both used as upper-cases and as majuscules, (4) explain that majuscules that are supported by upper-cases will be transcoded by punycode.

- 4.4. Case mapping. One may regret that the French majuscules current support of Unicode, which is perfectly adequate in other circumstances yet inadequate in this case, is not discussed. This would explain the upgrade above.

**Author's current answer:** First, you want certain characters treated differently for some languages that use a given script than for others that use the same script. That is nearly impossible to think about, just because there is, in general, no way to know what language a particular label is supposed to be associated with, nor is there a way to know what top-level domain has the label in one of its subtrees (even if one could reliably associate top-level domains with languages).

-- Second, if I understand your latest note correctly, you would like to have those characters treated via some contextual rule ("CONTEXT0"). But the contextual rules yield either "valid" or "invalid" based on adjacent or nearby characters -- they do not provide different mappings, nor different rules for different languages (the latter at least partially for the reasons above).

-- And, finally, your suggestion requires treating capital letters (or at least some capital letters) as distinct from their lower-case forms, which would create massive inconsistencies with IDNA2003 (not just the two characters of inconsistency with which we have had such extensive debates) as well as inconsistencies with DNS and host table practices that go back to the 1970s. No matter how strong your justification, and even if it were not also tied to differential treatment for a particular language, I cannot imagine the WG (or the IETF more broadly) agreeing to that change.

- 4.5. "Examples of this are Yiddish, written with an extended Hebrew script, and Dhivehi (the official language of Maldives) which is written in the Thaana script (which is, in turn, derived from the Arabic script)" It seems that some explanation about Yiddish would be welcome so that the language will obtain the same support as Dhivehi and Thaana.
- 5. "Conversely, lookup applications are expected to reject labels that clearly violate global (protocol) rules (no one has ever seriously claimed that being liberal in what is accepted requires being stupid)." The remark between the parentheses is confusing: it possibly qualifies as "stupid" a behavior that is not recommended, but that is acceptable by the document set.
- "Application implementors should be aware that where DNS wildcards are used, the ability to successfully resolve a name does not guarantee that it was actually registered." In which terms is this specific to IDNA?
- 7.6. The Symbol Question. That part actually discusses the Unicode originated difficulties. Yet, the choice of Unicode has not yet been discussed.
- 8.2. (Andrew Sullivan) the reference for DNSSEC should probably refer to DNSSECbis, which is usually [RFC4033], [RFC4034], [RFC4035], since RFC2535 is obsolete. One can make a strong argument for also including NSEC3 [RFC5155], but I don't feel too strongly about that.

- 9. "Adding languages (or similar context) to IDNs generally, or to DNS matching in particular, would imply context dependent matching in DNS, which would be a very significant change to the DNS protocol itself". This sentence seems confusing. Natural languages are quoted throughout this IDNs document.
- 12. (Andrew Sullivan): Here is where I make my now-familiar plea for the removal of the following sentence "As is usual with IETF specifications, while the document represents rough consensus, it should not be assumed that all participants and contributors agree with all provisions."

## IDNA Mapping

- (Andrew Sullivan): In general, I think the document is in reasonably good shape. I was hoping, however, for it to have some advice to registries on what sort of considerations are worth taking into account when formulating policies around what will and will not be registered.  
In particular, I think it would be very helpful to outline what it would mean for characters to be possibly mappable, and also to outline the different strategies that are available for preventing different alternative mappings from ending up with a different resolution.  
The basic advice is that registries should try to detect whether a candidate label may be expected to be the result of some (plausible) application of a mapping appropriate in the probable user community, and similarly that when faced with a candidate IDN, registries should consider the probable user community and consider the plausible applications of mappings appropriate to that community. If the registry does not have the expertise to evaluate the probable user community for a given code point, then it should simply reject the code point outright. (This is, in effect, the advice that you should have a policy, it should be a policy based on knowledge of the use cases, and that it should default closed.)  
After the registry has detected whether the candidate label, there are two basic strategies it might follow:
  1. Detect and reserve. In this case, the registry detects potential mappings, and reserves other candidate labels that might be the result of such mappings. This reservation takes the form of preventing registration of that label.
  2. Detect and bundle. In this case, the registry detects the potential mappings, and creates identical entries in the registry conforming to those "alternative forms" of the candidate label. There is the potential for a very large number of these bundled labels.
- (Andrew Sullivan): There's also a great deal of material at the end of rationale that would be more appropriate in this document (or else it should just go away, I think).
- (Andrew Sullivan): I think the document is ready to go, assuming this is what we want to say. Yet its content is a little flabby as recommendations go. I can imagine a reader being a little surprised at this advice, for example: "These are a minimal set of mappings that an application should strongly consider doing. Of course, there are

many others that might be done." That boils down to, "You might want to do this. Or not. Or something else. Up to you." I know why we're saying this, but I would not be surprised if people object to such thin advice. If that's all the advice we want to offer, however, this is the right document and it should go ahead.

- Not sure that the terminology of "make sense" is adequate or clear.
- 1. Introduction - This document is supposed to be separated from the IDNA document set. It should then document what the IDNA protocol is. It seems that the IDNA2008 protocols boil down to "DNS domain names are to be expressed in LDH form. IDNA is a commonly agreed upon convention wherein if they are entered by the user in another form, applications are advised to convert them to UTF in order to filter and map them, as is discussed in the present document, as well as to transcode them in by using the punycode algorithm. Depending on the Registry policy, their registration can be carried out in the ITF and/or the transcoded ASCII form."
- 2.3. NFC is confirmed, NFKC is not discussed.

## IDNA Protocol

- As a general comment:
  - The SHOULD/MUST chains may be somewhat awkward. MUSTs are used in a protocol procedure and then an alternative to that procedure is pragmatically considered. It could be of interest to draft a MUST tree to consider which cases are, or are not, covered.
  - there is some confusion as to what the "string" is compared to the label and domain name, in which "Label" may be used instead of "U-Label" or sometimes "A-Label". Wouldn't it be better to review the text in qualifying the "labels" in order to be certain that all the cases are clearly covered?
- 1.para 1: (Martin Dürst): I missed the term IDNA2003 in Defs, it would have been useful. I didn't complain because I thought it had been 'deprecated'. Seeing it here, I think it should go back to Defs, and be actively used in Defs and Protocol at least, to simplify and clarify prose.
- 1.para 2: (Martin Dürst): "does not changes" -> "does not change"
- 1.para 2: (Martin Dürst): "IDNA does not depend on any changes to DNS servers, resolvers, or protocol elements" -> "IDNA does not depend on any changes to DNS servers, resolvers, or DNS protocol elements" or "IDNA does not depend on any changes to DNS (servers, resolvers, or protocol elements)" (Otherwise, it's possible to understand 'protocol elements' as being not limited to DNS.)
- 1.para 4: (Martin Dürst): ", that share some terminology, reference data and operations." -> "These two protocols share terminology, reference data and operations."

- 2.para 1: (Martin Dürst): "Terminology used in IDNA, but also in Unicode or other character set standards and the DNS, appears in [IDNA2008-Defs]." -> "Terminology used in IDNA appears in [IDNA2008-Defs]." (where else these terms are used, or where they are from, can be explained in Defs where necessary, but is absolutely irrelevant here)
- 2.para 1: (Martin Dürst): "Terminology that is required as part of the IDNA definition, including the definitions of "ACE", appears in that document as well." -> remove (first, the word 'definition' is used with two slightly different meanings, and second, I don't see the point of singling out "ACE".)
- 3.1. Requirement 2: (Martin Dürst): Equivalence is already defined in Defs. Please make sure there is only a single definition.
- 3.1. Requirement 2: (Martin Dürst): Why is it a MUST that U-Labels are compared without case-folding (even for ASCII chars?) or other steps?
- 3.1. Requirement 2: (Martin Dürst): "In many cases, validation may be important for other reasons and SHOULD be performed.": Is this restricted to when trying to compare? Or in general?
- 3.1. Requirement 3: (Martin Dürst): This does double duty, and should be removed. The alternative is to covert 3.1 into a conformance section as usual e.g. for ISO standards, but then a lot more rewriting will be necessary in all of 3.1.
- 3.2. "It does not apply to domain name slots which do not use the Letter/Digit/Hyphen (LDH) syntax rules." Confusing. Would some of the DN slots not accept both?
- 3.2. para 1: (Martin Dürst): "IDNA applies to": What does "applies to" mean?
- 3.2. para 2: (Martin Dürst): "Because it uses the DNS, IDNA applies" -> "Because IDNA uses the DNS, it applies", or even better "Because IDNA uses the DNS, IDNA applies" (repetitions don't hurt in standards, reference before referent does)
- 3.2. para 2: "unless those protocols and implementations of them" -> "unless those protocols and their implementations"
- 3.2. para 2:(Martin Dürst): "be aware of IDNs in Unicode" -> "be aware of IDNs" (whether they are in Unicode or not is irrelevant here)
- 3.2.1. The word CLASS only appears in the whole document set in two sentences: "DNA applies only to domain names in the NAME and RDATA fields of DNS resource records whose CLASS is IN. See [RFC 1034](#) [RFC1034] for precise definitions of these terms. The application of IDNA to DNS resource records depends entirely on the CLASS of the record, and not on the TYPE except as noted below." What about internationalized domain name in a non IN CLASS?

**Author's answer:** I have received no further input on this and will assume that the current text is ok unless I do.

**Additional Comment** My reading of that text was that it was a `_restriction_`, not a claim of fact. In other words, I interpreted that text as saying that, if a new class is invented, IDNA (if it were specified) would need to be specified separately for that class.

By definition, the only place domain name labels can appear is in the NAME and RDATA fields of resource records. It's important to remember that the RDATA field can have subfields (as it does in the SOA record). I think that's clear enough from the rest of the discussion, and because the documents already say, "You need to understand DNS too."

- 3.2.1. (Paul Hoffman): Section 3.2.1 says: "IDNA applies only to domain names in the NAME and RDATA fields of DNS resource records whose CLASS is IN." It would be good for the DNS-centric folks in the WG to verify that they think that this restriction is correct. Are there really no other fields where domain labels would appear?
- 3.2.1. para 1: (Martin Dürst): The first paragraph reads as if IDNA applied to domain names in e.g. TXT records in CLASS IN. I think it would help here to say exactly what is meant by "IDNA applies". In some sense, IDNA applies nowhere in DNS records, they are all just ASCII. In some sense (labels starting with xn-- are presumed to be IDNA labels; you can add an IDN (or a label thereof) to a DNS record by using A-labels), IDNA applies.
- 3.2.1. para 2: (Martin Dürst): The SVR discussion has significant overlap with Defs, please reduce.
- 4. "This section defines the procedure for registering an IDN. The procedure is implementation independent; any sequence of steps that produces exactly the same result for all labels is considered a valid implementation." A procedure does provide but does not define a result?
- 4. para 1:(Martin Dürst): "defines *the* procedure" ... : This would work better if there were really only one procedure, and it were written as a procedure. However, there are often variations, and different, often non-procedural ways in which things are expressed (e.g. 'labels must ...' instead of 'if a label doesn't satisfy x, abort')
- 4. para 2: (Martin Dürst): "the registration and lookup protocols (Section 5)" -> "the registration protocol (this section) and the lookup protocol (Section 5)" (shortcuts are the enemies of specifications)
- 4. para 2: (Martin Dürst): "while ... are very similar in most respects, they are different" -> "while ... are very similar in most respects, they are not identical"
- 4. para 2: (Martin Dürst): "follow the appropriate steps": appropriate appeals to value judgement, which isn't adequate here.

- 4.1. The obligation chain reads: "By the time a string enters the IDNA registration process [], it is expected to be in Unicode [], yet "registries [] SHOULD avoid any possible ambiguity by accepting registrations only for A-labels []."
- 4.1. (Andrew Sullivan): "By the time a string enters the IDNA registration process as described in this specification, it is expected to be in Unicode and MUST be in Unicode Normalization Form C (NFC [Unicode-UAX15])." The "expected to be" part is redundant, since we have a subsequent MUST.
- 4.1. (Andrew Sullivan): I found this slightly confusing: "The registry SHOULD permit submission of labels in A-label form and is encouraged to accept both the A-label form and the U-label one. If it does so,". The "does so" reference there is ambiguous: is it the submission of A-labels or the A-label+U-label case. The subsequent text suggests it's the former.
- 4.1. title: (Martin Dürst): Why suddenly "Process" instead of "Procedure"? Why not just "Input"? And why singular in the title, and then plural in the first line of the text?
- 4.1.(Martin Dürst): "are outside the scope of these protocols": How many protocols are there? Only one that's relevant here.
- 4.1. (Martin Dürst): Why is NFC a condition on the input? Please make it a validation step afterwards, to streamline things.
- 4.1. (Martin Dürst): "Entities responsible for zone files ("registries") are expected to accept only the exact string for which registration is requested, free of any mappings or local adjustments.": It's clear to me what we want here, but it's much better to write this as a condition on the later processing, rather than on input, something like: "Entities responsible for zone files ("registries") MUST NOT apply any mappings or local adjustments of any kind to the exact string for which registration is requested."
- 4.1. (Martin Dürst): "They SHOULD avoid any possible ambiguity by accepting registrations only for A-labels, possibly paired with the relevant U-labels so that they can verify the correspondence." This has to be improved. First, the SHOULD doesn't belong on the reason, and the reason, if anywhere, belongs at the end. Second, there are three possible ways input can come in, so let's list things up: "Entities responsible for zone files ("registries") MAY accept input in any of three forms:
  - 1) As a pair of A-label and U-label
  - 2) As an A-label only
  - 3) As an U-label only.
 1) and 2) are RECOMMENDED because the use of A-labels avoids any possibility for ambiguity. (the first sentence in 4.2.1 can then be removed)
- 4.2.1. (Martin Dürst): This is a complex jungle of conditions on input, conversions,... What should be done is:
  - a) extract the 'raw' (without any preconditions) U-label->A-label and A-label->U-label 'functions' into subsections e.g. in Section 3; these will serve as building blocks both in Section 4 and Section 5.

b) As the first step of the registration procedure, make sure we have both an A-label and an U-label. One way to write this is:

"4.2.1: Preprocessing

1) If the input contained an A-label and a U-label, check that they are equivalent (or whatever that was called; the conditions are somewhere in Defs). If the check fails, abort registration.

2) If the input contained an A-label, but no U-label, calculate the U-label according to @@@.

3) If the input contained an U-label, but no A-label, calculate the A-label according to @@@."

The above makes sure we have both an A-label and an U-label from here on.

Checking on these can be performed independently (e.g. length check on A-label, NFC check on U-label). Conversion to punycode is no longer needed in 4.4, because we simply put the A-label we have now into the zone (assumed we have passed all the checks up to here, of course).

- 4.2.1. (Martin Dürst): (probably not needed anyway) "both the A-label form and the U-label one" -> "both the A-label form and the U-label form"
- 4.2.1. (Martin Dürst): Word the A-label checks more clearly, and create section "4.2.2 A-label Validation"
- 4.2.3.2. "a combining mark or combining character" -> "a combining character" (combining marks are a special case of combining characters, and as such irrelevant here)
- 4.2.3.3. (Martin Dürst): "To check this, each code-point marked as CONTEXTJ and CONTEXTO in [IDNA2008-Tables] MUST have a non-null rule." Is this a requirement on Tables? Are there "null rules"? What purpose do they serve, what's the difference between them and DISALLOWED?
- 4.2.3.4. (Martin Dürst): What are "characters written from right to left"? Either we define this clearly here, or we leave it (or put it) in Bidi, but then we have to rewrite the sentence here (just requiring conformance to the conditions in Bidi).
- 4.2.4. (Martin Dürst): This is totally unnecessary, please remove. If we need a summary for what's essentially just about a page of text, we better give up.
- 4.3. Registry restriction inheritance is not alluded to.
- 4.3. (Martin Dürst): "Policies are likely to be informed by the local languages" -> "Policies are likely to be informed by the local scripts and languages" (IDNs are mostly a script issue, much less a language issue. ICANN has fixed their documents to avoid only talking about languages (they still could move a bit further to scripts), so let's not commit the same mistake here again.)
- 4.3. (Martin Dürst): "or the application of special restrictions to others": like what? Like that such a label can only be resolved on Tuesdays?



- 4.4. (Martin Dürst): The generic parts of the conversion need to go somewhere else (Section 3?). The actual conversion (or checking) needs to go at the start of 4.2. Then this section is empty and can be removed.
- 4.5. (Martin Dürst): "The A-label is registered in the DNS by insertion into a zone." -> "The label is registered in the DNS by inserting the A-label into a zone." (distinguish registration of the abstract thing from insertion of the concrete thing)
- 5. Does this repetition (already in IDNA Rationale) "the presence of wild cards in the DNS might cause a string that is not actually registered in the DNS to be successfully looked up." reflect what the BIDI documents slightly differently: "Wildcard create the odd situation where a label is "valid" (can be looked up successfully) without the zone owner knowing that this label exists. So an owner of a zone whose name starts with a digit and contains a wildcard has no way of controlling whether or not names with RTL labels in them are looked up in his zone."
- 5. (Martin Dürst): para 2: " The two steps described in Section 5.2 are required.": Superfluous. Make sure there's a MUST at the right place in that section. (Looking at 5.2, I have no clue what the two steps should be. This shows that indirect requirements like the above are rather unhelpful.)
- 5.1.(Martin Dürst): first paragraph: Although IDNs will often get extracted from IRIs or URIs, there are many cases where these constructs are not involved. Examples would be telnet or ping commands, and so on. So IRIs and URIs should be deemphasized more.
- 5.1.(Martin Dürst): "Processing in this step and the next two are local matters, to be accomplished prior to actual invocation of IDNA.": Again, which steps? Before, we supposedly had two steps in 5.2, now it looks as if we are talking about 5.2 and 5.3 as two steps. -> Create a subsection such as "Input preparation" or what where all the preliminary stuff goes in. Alternatively, talk about subsections, with subsection numbers for clear identification.
- 5.2. The case of a character that is not supported by Unicode is not discussed.
- 5.2. (Martin Dürst): "is not already Unicode" -> "is not already in Unicode" (in parallel to 'into' in the line before)
- 5.2. (Martin Dürst): "A Unicode string may require normalization as discussed in Section 4.1.": There is no "discussion" in 4.1 (and no need for discussion). Express the requirements here independently of Section 4.
- 5.3. (Wil Tan): section 5.3, A-label Input section to add the lowercasing step prior to using the Punycode decoding algorithm. The section on symmetry constraint (-def-10, section 2.3.2.1) should also have similar wordings.

- 5.3. (Martin Dürst): (just checking) "See the Name Server Considerations section of [IDNA2008-Rationale] for additional discussion on this topic.": From the context, Name Server doesn't look related (we are client-side here).
- 5.3. (Martin Dürst): "That conversion and testing SHOULD": Replace 'That' with something clearer and more precise.
- 5.3. (Martin Dürst): para 2: List up the alternatives that are possible. Avoid mishmash textual paragraphs.
- 5.4. The use of "U-Labels" in this part instead of "Labels" would probably clarify it.
- 5.4. (Paul Hoffman): Section 5.4 assumes that an application knows the version of Unicode that is being used in the application. We should state that assumption in 5.4 or maybe further up near the beginning of section 5.

**Author's answer:** It assumes that either the application or the operating system or library support keeps the two consistent. That range of options is the reason why Section 5.4 is not more explicit about which particular software elements or modules know what. If this is to be changed, I need suggestions about textual fixes that do not imply that the knowledge must be in the application itself.

**Further comment:** The first bullet in 5.4 is the first time that "version of Unicode" is mentioned, so the note is probably most effective right there. I propose adding: This requirement means that the application must use a list of unassigned characters that is matched to the version of Unicode that is being used for the other requirements in this section. It is not required that the application know which version of Unicode is being used; that information might be part of the operating environment in which the application is running.

- 5.4. (Paul Hoffman) The paragraph in section 5.4 that starts "This test may..." is out of date because the rules in the Bidi document no longer do inter-label checking. The whole paragraph can be removed.
- 5.4. (Paul Hoffman) In the light of this, does the WG want to change the requirement level for checking Bidi on lookup from SHOULD to MUST? Given the above, I see no reason why not.
- 5.4. "applying the test is likely to give much better information about the reason for a lookup failure -- information that may be usefully passed to the user when that is feasible -- than DNS resolution failure information alone" might this lead to the idea that they could also be carried in case of the failure to better document it?
- 5.4. "For all other strings, the lookup application MUST rely on the presence or absence of labels in the DNS to determine the validity of those labels and the validity of the characters they contain". Is it correct to assume that the first labels stand for "A-Label" and the second one stands for "their corresponding U-Labels"?
- 5.4. (Martin Dürst): para 1: Mishmash again. Most of this para is best removed.

- 5.4. (Martin Dürst): para 1: "Putative labels": Both in Section 4 and 5, labels are for the most part putative, because they don't conform to the definitions unless checked. Either before section 4, or once at the start (Input subsection) of both section 4 and section 5, say that for the most part, we are dealing with putative labels, but 'putative' isn't repeated all the time to make the text easier to read.
- 5.4. (Martin Dürst): page 12: Finally a bullet list. I almost thought that the author didn't know how to create bullet lists, or was of the opinion that bullet lists don't have a place in spec. Quite to the contrary, please make sure there are much more bullet lists. It will make everything much easier to read and clearer.
- 5.4. (Martin Dürst): "Labels that are not in NFC form as defined in [Unicode-UAX15].": There is only one definition of NFC, but the sentence suggests there are several. Please change to "Labels that are not in NFC [Unicode-UAX15]."
- 5.4. (Martin Dürst): Please move bullet 1 (UNASSIGNED) and bullet 4 (DISALLOWED) and all the other table-related bullets together. I think it's best to put UNASSIGNED last (and mention that this is the category most subject to change).
- 5.4. (Martin Dürst): Streamline the wording used to refer to Tables and a category. Currently, we have: "in the UNASSIGNED category of [IDNA2008-Tables]" - "in the "DISALLOWED" category in the permitted character table [IDNA2008-Tables]" that are identified in [IDNA2008-Tables] as "CONTEXTJ"
- 5.4. (Martin Dürst): "Labels whose first character is a combining mark (see Section 4.2.3.2).": Refer directly to the relevant Unicode definition, rather than to section 4.2.3.2 (which contains a MUST, which is already implicit here).
- 5.4. (Martin Dürst): "In any event, lookup applications should avoid attempting to resolve labels that are invalid under that test.": Remove. We already have a SHOULD, no need for a should on top of that.
- 5.4. (Martin Dürst): last para: I assume this is e.g. about labels with mixed scripts,... What it essentially seems to say is that a browser may warn users if it detects mixed scripts, but if the user still wants to see the page, s/he is entitled to it. In such a context, the word 'validity' seems quite a bit out of place; it would be better to speak about 'other tests' or some such in a more general way.
- 5.5. (Martin Dürst): para 1: "using the Punycode algorithm (with the ACE prefix added)": The parenthetical seems to suggest that addition or not of the ACE prefix is an (optional) part of the Punycode algorithm, but [RFC 3492](#) does not define the prefix, nor is the addition of the prefix part of the punycode algorithm. -> Convert parenthetical to a clause or sentence ("... and then adding the ACE prefix." or so).
- 5.5. (Martin Dürst): rest from second sentence in para 1: As said in my comments on Section 4, a summary is unnecessary. Also, it has nothing to do with punycode conversion. In addition, the second bullet point is confusing, because an A-label (checked or not) cannot be punycode-converted again. -> remove

- 5.6. (Martin Dürst): "That ... string" -> "The string resulting from the conversion in Section 5.5"
- 5.6. (Martin Dürst): "That lookup" -> "The lookup"
- 5.7. (Martin Dürst): What about (streamlined): Security Considerations for this version of IDNA are described in [IDNA2008-Defs], except for the special issues associated with right to left scripts and characters, which are discussed in [IDNA2008-BIDI].
- 7. IANA Considerations - There is no commitment from UNICODE to not update those Unicode documents that are accepted as normative in the IDNA documentation set. Should their copy at the time of the publication of this set not be stored by the IANA?
- 8. (Andrew Sullivan): I suggest "This second-generation version would not have been possible without the work that went into that first version, due to its authors ..." Or something like that.
- 8./9.(Martin Dürst): These should be merged. The text explains it all.
- 8. (Martin Dürst): "Hoffman and Costello ... should not be held responsible for any errors or omissions.": Remove, this is implicitly clear, in the end it's the WG and the IETF that's responsible. Similar for "As is usual with IETF specifications, while the document represents rough consensus, it should not be assumed that all participants and contributors agree with all provisions."
- 9. (Andrew Sullivan): I object very strongly to the inclusion of the sentence, "As is usual with IETF specifications, while the document represents rough consensus, it should not be assumed that all participants and contributors agree with all provisions." Rough consensus is always rough on everyone, but if you are a participant who urges this sentence on the product of the WG, I ask you to reconsider. It is unworthy of your effort and the efforts of your colleagues. It would be better to have an outright flamewar on the mailing list than to have that sort of not-with-a-bang-but-a-whimper remark live forever in the WG output. If we as a WG really have such deep disagreements that we have to send drafts with this sort of disclaimer to the IESG, I feel pretty uncomfortable that the WG has in fact reached consensus.
- References. (Martin Dürst): [Unicode-RegEx], [Unicode-Scripts], [Unicode-UAX15] (and maybe others): Unicode data files don't have explicit authors, but Unicode TRs (and similar stuff) has authors/editors, same as RFCs. Please don't drop this information.

## IDNA BIDI

- (Paul Hoffman): This draft is not yet ready for publication. The text is still very confusing about the relationship between this document and [RFC 3454](#). In many places, the wording makes it sound like this new algorithm is a replacement for that in [RFC 3454](#), which it is not: it is the algorithm to use with IDNA2008. I would be happy to do a thorough edit to remove this ambiguity and make it clear that, while the algorithm is an improvement on the old one, it does not "change" or "fix" or "replace" the old one.

**Author's current answer:** It changes the old definition in exactly the same meaning of the word "change" as the way IDNA2008 changes IDNA2003.

**Author's current answer:** I'll send you the XML under separate cover; if you can make the changes you feel are needed, I can see if I agree with them.

- (Paul Hoffman) The terminology section needs to define "the end of the label". Tests 3 and 6 of section 2 are confusing without some definition.

**Author's current answer:** Will add "beginning" and "end" to the paragraph that says "<t>In this memo, we use "network order" to describe the sequence of characters as transmitted on the wire or stored in a file; the terms "first", "next", "previous", "before" and "after" are used to refer to the relationship of characters and labels in network order.</t>

- Abstract, and potentially elsewhere. (Martin Dürst): Avoid the word 'new'. RFCs are archival documents.
- 1.1. Advisable or not to specify "when U-labels" instead of "labels" ?

**Author's current answer:** The first paragraph says that the document is about U-labels. This should not need repeating.

- 1.1. (Martin Dürst): para 2: "When labels satisfy the rule, and when certain other conditions are satisfied, they can be used with a minimal chance of these labels being displayed in a confusing way by a bidirectional display algorithm.": "they" .. "these labels" is confusing. What about "When labels satisfy the rule, and when certain other conditions are satisfied, there is only a minimal chance that these labels will be displayed in a confusing way by a bidirectional display algorithm."
- 1.1. (Martin Dürst): "A bidirectional display algorithm": How many of them do we have? (I only know one, the Unicode one (with some minor variants)). How many of them have been used for testing/verification?
- 1.1. (Martin Dürst): para 3: what exactly is a "right-to-left character"?
- 1.2. (Andrew Sullivan): "While the document proposes completely new text, most reasonable labels that were allowed under the old criterion will also be allowed under the new criterion, so the operational impact of the rule change is limited." It would be nice here, I suggest, to offer some definition of what labels fall outside "most reasonable labels". The description sounds too much like, "The labels we

picked when we wrote this," which is indubitably not the impression anyone intended.

**Author's current answer:** I added some examples (mixtures of numerals, and AN inside LTR labels). Hope it helps.

- 1.2. (Martin Dürst): This section ideally should also be moved to after Section 2.
- 1.2. (Martin Dürst): para 1: "The IDNA specification "Stringprep"": change to something like "Stringprep, part of IDNA2003". Otherwise, it's not clear that this is an old spec.
- 1.2. (Martin Dürst): para 4: "However, this makes certain words" -> "However, this made certain words" (past tense)
- 1.2. (Martin Dürst): para 7: "While the document specifies rules" -> "While this document specifies rules"
- 1.2. (Martin Dürst): para 7: "(the most important being label that mix Arabic and European digits (AN and EN) inside an RTL label, and labels that use AN in an LTR label)": Very weird. Such cases may not be completely impossible, but they are much less frequent than e.g. Arabic numbers inside Arabic letters, European numbers inside Arabic letters, and so on. There was even a strong movement to prohibit number mixing at the protocol level; this would never have happened if such mixing would have been deemed to be "most important". Also, after looking at the actual conditions, we either have an RTL label, which by condition 4 excludes mixing EN and AN, or we have an LTR label, which by condition 5 excludes AN and therefore the mixture of EN and AN.
- 1.3. (Martin Dürst): title: "Layout" -> "Structure" or "Organization"
- 1.3. (Martin Dürst): para 1: Change from "bidi test" to "bidi rule". (or unify otherwise)
- 1.3. (Martin Dürst): para 1: ", that" -> ", which"
- 1.3. (Martin Dürst): para 1: "no matter what the direction of the label is": What does this mean? It could either mean that you can apply the test forwards or backwards, or it could mean that it doesn't depend on what directionality the characters in the label have, or whatever. In the later case, I'd write e.g.: "This test [->rule, see above and below] can be applied to any kind of label, but becomes trivial if the input is guaranteed to contain only LTR characters."
- 1.3. (Martin Dürst): "The primary initial use of that test": "that test" -> "this test" (this sentence talks about relationship with other documents, so it's the test in this document, not the test in that other section)
- 1.3. (Martin Dürst): para 2: "a BIDI rule" -> "the BIDI rule"

- 1.3. (Martin Dürst): para 3: "new rule proposed here" -> "new rule proposed" (we are talking about document organization, so it's "the rule in that other section over there", so "here" doesn't fit)
- 1.3. (Martin Dürst): para 4: "Section 5 to Section 9 describe" -> "Section 5 to Section 7 describe": Section 8 is IANA consideration.
- 1.4. (Mark Davis) "An RTL label is a label that contains at least one character of type R or AL." I believe you should also add "AN". There are cases where it affects ordering. What I mean is that if you had AN + L in a label (not nec in that order), you wouldn't even count it as a BIDI domain name, and thus none of the bidi doc would apply (according to the text). Yet such labels would be legal according to protocol, and I think they can cause reordering, and could thus cause the kind of visual confusion that BIDI is supposed to prevent.

**Author's current answer:** Good point. I'll modify.

- 1.4. BIDI properties come from Unicode. They might not be complete or could be completed in the future. What then?

**Author's current answer:** See section 7.2, "This memo does not propose a solution for this problem".

- 1.4. (Andrew Sullivan): There are some terms defined in Para 1.4. I think it would be way helpful to a naive reader to be directed to defs at the beginning of this section first, and then to say "there are specific BIDI-only terms also defined here". So I'd move the reference that's at the end of this section to the beginning.
- 1.4. (Andrew Sullivan): The third paragraph now ends with a comma, so it looks like something was supposed to be added and wasn't. Or is this just a typo?

**Author's current answer:** Typo, fixed.

- 1.4. (Andrew Sullivan): I find this peculiar: A "Bidi domain name" is a domain name that contains at least one RTL label. If a domain name is RTL.RTL.RTL, it qualifies under this definition, even though there is no bidirectionality (all labels have the same directionality). Explaining why this is still "bidi" would leave me less confused.

**Author's current answer:** Added some text to say "adding a separate RTL-name category would just make the spec more complicated".

- 1.4. (Martin Dürst): "non spacing" -> "nonspacing"
- 1.4. (Martin Dürst): "The directionality of such examples" -> "The display order of such examples"
- 1.4. (Martin Dürst): "it means ..., approximately" -> "it approximately means"

- 1.4. (Martin Dürst): "An RTL label": This seems to be the definition that Protocol might want to refer to.
- 1.4. (Martin Dürst): 'Having a separate category of "RTL domain names" would not make this specification simpler, so has not been done.' -> 'Providing a separate category of "RTL domain names" would not make this specification simpler.'
- 2. A replacement for the [RFC 3454](#) BIDI rule: it would probably be good to indicate the applying order.

**Author's current answer:** The 6 conditions can be checked in any order. All must be satisfied in order to make the test pass; different implementations may find that different checking orders make the code more or less efficient

- 2. (Martin Dürst): (title), and elsewhere: Both "Bidi rule" and "Bidi test" are used, that's confusing. The term is always in singular. The document works that way in general, but "The following test" at the start of Section 2 is confusing, because the only 'tests' that one can see are the ones labeled 1. to 6. Maybe use something like "In order to pass the BIDI test, the following conditions 1. to 6. must all be satisfied."
- 2. (Martin Dürst): conditions 2/4: Why are BN (control characters) allowed in RTL but not in LTR?
- 2.1. (Andrew Sullivan): Rule 1 in Para. 2 says, "The first character must be a character with BIDI property L, R or AL." I can't tell whether that must is a requirement or a statement of fact that is entailed by other IDNA rules. If it's a requirement, it presumably ought to be a 2119 MUST; even if not, it seems that we have to know what to do in case the first character doesn't match this rule. If it's an entailment, it'd help to make that plain, which could be done by restating it, "The first character will be a character with BIDI property L, R, or AL due to [reason reference]."

**Author's current answer:** I'm not sure what an entailment is - but this is a rule. In order to execute the text, you must look at the first character and check its BIDI property. This document isn't using 2119 (wow); I used to have the reference way back when, but it didn't seem to help any, so I removed it.

- 3. (Andrew Sullivan): "One specific requirement was thought to be problematic, but turned out to be satisfied by a string that obeys the proposed rules:
  - The Character Grouping requirement should be satisfied when directional controls (LRE, RLE, RLO, LRO, PDF) are used in the same paragraph (outside of the labels). Because these controls affect presentation order in non-obvious ways, by affecting the "sor" and "eor" properties of the Unicode BIDI algorithm, the conditions above require extra testing in order to figure out whether or not they influence the display of the domain name. Testing found that for the strings allowed under the rule presented in this document, directional controls do not influence the display of the domain name."



comes after the discussion of things considered and rejected. This leaves me confused about whether the text above is in fact a requirement or not. If it is a requirement, then I'd move this segment to the part before the rejected requirements.

**Author's current answer:** Added a little more text to explain the status of the requirement - it's a "nice to know".

- 3. (Martin Dürst): "A requirement" -> "The requirement" (see above)
- 3. (Martin Dürst): para 2: As this restricts things to the Unicode bidi algorithm, please say this earlier. (see above)
- 3. (Martin Dürst): para 3: "requirements proposed" -> "requirements" (we are working on finalizing this document, we are no longer in the proposal stage)
- 3. (Martin Dürst): requirement 2: Is the choice of 'characters delimiting the labels' open, is this only the ASCII dot, is this a small set (I'm interested in this both for spec clarity and because the answer might strongly affect draft-duerst-iri-bis).
- 3. (Martin Dürst): 'possible requirement' related to directionality controls: "(outside of the labels)" -> "(outside, but potentially directly adjacent of the labels)" (does this include cases with directionality controls inside a domain name, i.e. before/after a dot?) "the conditions above require extra testing" -> "the conditions above required extra testing"
- 3. (Martin Dürst): 'Delimiterchars': FULL STOP not allowed in domain names?????
- 4.1. (Martin Dürst): Thaana 'Computer' example: "UBIUFI" -> "UBUFI"
- 4.2. (Martin Dürst): This section could be shortened considerably. "Greater latitude here than ... Dhivehi." is irrelevant; as long as a significant part of a language's words cannot be used in IDN, there's a problem. The subsection is interesting for people interested in Yiddish, but the average reader of the spec will try to find something relevant for the algorithm, and mostly be more confused than enlightened.
- 4.2.3.4 (James Mitchell): If the proposed label contains any characters that are written from right to left it MUST meet the "bidi" criteria [IDNA2008-BIDI]. The above implies that the label must meet the BIDI criteria, however BIDI criteria is applied to a BIDI domain name. From draft-ietf-idnabis-bidi-04, The following test has been developed for labels in BIDI domain names and A "Bidi domain name" is a domain name that contains at least one RTL label. I hesitate to provide alternative text for until the following question has been answered.

The protocol states that the proposed label contains any characters written from right to left it MUST meet the bidi criteria. It does not impose such requirements on labels containing no right to left characters. Consider the registration of label 123abc in a zone containing labels written right to left. The label does not contain any right

to left characters, therefore does not have to meet the BIDI criteria. However this name is a BIDI domain name, yet such a name would fail as the first label (LTR) does not begin with BIDI property L. Is the label intended to be valid for registration given a right to left zone?

**Author's current answer:** In your specific case, registering "123abc" in the RTL zone "ABC" (usual convention applies) will lead to a domain name (network order: 123abc.ABC) that, on its own, will display as "123abc.CBA" in an RTL context, but if prepended by "DEF:", forming the network order string DEF:123abc.ABC, will display as "CBA.abc123:FED" - which may be surprising to some. (I'm only 90% confident on this - people more used to bidi in practice may be more confident).

The WG has rejected inter-label tests, therefore all tests defined by the protocol as normative (MUST or SHOULD) apply only to one label at a time. Given the WG decision, I tried to make the BIDI document quite clear that certain properties can only be guaranteed for domain names where all the labels meet the test, but this is a case where people have to read the warnings and do something reasonable, rather than having the rules define that being unreasonable is forbidden. "Warning: Contains hot liquids".

I am not concerned about the display ordering of the name in question. The issue is a mismatch between the registration and lookup protocols.

The registration protocol asks the question of the label '123abc', which left-to-right is not required to satisfy BIDI. However, the lookup protocol says that one SHOULD apply the BIDI test (on I assume the name). Applying the BIDI test to this name will fail and the name will not be looked up.

As a registry, should I allow registration of the name 123abc.RTL?

**Author's current answer:** My recommendation is that you should (as a registry) establish policy that says "it is not allowed". The protocol does not require you to do so.

As an application, should I lookup the name 123abc.RTL?

**Author's current answer:** The protocol does not say that you can't. For the obvious reasons, I think it's a legitimate implementor decision to decide not to.

- 4.2.3.4 (James Mitchell): I note also the inconsistent use of the term BIDI. In draft-ietf-idnabis-protocol-14 section 4.2.3.4. it is quoted and in lower-case, whereas the draft-ietf-idnabis-bidi-04 uses the upper-case version extensively. Also, within draft-ietf-idnabis-bidi-04 the term "Bidi domain name" in Section 1.4 is inconsistent with BIDI domain names in Section 2, and the tem Bidi rule in Section 10 is inconsistent with the several other occurrences in the document.

**Author's current answer:** Thanks, I tried to normalize it to uppercase in an earlier round, but didn't remain consistent in later edits. Will fix!

- 4.3. (Martin Dürst): "(with the 5 being considered right-to-left because of the leading ALEF)": No, the 5 itself is never right-to-left. Change to "(the overall directionality being right-to-left because of the leading ALEF)"
- 4.3. (Martin Dürst): "but barring them both seems to require justification" -> "but barring them both seems unnecessary" or "but barring them both turned out to be unnecessary"
- 5. (Martin Dürst): "Even if a label is registered under a "safe" label,": 'under' should be explained more clearly (I assume this refers to the hierarchical relationship in the DNS)
- 5. (Martin Dürst): last paragraph: It would be better to change this into a SHOULD, such as "Where implementations see a way to avoid ..., they SHOULD avoid". That will bring this issue on the radar screen of implementers, whereas it currently will just be glossed over.
- 6 (Mark Davis)"Rules can also be specified at the protocol level, but while the example above involves right-to-left characters, this is not inherently a BIDI problem." I think the issue is that the word "can" was appropriate for when this was a proposal, but the situation is different in heading for release; the "can" should be changed according to what you mean.
  - If you are referring to the situation as of when these are all released, then the rules either "are specified" or they are not (in which case the statement removed).
  - If you are referring to a future time, then the "can" becomes "could" (or for clarity, adding "in a future version" or some such language).
  - If you are referring to what could have been done, then "could also have been" would be appropriate.

Because I can't tell what you want to say, I don't know which of these you would mean.

**Author's current answer:** There's no guarantee of synchronicity in further updates, so the situation isn't really all that different (BIDI doesn't place any constraint on future -tables), but I'll change to "are".

- 6. (Martin Dürst): first paragraph: "All other issues with these scripts": What scripts???
- 6. (Martin Dürst): "wishes to create rules for the mixing of digits" -> "wishes to create rules against the mixing of digits" or "wishes to restrict the mixing of digits"

- 6. (Martin Dürst): "Rules are also specified at the protocol level, but while the example above involves right-to-left characters, this is not inherently a BIDI problem." -> "This example is not inherently a BIDI problem, so such restrictions are not specified at the protocol level." ("Rules are also specified at the protocol level" is inherently vague; it seems to mean "Some rules against mixing digits are also specified at the protocol level, but only when this is necessary to avoid a BIDI problem.")
- 6. (Martin Dürst): "It is unrealistic to expect that applications will display domain names using embedded formatting codes between their labels (for one thing, no reliable algorithms for identifying domain names in running text exist);": Please add that it is also unrealistic that formatting codes are removed before IDNA processing, and that allowing formatting codes could lead to many kinds of 'mischief' that would go against the two requirements in section 3.
- 6. (Martin Dürst): "which might surprise someone expecting to see labels displayed in hierarchical order.": Please add that this may not be such a big problem to general users familiar with BIDI, because they are used to seeing/reading a sequence of RTL units (e.g. words) from right to left. (for wording alternatives, see <http://tools.ietf.org/html/rfc3987#section-4.4>, first para, \*second para\*, ...)
- 7. Does that restriction mean that telephone numbers cannot be registered in BIDI zones?

**Author's current answer:** If the registry desires that domain names behave sensibly, yes; if the registry only desires that domain names pass the test, no. There are no inter-label tests.

- 7.1 (Martin Dürst): Bullet points 1 and 2 are major, whereas bullet point 3 is really farfetched (not impossible just because there is no guarantee against weird implementations). It would be good to indicate that somehow. (this includes the paragraph following bullet point 3)
- 7.1. (Martin Dürst): "The editors believe": change to something less specific; this is a WG document, we either have rough consensus or we don't. (I for one fully agree with this point)
- 7.2. (Martin Dürst): This should be slightly reworded to more clearly send the message that changes to Unicode bidi properties, while not totally impossible, are expected to be rare, and to affect mostly symbols and the like, which will limit their effect on what the BIDI rule(/test) allows and what not.
- 8. IANA considerations. Same remark as in the Protocol case.

**Author's current answer:** The Unicode Consortium does not make changes to published versions of its standards; I believe we can trust them to keep version 5.1 available for a while.

Moreover, the section above then states: "the determination of validity for any string depends on the Unicode BIDI property values, which are not declared immutable by the Unicode Consortium."

**Author's current answer:** See section 7.2.

- 8. (Martin Dürst): "It is possible that differences in the interpretation of the specification": Wrong. There are no differences in interpretation for the old spec. There are no differences in the interpretation of the new spec. There are differences in the specs themselves.

## IDNA Tables

- disorder in paragraphs.

**Author's current answer:** The exact order of the sections will be decided upon (and sections will potentially be moved around) at the time of publication by the RFC Editor while doing other formatting changes.

I have been in contact with the other document editors, and our suggestion to our wg chair is to *\*NOT\** risk destroying the actual content of the messages by moving things around at the time of last call (both wg and IETF).

- (Gihan Dias): Tamil digits. John Klensin: "I see a considerable difference between, e.g.,  
- "exclude Tamil numerals"  
and

- "this character looks like that character, so exclude one of them".

The latter is clearly part of a case-by-case character analysis. The former, whatever it might be, is a decision about a class of characters, whether Unicode's selection of properties identifies it as a class or not.

The Sri Lanka IDN Task Force considered the document draft-ietf- idnabis-tables-06.txt and has an issue with the inclusion on the Tamil Digits as valid IDNA characters.

```
OBE6..0BEF ; PVALID # TAMIL DIGIT ZERO..TAMIL DIGIT NINE
```

However, we consider that the potential for confusion is sufficient that they be disallowed in the protocol, and request that they be excluded.

**Author's current answer:** Given the current rules, and the properties in the Unicode Database, the only way to treat OBE6..0BEF as DISALLOWED is to add them to the exceptions table one by one. I.e. add them to section "2.6. Exceptions (F)" with the explicit value DISALLOWED.

- 1. Introduction. "In particular, some combinations of allowed code points are not advisable for use in IDNs due to rules specific to a script or class of characters"

introduces the concept of a "class of characters", but does not document it. IDNA Rationale 7.1.3 states "Maintain IDNA and Unicode tables that are consistent with regard to versions, i.e., unless the application actually executes the classification rules in [IDNA2008-Tables]" yet the only time "classifications (rules) appears" in IDNA Tables is in "4. Code points" as "The Categories and Rules defined in Section 2 and Section 3 apply to all Unicode code points. The table in Appendix B shows, for illustrative purposes, the consequences of the categories and classification rules, and the resulting property values."

What is a "class of characters"?

- 1. Introduction ends with " This document is part of a series that, together, constitute a proposal for updating the IDNA standards to resolve issues uncovered in recent years, cover a broader range of scripts, and provide for migration to newer versions of Unicode. See [IDNA2008-rationale] for a broader discussion. " Should this not be removed or edited?
- 2.1. "For more information, see section 4.5 of The Unicode Standard [Unicode5]." Is it also the case in Unicode 5.1? Shouldn't this document be stored by the IANA?
- 2.2. NFKC or NFC?
- 2.2. (Paul Hoffman) Section 2.2 uses NFKC, but the protocol itself uses NFC. I think it is useful to make a note to this effect in Section 2.2.
- 2.8. (Andrew Sullivan): "JoinControl (H)  
H: Join\_Control(cp) = True  
This category consists of Join Control characters (i.e., they are not in LetterDigits (Section 2.1)) but are still required in IDN labels under some circumstances. They require extended special treatment in Lookup and Resolution."  
I think we previously agreed just to call the action "lookup". Strictly, all the special treatment is part of the lookup process, but not the resolution process (which is a straight DNS activity that happens to be using an A-label as its QNAME). As I've argued before, I want the documents to stay very far away from any suggestion that they are changing the operation of the DNS as such.
- 2.10. "It should be noted that Unicode distinguishes between 'unassigned code points' and 'unassigned characters'". Can the differences (nature and in relation to IDNA) between the characters and codepoints be explained here?
- 5. IANA consideration. It is suggested that IANA should retain online copies of the version of external documents that are normatively referenced in the IETF documents.
- 7. (Andrew Sullivan): " As is usual with IETF specifications, while the document represents rough consensus, it should not be assumed that all participants and contributors agree with all provisions." I'll spare participants my speech on why this is a bad thing this time.

- "A table from which that registry can be initialized, and some further discussion, appears in Appendix A. " - Who is to decide and maintain the table and according to which rules/procedures?
- Appendix A. as a comment, we do not understand, from the presented kind of logic, as to why:
  - Tamil digits cannot be made subject to a rule and added to CONTEXTO?
  - The same for French majuscules?
  - The same for any zone specific restriction?

It seems implied that the logic should be the same on the sending and receiving end? The receiving end is only for decoding what the sending end chose to encode in its own context. That context needs to be considered and supported. If my application is in Tamil or French, it knows it and can be demanded to proceed accordingly.

- Appendix A (Andrew Sullivan): paragraph "Note that "Before" and "After" do not refer to the visual display order of the character in a label, which may be reversed or otherwise modified by the bidirectional algorithm for labels including characters from scripts written right-to-left." might benefit from the addition of another sentence, "Instead, 'Before' and 'After' refer to the network order of the character in the label."
- Appendix A (Andrew Sullivan): "Appendix A.7. KATAKANA MIDDLE DOT  
Code point:  
U+30FB  
Overview:  
Note that the Script of Katakana Middle Dot is not any of "Hiragana", "Katakana" or "Han". The effect of this rule is to require at least one character in the label to be in one of those scripts. ...." there is no "End For" as called for in the pseudocode definition.

# Annex 6.

## Punyplus

This memo documents an IDNA proposition (AKA IDNAPLUS), which supports upper and lowercases and new possibilities for the Internet name space.

### Contents

- [1 Introduction](#)
- [2 Discussion list](#)
- [3 The Interplus approach](#)
- [4 Punyplus algorithm](#)
  - [4.1 Support as a regular character](#)
  - [4.2 Support through a dedicated format](#)
- [5 Value of the UMI codepoint](#)
  - [5.1 To use a DISALLOWED codepoint](#)
  - [5.2 To use an UNASSIGNED codepoint](#)
  - [5.3 To use a PVALID codepoint](#)
  - [5.4 To register a dedicated codepoint](#)
  - [5.5 Solution](#)
- [6 Naming added services](#)
- [7 Further extensions](#)
- [8 Multilingualism section](#)
- [9 Security section](#)
- [10 IANA section](#)
- [11 Acknowledgments](#)
- [12 References](#)

### Introduction

According to the current WG-IDNABIS proposal, which is based on the punycode support algorithm, the entries in domain name locations should be subjected to a systematic preceding lowercasing without later restoration.

This means that uppercases are not supported on an end-to-end basis and that uppercased domain names are resolved as if they were their lowercase equivalent. Therefore, half of all the characters of the world's alphabets are thereby ignored by the IDNA. The WG-IDNABIS considers this to be perfectly acceptable. The WG-IDNABIS also believes that this does not significantly affect the overall "internationalized" domain name usage.

The resulting semantic loss by orthotypographic limitation makes IDNA unfit for:



- a good steganographic use of the namespace (mnemonics, encryption, etc.)
- a direct support of the Semantic Addressing System (SAS) that the IUCG is exploring.

This memo documents a punyplus algorithm that fully supports the SAS and offers a strictly compliant IDNA option.

## Discussion list

This proposition can be discussed on the [iucg@ietf.org](mailto:iucg@ietf.org).

## The Interplus approach

The Interplus is an Internet "better use" architecture to bring users better Facilitation services. It deploys additional layers on the user side in a manner that is transparent to the Internet layers' operations. It is in this way that it can support extended network functions (active content support) and additional services (such as the support of ambient content).

The Interplus architecture includes an ML-DNS (multi-layers domain name system) pile. This pile comprises three naming layers:

- User Domain Names (UDNs) that can be case sensitive full UTF-8
- Application Domain Names (ADNs) that can be case sensitive NFC
- Internet Domain Names (IDNs) that comprise alphadecimal (0-Z) labels (case insensitive by nature) in using strong (".") separators, and a weak ("--") separator between the sub-labels.

Empty sub-labels (resulting in a "--" separation) are understood as the operationally validated way to support the Internet Presentation virtual layer, in which they separate a presentation prefix (or suffix) from the naming payload. "xn--" is understood as the "extended names presentation" prefix.

## Punyplus algorithm

The punyplus algorithm is used to convert (back and forth) ADNs into IDNs. An option of punyplus conforms to the punycode algorithm and does not support uppercases.

To reliably obtain this, punyplus is a strict punycode copy with the added capacity, (as a removable option) to insert the "^" prefix to signal the replacement of an uppercase by a lowercase. There are then two possible ways to support this insert.

## Support as a regular character

When executing the punycode algorithm itself, it converts "^" as a non-ASCII (UMI) uppercase metadata indicator and then handles its code point in the regular way.

When converting back to non-ASCII, it restores the UMI as "^", or deletes it if the next character is restored as an uppercase.

Exemple:

UDN: User domain name : Etat.fra

ADN: Application domain name : ^etat.fra

IDN: Internet domain name : xn--etat-abc.fra (abc depending on the choice of UMI).

## Support through a dedicated format

The "^" insert is replaced by the "---1" sequence. The advantage of this solution is its independence from the coding system, that it is open to the support of other needs by sequence of "---x" or "---zz" type, its readability is simple and direct entries can be easily manually made. Its disadvantage is the user of four or five bytes to send a metadata.

Exemple:

UDN: User domain name : Etat.fra

ADN: Application domain name : ^etat.fra

IDN: Internet domain name : xn--e---1tat.fra

## Value of the UMI codepoint

There are four possibilities for the choice of the UMI code point:

### To use a **DISALLOWED** codepoint

This will be transparent to IDNA unaware applications and uncertain with IDNA2003 applications, but blocked by IDNA2008 compliant applications. There is no risk of confusion with a lowercase described destination.

### To use an **UNASSIGNED** codepoint

This will be transparent to IDNA unaware applications and uncertain with IDNA2003 applications, but blocked by IDNA2008 compliant applications. A conflict may result from a further edition of ISO 10646. Otherwise, there is no risk of confusion with a lowercase described destination.

## **To use a PVALID codepoint**

The code point would be out of regular use sequence, therefore signaling a special use. This will be transparent to IDNA unaware applications and accepted, but not necessarily understood, by IDNA2003 and IDNA2008 compatible applications. There is no risk of confusion with a lowercase described destination.

## **To register a dedicated codepoint**

This will be transparent to IDNA unaware applications and uncertain with IDNA2003 applications, but blocked by IDNA2008 compliant applications. There is no risk of confusion with a lowercase described destination. A further IDNA update could make that code point PVALID and extend IDNAPLUS to all the IDNA users.

## **Solution**

These possibilities will be submitted to tests and discussion with users.

## **Naming added services**

The punyplus algorithm being transparent to pure ASCII UDNs/ADNs, it is applicable to all Internet domain names. It can therefore be used as an ML-DNS naming command parser and universal UDN/DNA/IDN converter. Its concerted generalization will therefore participate in an ambient canonicalization of the naming space.

## **Further extensions**

The punyplus algorithm will progressively be extended to support DNS related security projects, linguistic mail addresses, DNS class selection, semantic addressing and Netix (interplus explored interapplication system) commands.

## **Multilinguistics section**

Multilinguistics is understood as the cybernetics of linguistic diversity. This memo enables improved Internet multilinguistic support.

## **Security section**

The ability to use a larger number of code points that are denied; or an additional alphadecimal character in naming (corresponding to the "power" sign) that is blocked by all the naming functions, which in turn should not create security issues for those applications that are properly written or those interfaced by properly written applications.

The inability to use uppercases dramatically reduces the interest of randomly generated or encrypted domain names in access protection strategies.

The global stability of the Internet will probably be affected by some uses taking, like IDNA, a better advantage from its protocols. This should motivate a wide separated study. The author things that the punyplus practice may bring a positive added experience in the study and the solving of the new problems that may be encountered.

## **IANA section**

The selected value of the UMI should be recorded by the IANA.

## **Acknowledgments**

I particularly thank the members of the france@large center of expertise on IDNs for their dedicated support and pertinent contributions.

## **References**

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, [RFC 2119](#), March 1997.

## Annex 7.

### Projet.FRA report

Dear IESG Members,

I understand that there is an IESG conference call on Jan 7 that may discuss IDNA. I had hoped that I could publish a general framework I\_D prior to that date on the practical evolution of Internet usage, in turn covering IDNA as a working example. A family emergency has delayed that work by a few days. This mail is expressed within an IDNA perspective. FYI, I chair Projet.FRA for a French speaking namespace.

IDNA has been specified as permitting a support of IDNs without any DNS change. I identify five layers in these proceedings, which are:

1. the internet layer
2. the IDNA interface protocol
3. the user implementation of this protocol
4. the impact on the network naming topology
5. the technical and political interinfluence of the resulting changes.

### Contents

- [1 The Internet layer:](#)
- [2 The IDNA interface protocol:](#)
- [3 The user implementation of this protocol](#)
- [4 The impact on the network naming topology](#)
  - [4.1 Impact on the virtual root file](#)
  - [4.2 Impact on the Internet usage architecture](#)
    - [4.2.1 Principles](#)
    - [4.2.2 Requirements](#)
    - [4.2.3 Interplus](#)
- [5 The technical and political interinfluence of the resulting changes](#)
- [6 Comments](#)

### The Internet layer:

The internet is to provide regular Internet transport, network, and DNS LDH resolution services. This is to remain unaffected. This has been respected thus far.

## The IDNA interface protocol:

This was controverted between Unicode, the IETF culture, and us. Moreover, the proper support of French is turning out to be the most complex issue due to the use of the same Roman charset as English, while the French key concept of "majuscules" is absent in English and Unicode. The success of IDNA2008 is to have reached a workable consensus between these three positions (however proper French [Latin languages] orthotypography obliges ".FRA" to find an additional way to support them). We designate IDNA2010 a BCP Draft project that would document the various ways IDNA2008 is/can be implemented by Zone Managers. For the time being, the related `workon@idna2010.org` mailing list is only for the informational purposes of its respective members. Active debate will only be started in coordination with the WG/IDNABIS Chair, so that there is no confusion. To date, there are people of various origins on the list, but none from ICANN.

## The user implementation of this protocol

In suppressing Nameprep, IDNA2008 means progress in flexibility. However, this flexibility may lead different applications on the same machine to resolve the same IDN to different IP addresses. This means that there is not only a problem of transition, but also a problem in the architecture of the user's machine in order to address this risk. As initially indicated, I had planned to address this need in full compatibility with IDNA2008 via the ML-DNS approach (cf. *infra*) I announced, as soon as IDNA2008 would be approved by the IESG. Lisa Dussault preferred raising the question once IDNA2008 was completed but still not yet approved. This provides you with full control over the decision, but you now need to know the resulting upward and downward contexts better.

The simplest and most rewarding solution is an "IDNApplication" that:

1. respects the IDNA architectural principle of U-Labels that are to be dealt with at the user application layer.
2. intercepts all the domain name entries
3. transcodes them before sending them to the DNS in:
  - leaving ASCII domain names unchanged.
  - differentiating the few TLDs supporting IDNA2003 in order to address them specifically.
  - respecting IDNA2008 otherwise.

In this way all applications and protocols will be provided transparently with the same adequately formatted labels. In addition, IDNA2003 transition schemes will be possible at dates that are decided by each of the concerned TLD Managers.

The fastest and easiest ways to deploy this IDNApplication are either:

- to use an OPES like front-end to existing nameservers (work on this was foreseen but not completed by the WG/OPES)
- or to embed the support of punycode (or rather, "punyplus" (<http://tools.ietf.org/html/draft-iucg-punyplus-03.txt>) within new various nameservers releases.

## The impact on the network naming topology

There are three quick (and, therefore, stabilizing) large scale and easy to disseminate deployment strategies:

- ISP "patching", as is currently and similarly done for Chinese Domain and Key Names.
- public DNS services, such as Google's PDNS proposition that may help enforce that type of service immediately on a large scale.
- implementation of non-external-caching DNS servers at the user's machine.

### Impact on the virtual root file

This happens while a demand for many more IDNg/ccTLDs than ICANN is prepared to accept in their Root file. In addition, ICANN in Seoul has strategically and commercially favored some non-Roman IDNccTLDs that will enter the NTIA root file through the "Fast Track" experiment; private, civil, and government IDNgTLDs projects that were previously incited to be planned and to apply will probably be delayed for several years.

The solution for these rebuked TLDs candidates (like .FRA) is very simple: to be declared as ULDs at the user's nameservers (for improved clarity, we call "user level domain" the top level or New.Net style domains that are declared by users). This means, in other words, local root files with all the TLDs and ULDs that each user needs. At a PDNS, this will call for the support of all of them. All of this amounts to acknowledging the need to manage the virtual root file that has already been in use for years (cf. the way Chinese TLDs are declared in top level nameservers).

### Impact on the Internet usage architecture

As indicated, the user innovative project (.FRA) introduced by france@large, together with the IUCG early @large participants, committed to document an IDNA2008 extension called ML-DNS (multi-layer). ML-DNS is part of an Internet Usage architectural framework that is able to provide a comprehensive and robust basis to the semiotic strata (Intersem) of which .FRA is interested in order to support intercomprehension facilitation (Internet of the Users and of their thoughts).

## Principles

Its principles are:

- (1) the strict equivalence (synonymy) of a domain name pile ranging from the UDN (user domain name) to the IDN (Internet Domain Name as documented by IDNA2008)
- (2) a generalized extended value IDNA consistent syntax (our rationale is that languages are supported through the presentation layer. If IDNA works, it means that it uses the Internet presentation layer, i.e. in this specific case, the "xn--" presentation).

## Requirements

Our .FRA (as many other TLD projects including the Multilinc set of one sociolinguistic per ISO 639-6/LS 640 linguistic entity) requirements are:

- French (Latin languages) majuscules support (and other similar particularities in other scripts),
- extension of the Internet passive content support to ambient and active contents,
- complete access to presentations and classes,
- a "commuting cloud" of network services open capacity,
- Unicode TR46 considerations
- etc.

## Interplus

All of this is embodied through the "Interplus" (internet plus plugged layers on the user side), which does not change a single Internet bit and conceptually adds on the user side:

- two real layers:
  - an interapplication communications overlay.
  - and a pseudo-network service area. Its general purpose is to support network extended services that users may easily "interplug" for an optional "smart network" experience. Software slots may support services such as:
    - ML-DNS nameserver
    - an Application Firewall
    - Virus protection, social network, Intersem metastructural distributed referential services (MDRS)
- the virtual presentation layer (managed by the ML-DNS along the the "xx--" format, with the "x--" format that is being used for a "Netix" interapplication command set).



## The technical and political interinfluence of the resulting changes

ICANN has imposed pressure on the IDNA text finalization in order to match the Seoul date. This led to some transition confusion (now sorted out) between the WG/IDNABIS and the workon@idna2010.org mailing list. Then, the ICANN Seoul announcements were both perceived as an ICANN opposition to the open deployment of IDNA and as a technically premature move, since Fast Track is more of a legal work than a technical test, and there has been no consideration as of yet of IDNA2008 practical implementation and usage issues.

ICANN participants in WG/IDNABIS either did not perceive the stakes or were not entitled to comment. They did not answer the various questions that we raised regarding their position(s) on IDNA2008 and ULDs.

As a result, we may have:

### (1) four different works on the way in order to implement IDNA2008:

- ICANN Guidelines supporting closed committees representing approx. 120 second level zone managers among millions.
- IUCG workon@idna2010.org open mailing list, targeting an informational BCP for concerted and interoperable implementations of IDNA2008
- some major Public DNS services unilateral policies
- idem for some national deployments.

(2) **confusion over the namespace**, with ULDs popping up here and there without any coordination. Local root dissemination means a heterarchic naming structure. This was foreseen in the ICANN ICP-3 document, but we are the only ones to have run a community test-bed as requested by ICANN. If we want to keep the DNS namespace stable, we need some virtual root cooperative management through an IDNAlliance.

We also have to take into account that:

- class usage is off-the-shelves of any IDNApplication solution
- I only describe the simplest thing that I am to deploy as the Project.FRA Chair in order to get the .FRA namespace operational, at NO change whatsoever except for the loading of a slightly enhanced version of Bind on the participating machines. Everyone else can do the same, and many other TLD project may wish and have better funding to do it. There is NO change in any

way. This is simply a normal reading of the RFCs from an IDNA2008 point of view. The Internet legacy turns out being still more powerful than expected.

## Comments

The decision is yours. You can either approve or delay IDNA2008 and request more information.

- IMHO, if you do not approve IDNA2008, which is certainly excellent work, you will cast suspicion on the entire IDNA scheme, as ICANN will have to delay Fast Track due to an IESG decision. Furthermore, this would delay what needs to be built over IDNA2008. During that time, people would certainly start building it up over IDNA2003, thereby initiating a terrible mess.
- this is why we have started building a technical position against IDNA2008 under the terms imposed by ICANN's urgency (lack of consideration of the abovementioned points). If you approve IDNA (as you technically should), I will appeal against your decision (I fully approve) due to its non considered specific context. I will upset many. However, my expectation is that it would provide many more people with a four to six month respite in order to better organize and work out an IDNA2010 BCP proposition or start a WG/VIRTUALROOT, in turn preventing the current 500 ICANN TLD candidates (and probably many more) from deploying as uncoordinated and most probably conflicting ULDs. In this case, some will initiate lawsuits against others (this has already started): this will result in an unknown US jurisprudence on TLD and ULD attribution that, by essence, no ULD will respect outside the USA.

## Annex 8.

### Mail exchange with Patrik Falström

The AD had asked questions. Patrik Falstrom addressed them as a new thread (as PF1). I myself (JFC) asked questions to Patrik Falström's whose responses (as PF2) are listed in a first part. Not to break the rythm of his valuable responses, I mention (as [JFC.X]) my own answers that I list afterward. Patrik Falström did not answered them yet.

**On 9 feb 2010, at 16.06, jean-michel bernier de portzamparc wrote:**

As requested by Patrik Fälström, I copy these questions to the WG list.

**Patrik Falström responded (PF2):**

It was requested by **JFC**, as I did not want to respond to this private mail in private, and he can not post to this list. So he requested you to forward this to the list.

---

**At 23:15 08/02/2010, Patrik Fältström wrote (PF1):**

Part from standing behind what Vint wrote in his response, let me answer more directly the questions from Lisa.[The Applications AD]

I am a developer/designer of mostly web based software for management of domain names, web sites, email services and what not. I will (and want to be able to) in my interfaces first of all be pretty sure over what data comes to me as a "server". I.e. I want to know what the end user really wanted to give me (I did explicitly not say "typed" here for reasons I hope people understand).

**JFC:** I understand that you are on the server end. You are related to a user and you do not know if this is a direct relation, or it goes through a browser, an iPhone, a Client, a User Agent. Correct? Yet, you also consider the case where you may have to develop the User Client/Agent and you control things down to the user's fingers?

**PF2:** My point is that what people historically believe is "the user interface" sometimes is a split relationship between what today is a client and a server, where the human have access to the client, and the access to the server is from the client via some protocol like HTTP.

I.e. we have often said that potential mappings should be done in the user interface, and I agree with this, but I wanted to give one example where it technically is very hard to do what is the user interface. IF there is a mandate on some mapping, what software should do it? What software will be compliant, and what software will not be compliant?

It is not as easy to answer those questions as some people might think.

[JFC.1]

---

**PF1:** Now, if the web browser (if the web browser was used) did some trick and converted some characters, then I do not get the characters the end user typed (for example), but regardless of this, I will do whatever I want to do to make the end user happy. Exactly what that implies will differ depending on context. Is this domain name management or management of a web site, or the management of email addresses or sites?

**JFC:** So you mean that the reference is what the user wants to achieve?

**PF2:** To some degree, yes.

A user should not be surprised. But then what is surprised?

[JFC.2]

**JFC:** And your understanding of what he wants to achieve will depend on different context inputs, and obviously of the protocol being used and the task to be performed?

**PF2:** Potentially, yes.

[JFC.3]

**JFC:** Also from the zone?

**PF2:** No. And the reason is because of the following:

To be able to give context dependent responses (such as language, username or whatever), the selection mechanism must get information about the context.

And that is not possible to do in the DNS protocol.

[JFC.4]

So some layers in the architecture must be exactly the same, all over the world, for all users, or else it will be impossible to get the base interoperability that is needed to build the systems people wants.

[JFC.5]

Another example, the HTTP and HTML protocols are the same all over the place, but HTTP include a negotiation phase regarding screen size, type of browser etc that makes it possible for an HTTP server to give back different data depending on such information.

HTTP is not different. HTML is not different.

[JFC.6]

---

**PF1:** Is this mapping, if applied, to be implemented in the browser or in the web server? I would like to get as raw data as possible, without having the browser doing anything, so that I know as much as possible on the server side. So I can "help" users the same way browsers today "help" users by adding ".COM" TLD if that is not added. "guessing" I think it can be called.

**JFC:** I understand this is in the case where you have to develop the user client interface?

**PF2:** Sometimes yes, sometimes no. If you for example on a web server use Wordpress software, you as a user of the site do not develop the software. Are you developing the user client interface? I do not know. I would claim I am not doing that. Do the one creating the wordpress theme do that? I do not know.

It is blurry.

But I want to trust all of those developers "to do the right thing".

[JFC.7]

---

**PF1:** And if HTTP/Web browser is not in use, but instead for example an iPhone application that then interact over JSON/Netstring/TCP with some server API, then I can in a more direct way control what is sent to the server (if I also write the client). I will only be confused (possibly) over the mapping that is in the operating system / user interface libraries.

**JFC:** OK. Did you think of a method for you to clarify that possible confusion?

**PF2:** Only by giving me the ability to implement mapping or not depending on what "makes sense" for me, and my application.

[JFC.8]

I am VERY nervous over having any kind of mandatory mappings. Mandatory for whom? Who must implement it?

[JFC.9]

---

**PF1:** All of this might sounds like if it will make things messier for the user, but in reality I do not think so. \*NO\* developer have the interest in making this messier for the user. And exactly what it means to make it easier for the user depends on context. Very much.

Specifically when we have so many protocols and environments and contexts where domain names are in use.

**JFC:** As I said, I am not that much competent in Web level applications, rather with user level interapplications. Which documentation do you use to identify how you should understand what to do with a given protocol, in a given environment, for a given context?

**PF2:** I will use whatever I can find. Often I am looking at what is done in local language communities. At the moment IANA Language tables.

[JFC.10]

---

**PF1:** Saying just "no mapping at registration time" is not good enough for me. What about if you want to change the holder or adminc or techc of a domain? Should you do mapping then, if the domain name is typed by the user or not? Is a transfer of a domain name "registration time"?

**JFC:** Good point. What if the Registry software does not consider holder/adminc/techc and uses a different system?

**PF2:** They all do. They might use different terminology.

[JFC.11]

**JFC:** What is the Registry only use local script/language interfaces?

Many only support interfaces in local script and local language. I do not see any problem with that.

[JFC.12]

---

**PF1:** No, I rather see that I as a developer can do whatever I want to make things easier for the users, and that might include that I try to do some mappings of codepoints that are not PVALID, or complete changes of some things that might be domain names (hard to know sometimes) to something that are domain names (different separators between labels for example).

**JFC:** and French/Latin language majuscule support?

**PF2:** I can not be more precise than what I was above.

---

**PF1:** So to answer your question Lisa: No, it is not clear what characters are suggested to be mapped, but we do not even know where or who has the responsibility to do the mapping -- if any. In the case of HTTP, is that done in the web browser or on the server side, in the cgi?

**JFC:** I have some difficulty here: how can an IDN be mapped on the server side? Should it not necessarily be mapped on the user side so it can be properly resolved to that very server?

**PF2:** "server" as "server side" in HTTP. Not server side as in DNS server.

---

**PF1:** For example how do a web browser know whether the ajax / json call is to be "a registration" or not? A situation when mapping should explicitly not occur? It can not, and because of that browsers should not include any mapping at all. Ever.

That is one view, and of course you can find the arguments the other way around as well.

**JFC:** I agree with your position. Would you know of an online document/discussion on this matter, or a con/pro list?

**PF2:** I think the IDNA mailing list archive is a good source...

[JFC.13]

---

**PF1:** And this is why I think the current mapping document is as good as we can get it -- and why I think (like Pete) TR46 is too strong. Because it is very wrong to explicitly ask for mappings the way it does.

**JFC:** Do you know if other possibilities have been expressed by W3C, or SDOs?

**PF2:** TR46 is one of the documents. I know ITU and UNESCO also are doing work, but they have been waiting for the RFCs to be published. ISOC might come out with something. ICANN might write recommendations. Etc.

[JFC.14]

---

**PF1:** What *\*might\** be ok and good to do is to say what codepoints can be mapped in various contexts to other characters, but I think for example the definition of bundles in the IANA registries to some degree do that. When not mapping from PVALID to other PVALID of course. So that people know that IF one work in a specific context (certain language for example), THEN it MIGHT be interesting to map from A to B.

**JFC:** Do you consider the possibility of mapping on a TLD basis?

**PF2:** No.

I think the user wants mapping based on other criteria.

[JFC.15]

But of course the TLDs already today via the language tables do implement in some cases bundling and various policy based issues.

[JFC.16]

---

**PF1:** So I would say the mapping document we have is good.

**JFC:** How would you qualify it? a reminder, a protection against too stringent or lax behaviours, a guidance, a reference, a open link to further work? Is its informational nature not a problem while it completes a standard track work?

**PF2:** We do not know yet. Let people try, and we see. Having guidelines is ok, and probably what developers need.

But mandatory rules, nope.

[JFC.17]

Patrik

---

JFC questions not yet addressed by Patrik Falstrom.

---

[JFC.1]

Correct. This is where I identify there are four possible IDNA user localisations:

1. in the user interface to the client
2. in the user client.
3. as a unique IDNApplication that serve the client
4. as a part of an IUI (Internet Use Interface) in between the client and the plug to the network where the server is plugged.

1 & 2 are OK for testing, not for user operations. There are as many experiences and possible different resolutions as there are applications or clients.



3. is more secure, can have a control panel, support local config files. But user applications must know that service application and how to dialog with it.

Only 4 permit localization issues to be locally centrally managed and networkization issues dealt through it with the ISP in order to permit a semantic relation with the server cleared from the context and identical with every existing protocols (the IUI looks to them as a pseudo-network they can ignore). You may have scores of applications using IDN without knowing it if they are UTF transparent.

---

[JFC.2]

We agree. And this is where I say that the IUI must be intelligent enough to be able to make a difference between localization and personalization issues. Also, it means that there might actually be different Ulevels depending on personal, local, etc. mapping (i.e. for example a simple equivalent to Host.txt being supported)

This is why I refer to a Multi-Layer Domain Name System pile with

- what the user actually types,
- which can be different from what the client will accept
- that can be different from the real Ulevel
- and then the actual Alabel

All of them must strictly polynym (synonym in their context) for the user (this may change from users to users). This can be very common if the resolution may also be made outside of the DNS in other environments also using the IDN (IRI, semantic addressing, etc.) This is why I prefer to refer to IDNs as "Internet Domain Names" with a single simple syntax that covers DN and IntlDNs alike. This permits to consider that IntlDNs only are internet domain name of the "xn" presentation. The IUI having the job of sorting presentations and classes.

Presentations are simply the two letters before double dash (IDNA uses the xn-- "extended names" class). Classes other DNSupported pay loads can be introduced by single letter + --.

1. the Internet presentation layer emerges from this, with years of validation behind by i-DNs and IDNA2003 2. I have a universal syntax available to support presentations and any semantic. To make it universal I consider that Internet Domain Names syntax is alphadigital labels and sub-labels with "." (or something else) as intelabel, and "-" or (something else) as intersublabel. Double dash corresponds to an empty sublabel that separates metadata payload from naming payload.

---

[JFC.3]

Agree.

---

[JFC.4]

Yes you can.

I can decide that .fr do not support French majuscules because it never had. And that .fra does, because it is designed to be semantically clear.

---

[JFC.5]

Correct. This is why the architecture needs to be more precise and support the needed flexibility up to the point IDNA2008 defines. This is this way that IDNA2008 defines multiplicity. It is the ultimate required point where architecture must be the same, all over the world, for all users. As such it defines the boarder of the IETF scope. This is why I identify:

- the Internet system, the IETF scope of responsibility : IDNA2008
- the Internet exotem, the user area that wants to use IDNA2008, documenting it is what I call IDNA2010 in order ICANN does not continue to use the word wrongly
- the Internet peritem, the fringe between the Internet system and the user's Internet exotem. Documenting the way this peritem information must be presented is what I call IDNA2012.

As Project.FRA I have a need. I know that to solve it I need to get my users to treat .fra differently from .fr.

How can I do that?

Very simply: in having my users using "punyplus" as an algorithm rather than "punycode". When they find a TLD or an upper level domain (can be several level) ("ULD", also a "User Level Domain") listed in their "netlocale" their UI is going to localize it as indicated in the netlocale file.

The same, if they need something special (for example kidprotection filtering) the punyplus algorithm can be used to filter as documented in a "personale" file. Punyplus and the netlocale file formats and possible management system need to be documented.

My suggestion is that IETF can use:

- MUSTs for its system
  - SHOULDs for the peritem fringe
  - MAY for the user exotem, like Mapping does.
- 

[JFC.6]

Absolutely. This kind of functions (server front end and IUI can be performed by OPES - the HTTP OPES has been specified).

Now, the problem you raise that it is not possible to do it through the DNS protocol is more a question of capacity if there are millions of TLDs to filter, update, etc. This is true on the Internet network side. This is not on a user side.

If instead of a unique unique root file, you consider a unique virtual root file that people can partly implement through their "personale" file. They just maintain the TLDs they are interested in. There are 25.000 sociolinguistic entities: I am able to read or grasp just a tiny few. I wish to filter out what I do not want (and not what Unicode wants, hence the dispute about [RFC 4646](#)).

---

[JFC.7 ]

You cannot. First a whole because some are breakers, phishers, pirates, etc.

This is why I tend to say, what the user feel as the "network" MUST be what his machine perceives as the network (this can include non-Internet networks). In order to achieve this, I consider that "user side network applications" must be considered as forming a dedicated network layer, where they appear as "smart local operating tasks" (slots) to clearly identify the from their brother at networking the Internet network applications (Mail, DNS, etc.).

They locally operate tasks the user consider as parts of his "smart Internet system", whatever it may really be. The user will access them through the browser or some dedicated local interface. Hence I call the slot layer "pseudo-network" as it is seen by the user as an intelligent network shell.

Now, I put that slot layer in the IUI, so there is a clean overlay :

- through the IUI the network expand on the User side.
- through the IUI the user controls the fringe of the network. This permits to differentiate :
  - dump traffic (passive content) from end to end
  - smart traffic (ambient and active content) from IUI to IUI

Then I observe such slots have much in common and that it could be nice for them to interoperate with other local slot servicing them, or with remote fellow slots to support metatraffic. This is why I add interapplication layer - between the Internet network application layer and the IUI network application layer.

Then I observe that there can be advantages in locating network applications in the Internet (global range network applications) and in the IUI (local range network applications).

Let take the example of the DNS. If I put the DNS in the IUI it is both part of the Internet DNS and dedicated to my machine/small-home-network (SnHn) or duplicated within my own Internet context (cybship: all the digital resources I control). To be sure there is no pollution I turn recursion down, making it secure. This means that now I can use my own local root, with just the TLD/ULDs I am interested in from the unique virtual root classes (I can add my own ULDs as China does with its own IDNgTLDs).

---

[JFC.8]

OK. This is where I say there is an urgent need for Internet Adminance (i.e. technically managing the internet in cooperation, so all this can be documented by who is actually responsible). But this calls for flexibility, simple and robust rules, etc. This is why I say the IUI needs an interapplication system that permits network applications. Naturally called it "Netix". And I can easily support it in using the ML-DNS syntax.

This way Netix in the IUI makes a complete interoperable addition to the Internet basis able to match the user's expectation for an intelligent open network of his own understanding, using its own classes and presentations, ML-DNS pile, interslot authentication

---

[JFC.9]

Sure !!! But having a mandatory syntax to express a free mapping semantic may help. Moreover it can be multilingual and users can enter parameters or use scripts.

---

[JFC.10]

Correct.

This is why I need all this and much more to be federated in a structured open repository that applications can directly access. A local customized IANA.

I call it the Multilingual Distributed Referential System (MDRS). It should be ISO 11179 compatible to interoperate with many other registries.

As you say, most is documented on a local language community basis. This is why I documented the ccTAG concept with people from ISO, MAAYA, Linguasphere, etc. They are just a reference grid, a continuation of ISO 3166 (we use to call ISO 3166-4). ISO 3166 includes for each country the administrative (i.e. normative) languages and scripts using ISO 639 and 15924 codes. So ccTAGs proceed in the logical categorization order: country, script, language, region - supported by the core document of international consistency, including Internet and ccTLDs.

The technical interest is that being polylingual (French/English) and listing the Govs, i.e. the standardization authorities, it can be understood as the French and English instantiations of a single multilingual standard with equal validity in every normative language (what we call "polynym", synonym in different context).

Due to the universal acceptance of ISO 3166 it continues the Internet infrastructure as it is since 1978. Very stable, accepted, intuitive. This means that I can tag every application linguistic instantiation in a congruent manner with its administrative forms, sales contracts, vendor rates and support, local regulations, etc.

Also, that through ISO 11179 compatibility the MDRS can be indefinitely expanded as a metastructure, ready to support the Intersem (semiotic/semantic internet) needs.

---

[JFC.11]

All ccTLDs do not adhere to the ICANN WHOIS concept which actually is illegal but tolerated in most of the countries (violation with local privacy laws). Zone Managers can be at any level and be sure that many ULD managers will not even know that ICANN exists or what it could have to do with their Private Class TLD. ICANN is only for class IN.

---

[JFC.12]

Sorry, I meant at application layer, not at user layer.

Internationalization means that local language is embedded in English. Multilingualization means that local language replaces English.

This is the conflict I had with Unicode because their Globalization is internationalization+localization+flangtag-filtering. In the fussy process we identified the existence of multilinguistics as the discipline of the cybernetics of the linguistic diversity. How languages interrelates among other languages and/or cultural entities.

---

[JFC.13]

OK

But I thought about something more formal and organised.

I compiled all the WG/LC remarks in a Draft. Indicating which had been answered. I lost the count of the non-answered ones.

<http://www.ietf.org/id/draft-iucg-wgidnabslc-01.txt> (or better reading:  
[http://wikidna.org/index.php?title=WG-IDNA\\_LC](http://wikidna.org/index.php?title=WG-IDNA_LC))

---

[JFC.14]

Dont you think a common work could be usefull ? Users are not interested in consensus, but in multiconsensus, i.e. a consensus in each of the position groups, and a general consensus on how each position bridges with the others. Beging in using a single terminology and possible comparison tables.

---

[JFC.15]

For example Project.FRA wants majuscule support based on .fra while .fr may npt mind or address the need differently.

---

[JFC.16]

I am interested in what I call mecalanguages. i.e. natural languages as to be spoken by machines. The risk is that they become the referents and that languages rigidifies or split (machine/human). But the risk is, now on an interactive basis, equivalent to dictionnaires and grammars. The mecalanguages must be able to adapt and support dolwn to avalects (languages of each individual avatar's styles) (hundreds of billions). The only serious non-physical personal authentication is based on avalects as it betrays the cerebral status.

Now, the challenge is to work on the invariants, i.e. the semantic traits or atoms, that can then be passed as metadata (as one passes Alabels) and restored through different mecalanguages. This gives the idea of Ulabels that could display differently depending on the context.

This is certainly possible as part of the ML-DNS (multilayer domain names pile system - the minimum of it being the Alabel/Ulabel pile. Domain Names cannot refer to invariant, but semantic addressing can do it while using the DNS as a support. This is one of the targets of the Project.FRA in using its namespace as a taxonomy for an open ontology.

---

[JFC.17]

Who in your opinion should work these guidleines after the experience of this WG, i.e. working relations with Unicode, IUCG, i-DNs, ICANN, ITU, etc. ?

We have a MAAYA Unesco meeting on 23/24 to prepare a book on multilingualism in cyberspace. I think we should start with some kind of ontology of multilingualism and cyberlinguism, i.e. the way languages behave together and the way cyberspace affects languages. But we will know better when we know what others may mean when we use the same words as us.

This is why I say that now we have built the Internet side IDNA, one has to consider the User side and the infrastructural side (guidelines, tables, information, registries, etc. ): where and how the developer information should be stored and found.

## Annex 9.

### AD and Chair point of view

#### Contents

[\[hide\]](#)

- [1 At 21:33 01/12/2009, Lisa Dusseault wrote](#)
- [2 At 22:04 01/12/2009, Lisa Dusseault](#)
- [3 At 16:06 02/12/2009, Lisa Dusseault wrote](#)
- [4 At 20:35 02/12/2009, Lisa Dusseault wrote](#)
- [5 At 00:15 03/12/2009, Lisa Dusseault wrote](#)
- [6 At 01:46 03/12/2009, Lisa Dusseault wrote](#)
- [7 At 05:01 03/12/2009, Lisa Dusseault wrote](#)
- [8 At 23:41 04/12/2009, Lisa Dusseault wrote](#)
- [9 The WG/IDNABIS Chair wrote on 2009/12/29](#)
- [10 france@large comments](#)

#### At 21:33 01/12/2009, Lisa Dusseault wrote

One example I discussed with Patrik yesterday, was whether locale might affect mapping. I'd like to get better insight into the general understanding of that.

1. Could locale determine whether a PVALID character should be mapped into another PVALID character prior to following the rules to turn into an ALABEL? I believe the consensus answer is probably SHOULD NOT or MUST NOT because that would make domains with that valid character unreachable by software using those locale rules.

2. Could locale determine whether, or how, a DISALLOWED character is mapped into a PVALID character prior to getting an ALABEL? For example:

- in locale Laputa, disallowed character  $\square$  (x230) is not used in the local language, so it's not mapped, an error occurs so a user seeing that in a Web page can't reach any actual domain
- in locale Balnibarbi,  $\square$  (x230) is considered to be the same as O, so it's mapped to 'o', and a Balnibarbi user reaches domains containing o's
- in locale Glubbubdrig,  $\square$  (x230) is considered to be the capitalized version of  $\bar{O}$  (x231) and a Glubbubdrig user reaches a different set of domains containing  $\bar{O}$ 's



Note that none of these locale rules would necessarily make domains containing  $\bar{O}$  completely unreachable to their users -- a Web page containing a link with  $\bar{O}$  would be looked up without mapping that character to another (assuming the conclusion of point 1 above).

If I'm reading between the lines correctly, communication is hampered between people who are writing under the assumption that this kind of locale-dependent scenario is going to happen, and people who are writing under the assumption that this kind of locale-dependent scenario ought to be forbidden and nobody would try such a crazy thing.

## **At 22:04 01/12/2009, Lisa Dusseault**

I don't believe we know what the WG consensus position is around how strongly pre-lookup mappings are recommended and in what use cases, and how compatible optional pre-lookup mappings are with IDNA2003 in-protocol mapping.

Agreeing with you to a large extent Vint, I believe there is a strong consensus in the WG that mappings "aren't part of the protocol", meaning mapping shouldn't be required in all cases. Certainly, registries shouldn't be required to automatically map any character into any other character before registering the mapped domain name: better to make the domain name holder be explicit about the exact domain name they are registering. We have also referred to use cases involving client software doing lookup of domains that are already supposed to be valid, in which context an invalid character would simply be a software error rather than a situation that required mapping. In this matter, we have consensus to do something different in IDNA2008 vs IDNA2003.

I believe we also do have consensus in the equivalence of A-label and U-label forms and transformation in either direction.

After that it gets confusingly nuanced. We even discussed having a different term for optional, pre-lookup mapping of user-typed-in domain names, including links found in HTML typed in by the HTML author. I believe we do have consensus in the WG that this type of helpful mapping is going to be implemented, for example in Web browsers, as being a user experience that is preferred over a dialog interruption or error.

So that leaves us with a great deal of progress in IDNA2008, and a lot that is decided by WG consensus, but still with a gap in our total consensus. Some people would prefer that optional, pre-lookup mapping be wide open. Others would prefer would that although pre-lookup mapping should be optional, IF it is done, it MUST be done one way. There's even room for a middle ground where there might be more than one recommended standard set of mappings, but they should be done as a matter of standardization and not as a matter of full implementer discretion, if at all possible. There is room for splitting the mappings territory up per-application: perhaps domains in HTTP URLs and browser address bars should be mapped one way, but other types of URLs could follow "no mappings allowed" rules.

Finally, we have the options around transition strategies (such as bundling recommendations) that might allow us to come to consensus regarding these optional pre-lookup mapping questions.

I appreciate all the discussion that is leading us to a better understanding of the issues around those last issues, and developing consensus regarding those last options.

## **At 16:06 02/12/2009, Lisa Dusseault wrote**

I'd like to try to unpack some of the different use cases we're talking about a little more.

ISTM that use cases where the person following the link is the person who is typing it in, are use cases that locale-dependent mapping might be most useful. If I'm in a locale where `Ö` (x230) is considered to be the capitalized version of `o` (ASCII `o`), it might very well be most helpful to make that mapping. Use cases where the same user is typing in the domain names that then looks them up include:

- typing links in the address bar
- typing mail address in the To field of an email
- Writing a Web page, blog post or email, wherein I check that the links work before posting/sending my document

In contrast, the use case where the person looking up the domain `FÖÖ.example` is not the person who typed it in, then in most cases we no longer know the intent or locale of the person who typed in the domain. It may be the same locale as the person who is looking up the domain but it may not be. The person who typed in the domain may have intended `föö.example` or `foo.example`, and may have tested that before sending/posting the link, but we no longer have that information. Use cases include:

- Following a HTTP link in any Web page, document, blog post, email, etc
- Using a `mailto` link (explicit or implicit), e.g. when one person sends me another person's email address

We probably would all agree that people follow links while Web browsing far more often than they type them in, and even when typing in, auto-complete probably drastically reduces the new cases of from-scratch mapping and lookup.

However, we probably have quite different assumptions about how much Internet activity takes place among users of a consistent locale. Can we assume that Patrik wants `Ö` interpreted as `o` because he communicates mostly in Swedish with Swedish users and mostly reads Swedish Web pages? Or must we assume that Patrik also gets email from German and Swiss senders, and also reads Web pages (perhaps in English!) written by German users who expected different mappings? How do we vary on our model of a user, and whether we're using ourselves as hypothetical examples or not.

One slightly more solid question for browsers is, would it be entirely crazy to have different mapping algorithms for typed-in domain names than for links followed? There might be a locale-dependent mapping as well as a global mapping. (I assume that having every established locale mapping installed would be complete craziness.)

Another question is: when posted links are followed, how often do we know the locale where the link was authored? Not that the browser following the link would necessarily be able to apply the mappings of the locale in which it was authored, but would it be slightly better to apply a global mapping than a mapping from a different locale?

Do any authoring software clients fix up links as the user types? When I type a link in a document, the authoring software often makes that link active. Is there any software that automatically lower-cases? If so, would such software also be likely to map to PVALID characters before the doc is finished?

### **At 20:35 02/12/2009, Lisa Dusseault wrote**

I don't see where that principle begins and ends. If it were entirely true, zone operators would be entirely free to register xn--garbage as a domain, entirely free to register DISALLOWED characters.

Since it is in the purview of IDNA designers to make a character valid or disallowed, wouldn't also be in the purview of IDNA designers to say something like "This character becomes valid on date 01/01/2020", right? So why wouldn't it be in the purview of IDNA designers to also say something like "This character becomes fully valid on date 01/01/2020, but until then, may be registered if it is bundled with another sequence"?

Doncha love slippery slopes

### **At 00:15 03/12/2009, Lisa Dusseault wrote**

*Would someone who understands IETF process better than I do please explain why the discussion of that character needs to proceed any further?*

Even if this is a rhetorical question, I'll bite. It's because the IETF makes decisions by rough consensus and running code. Rough consensus is among informed participants as well as experts and people in certain positions of authority or responsibility. Running code certainly brings in browser/client implementation history and current client implementation concerns. It is not only operators of the countries where those languages are most spoken, that have collateral effects from the status of the characters of those languages in IDNA.

### **At 01:46 03/12/2009, Lisa Dusseault wrote**

Your argument is a fair argument to try to influence the consensus, but it is not a fair argument to overrule a consensus (not that I'm saying you are doing that). We do rely on most IETF participants being constructive most of the time. We just have no other process besides appealing to people who are forming the consensus to look at running code, reality, security concerns and so on.

### **At 05:01 03/12/2009, Lisa Dusseault wrote**

I did claim that running code was an important influence on a WG consensus, by tradition, habit and official encouragement.

- First, that doesn't mean it's the only influence, or the most important influence in some considerations.
- Second, the term is used both strictly, as in to demonstrate interoperability directly, and also loosely. By loose interpretations of "running code" you could claim that how a country uses a character in printed books was a form of deployed code that can be used as factual evidence.

### **At 23:41 04/12/2009, Lisa Dusseault wrote**

I've been having trouble understanding how TRANSITIONAL characters that could not be mapped or looked up by clients would help us out here. (You're not the only one Erik, I was just thinking of this offline)

If I understand correctly, Web browsers in particular visit a corpus of Web pages with links that they would like to not only continue to support, but also continue to resolve to the same host (absent other changes). So on the day that somebody upgrades to the new version of a browser with IDNA2008, their link destinations for TRANSITIONAL characters do not change.

"Do not change" also means that the user doesn't suddenly start getting an error trying to follow a link with a TRANSITIONAL character in the domain.

I agree that in some models, an error is better than going to an indeterminate destination. But only in some models. To the user, upgrading their browser and suddenly having links with `␣` in domains fail where it succeeded the day before, does not seem like a real upgrade.

### **The WG/IDNABIS Chair wrote on 2009/12/29**

It has been suggested that a better forum in which to deal with IDNA2003 and IDNA2008 incompatibility is the ICANN IDN Guidelines Committee. That may be a better forum with broader participation than the IDNABIS working group in which the TR46 proposal or other

proposals may be discussed. If we adopt Cary Karp's offer, your observations, below, would be input into the Guidelines committee discussions.

For years, none of the france@large participants to the writing of the following note has succeeded to participate to the ICANN process on an individual basis or through an ICANN constituency.

## france@large comments

Dear Lisa,

It took a few days for france@large to consider all your mails in attachment. We found that they raise three fundamental problems that have not been discussed yet but that now needs to be addressed prior to any publication of the IDNA2008 documentation set.

These problems are the network neutrality, security, and operative independence that are to be enhanced or at least respected when introducing IDNA2008. However, IDNA2008 is Internet neutral; it can be tested, introduced, and supported in a non-neutral manner. IDNA2008 is both a protocol and architecture. The protocol security aspects are covered in the current security section, but the architectural security aspects are not discussed.

Moreover:

- IDNA2008 represents a substantial technical progress of the Unicode globalization towards a Multilingual Internet (represented for example by an independence from a Unicode specific version and reduced mapping at protocol level). This means a better support of specific namespaces that currently are, can be operated, or are to be built, outside of the area of ICANN's governance (as prepared by the ICANN ICP-3 document, france@large has applied through a two-year community test). This in particular includes non-IDNA presentations, non class "IN" classes.
- [RFC 3935](#) assigns the IETF with the responsibility for a much broader Internet than the ICANN footprint, i.e. "A large, heterogeneous collection of interconnected systems that can be used for communication of many different types between any interested parties connected to it. The term includes both the "core Internet" (ISP networks) and "edge Internet"(corporate and private networks, often connected via firewalls, NAT boxes, application layer gateways and similar devices). The Internet is a truly global network, reaching into just about every country in the world. The IETF community wants the Internet to succeed because we believe that the existence of the Internet, and its influence on economics, communication, and education, will help us to build a better human society."

We do not have the time, technical capacity or experimental background now to tackle this issue at an IETF standardization level. However, the IETF MUST make it clear that if the resolution of IDNA operational and architectural issues is delayed and possibly delegated to ICANN's (\*) IDN Guidelines Committee (cf. WG Chair 19

December 2009 mail) or is assumed by the IUCG open special interest group for an IDNA2010 BCP on the implementation, transition, and deployment of IDNA2008 (<http://idna2010.org>), some principles must be clearly and consistently spelled and be respected by these entities because they belong to the IETF rights, duties, and values. The first of them is that these entities must be open to any interested individual.

- The WG/IDNABIS Charter states: "It is recognized that some explicit exceptions may be necessary in any case, but attempts will be made to minimize these exceptions." At the WG/IDNABIS, we considered the technical exceptions within the protocol definitions and minimized them. However, there are other exceptions that may arise in the protocol experimentation, deployment, and various applications that can lead (as you have yourself noted) to conflicts on the same user machine, within the same usage, the same application or the same network governance area. Our mission is to document how to avoid such technical, security and operational problems or to list them in the security section with the principles that should be followed in order to best protect Internet Users.
- The WG/IDNABIS Charter also states: "The constraints of the original IDN WG still apply to IDNABIS, namely to avoid disturbing the current use and operation of the domain name system". This is true at every level of the DNS: this is also true for the first level. This means that this WG must ensure that its document set and its consequences are not possibly used to disturb the current use and operation of the domain name system at architectural, security, and operational levels. This is not the case. The documents should at least include the mention of these risks, and how they should be defeated or that further work on the issue is to be carried.

The IETF is not involved with the "current use and operation of the domain name system", except for one major exception. This exception is the management of the IANA registry. This is described in [RFC 2860](#) and [RFC 5226](#) (which obsoletes [RFC 2424](#)). There are two main cases involving the IETF's direct responsibility to the point where it can exercise its ability to cancel the IETF/ICANN MOU:

- in case of experimentation by ICANN and/or IETF
- in case of a conflict between the standard and the way ICANN applies an IETF standard.

There are non-ICANN entities which may de facto or legally share or assume the IANA's role to the legitimate benefit of some IDN internet communities; they will be called IDNO (IDN Organizations).

There are expressed concerns and demonstrated situations where ICANN, or other IDNO, that the current use and operation of the domain name system might be disturbed. There are three classes of (possible) situations where this happens:

- the lack of distribution equilibrium between the kind of potential difficulties to experiment, analyze, and address that is due to the limitations imposed by ICANN to participate in its experimentation which is restricted to non-roman IDNccTLDs. It is to be expected that most of the new difficulties to experiment will be located
  - at IDNgTLDs along with more innovative uses and the lack of a governmental authority to sustain their naming policy,
  - at bidi TLDs,
  - and at roman TLD where the orthotypography is not fully supported (as Latin language TLDs, and especially French).
- These TLDs MUST either be a part of the ICANN experimentation on an equal footing basis, or an IETF sponsored complementary experimentation should be organized by adding a batch of experimental TLDs in the root file towards to warranty an equal chance of access to experimental operations and sales, on a first come first serve basis.
- the commercial discrimination that is imposed by ICANN's internal rules between candidates to IDNg/ccTLD on IDNA and root capacity related technical grounds (after two years of an official public campaign denying such limitations and an offer to register gTLD candidates (with 500 candidates being persistently rumored)). This is in clear contradiction with the WTO rules and the international agreements concerning the Technical Barriers to Trade. Due to the [RFC 5226](#) (2424) on the "first come, first serve" principle in IANA registration and the [RFC 2860](#), the IETF could find itself involved in a lengthy international commerce case that could affect the credibility of its international responsibility capacity and further delay IDNs for years to come.
- new Internet architectural readings, such as "Interplus", which was introduced by the IUCG, based upon IDNA2008 findings. This WG should add in the security section that this potential risk has been identified, the IUCG was requested to document "Interplus like" architectural frameworks, and it will rest with ICANN, IETF, and the appropriate IDNO and Internet User representations to address it.

[RFC 3935](#) assigns the IETF with the goal to make the Internet work better:

"The mission of the IETF is to produce high quality, relevant technical and engineering documents that influence the way people design, use, and manage the Internet in such a way as to make the Internet work better. These documents include protocol standards, best current practices, and informational documents of various kinds".

The standards are well defined in [RFC 3935](#):

"As used here, the term describes a specification of a protocol, system behaviour or procedure that has a unique identifier, and where the IETF has agreed that "if you want to do this thing, this is the description of how to do it". It does not imply any attempt by the IETF to mandate its use, or any attempt to police its usage - only that "if you say that you are doing this according to this standard, do it this way". The benefit of a standard to the

Internet is in interoperability - that multiple products implementing a standard are able to work together in order to deliver valuable functions to the Internet's users."

As a result, this WG should make it clear that any derivative work where additional neutrality, security, or operational additional constraints that is based upon privately contracted situations that would not fully respect "its description of how to do it", including its principles for derivative work, might not be interoperable and, therefore, should not be attempted by ICANN and IDNA having entered into a comparable MoU with the IETF. This restriction is a consequence of the IETF [RFC 2860](#) obligations to properly advise ICANN.

[RFC 3935](#) also states:

"The Internet isn't value-neutral, and neither is the IETF. We want the Internet to be useful for communities that share our commitment to openness and fairness. We embrace technical concepts such as decentralized control, edge-user empowerment and sharing of resources, because those concepts resonate with the core values of the IETF community. These concepts have little to do with the technology that's possible, and much to do with the technology that we choose to create."

What is discussed in this very memo shows that IDNA2008 may lead to important, yet potentially disruptive, innovations that the IETF, ICANN, IDNOs and Internet Users should collectively evaluate and canalize. IDNA2008 permits misuses that they all should oppose in unison.



## Annex 10.

### IAB recommendations and considerations

Taking into account the issues above, it would seem inappropriate for an application to convert a name to Punycode when it does not know whether DNS will be used by the name resolution library, or whether the name exists in a private name space that uses UTF-8, or in the global DNS that uses Punycode.

Instead, conversion to Punycode, UTF-8, or whatever other encoding, should be done only by an entity that knows which protocol will be used (e.g., the DNS resolver, or `getaddrinfo` upon deciding to pass the name to DNS), rather than by general applications that call protocol-independent name resolution APIs. Similarly, even when DNS is used, the conversion to Punycode should be done only by an entity that knows which name space will be used.

That is, a more intelligent DNS resolver would be more liberal in what it would accept from an application and be able to query for both a Punycode name (e.g., over the Internet) and a UTF-8 name (e.g., over a corporate network with a private name space) in case the server only recognized one. However, we might also take into account that the various resolution behaviors discussed earlier could also occur with record updates (e.g., with Dynamic Update [RFC2136]), resulting in some names being registered in a local network's private name space by applications doing Punycode conversion, and other names being registered using UTF-8. Hence a name might have to be queried with both encodings to be sure to succeed without changes to DNS servers.

Similarly, a more intelligent stub resolver would also be more liberal in what it would accept from a response as the value of a record (e.g., PTR) in that it would accept either UTF-8 or Punycode and convert them to whatever encoding is used by the application APIs to return strings to applications.

Indeed the choice of conversion within the resolver libraries is consistent with the quote from [RFC3490] section 6.2 stating that Punycode conversion "might be performed inside these new versions of the resolver libraries".

That said, some application-layer protocols may be defined to use Punycode rather than UTF-8 as recommended by [RFC2277]. In this case, an application may receive a Punycode name and want to pass it to name resolution APIs. Again the recommendation is that a resolver library be more liberal in what it would accept from an application would mean that such a name would be accepted and re-encoded as needed, rather than requiring the application to do so.

Finally, the question remains about what a DNS server should do to handle cases where some existing applications or hosts do Punycode queries within the local network using a private name space, and other existing applications or hosts send UTF-8 queries. It is undesirable to store different records for different encodings of the same name, since this

introduces the possibility for inconsistency between them. Instead, a new DNS server could treat encoding- conversion in the same way as case-insensitive comparison which a DNS server is already required to do. Two encodings are, in this sense, two representations of the same name, just as two case-different strings are. However, whereas case comparison of non-ASCII characters is complicated by ambiguities (see [RFC4690]), encoding conversion between Punycode and UTF-8 is unambiguous.

## **Normalization/mapping issues**

[EDITOR'S NOTE: There are also normalization/mapping issues which the next version of this document may explore. Currently we only explore encoding issues.]

## **Security Considerations**

Having applications convert names to Punycode before calling name resolution can result in security vulnerabilities. If the name is resolved by protocols or in zones for which records are registered using other encoding schemes, an attacker can claim the Punycode version of the same name and hence trick the victim into accessing a different destination. This can be done for any non-ASCII name, even when there is no possible confusion due to case, language, or other issues. Other types of confusion beyond those resulting simply from the choice of encoding scheme are discussed in [RFC4690].

# Annex 11.

## IDNApplication Medley

This is just a documented list of topics that I believe related to the IDNA2010 work to be done to acclimate the IDNA2008 tool-kit with users.

### Contents

- [1 ICANN position regarding IDNA2008 implementations](#)
- [2 Microsoft IRT the French language orthotypography](#)
- [3 Developer information](#)
  - [3.1 091104 14:31 Erik van der Poel](#)
  - [3.2 091104 15:38 Gihan Dias](#)
  - [3.3 091104 15:47 Andrew Sullivan](#)
  - [3.4 091104 15:59 Vint Cerf](#)
  - [3.5 091104 16:16 Gihan Dias](#)
  - [3.6 091123 16:23 Erik van der Poel](#)
  - [3.7 091104 16:32 Andrew Sullivan <ajs@shinkuro.com>](#)
  - [3.8 091104 16:37 John C Klensin](#)
  - [3.9 091104 16:48 Andrew Sullivan](#)
- [4 Cross/inter label testing](#)
  - [4.1 091104 11:11 Vint Cerf](#)
- [5 Unicode TR46](#)
  - [5.1 20091027 Martin J. Dürst](#)
  - [5.2 20091226 MARK DAVIS](#)
- [6 "eszett" exception](#)
  - [6.1 Maryhofer - NIC.AT 20091025](#)
  - [6.2 Mark Davis, Unicode, 20101026](#)
- [7 Vint Cerf letter to Unicode](#)
  - [7.1 Letter to Unicode](#)
  - [7.2 Unicode response](#)
  - [7.3 Mark Davis, Unicode President](#)
    - [7.3.1 Problem](#)
    - [7.3.2 Proposal](#)
      - [7.3.2.1 The requirements are:](#)
      - [7.3.2.2 The goal is to](#)
    - [7.3.3 Scenarios](#)
  - [7.4 Acknowledgment of the WG/IDNABIS Chair](#)
- [8 Pete Resnick's Recipe](#)
- [9 IDNA 2008 in light of DYN DNS](#)
  - [9.1 John Klensin 200910025](#)
  - [9.2 Bernard Adoba 20091026](#)
  - [9.3 Vint Cerf 20091026](#)
- [10 RTL & Labels](#)

- [10.1 Abdulrahman I. ALGhadir 20091013](#)
- [10.2 Vint Cerf 20091013](#)
- [10.3 Abdulrahman I. ALGhadir 20091014](#)
- [11 Full stop Mapping](#)
  - [11.1 20091011 Sarmad Hussain](#)
  - [11.2 20091011 Vint Cerf](#)
  - [11.3 200910111703 Erik van der Poel](#)
  - [11.4 20091011 At 17:24 int Cerf](#)
  - [11.5 20091011 At 18:52 11/10/2009, Paul Hoffman](#)
  - [11.6 20091012 At 14:53 John C Klensin](#)
  - [11.7 20091012 At 15:22 Paul Hoffman](#)
  - [11.8 20091012 Shawn Steele](#)
  - [11.9 At 17:06 12/10/2009, Paul Hoffman wrote:](#)
  - [11.10 At 17:16 12/10/2009, Vint Cerf wrote:](#)
- [12 Bidi discussion](#)
  - [12.1 From: jefsey \[1\] Sent: 13/Feb/2010 4:15 PM](#)
  - [12.2 Slim Amamou <slim@alixsys.com> 14 février 2010 04:57](#)
  - [12.3 Abdulrahman I. ALGhadir <aghadir@citc.gov.sa> 14 février 2010 05:39](#)
  - [12.4 Slim Amamou <slim@alixsys.com> 14 février 2010 11:23](#)
  - [12.5 Vint Cerf <vint@google.com> 14 février 2010 15:36](#)
  - [12.6 Slim Amamou <slim@alixsys.com> 14 février 2010 20:41](#)
  - [12.7 Vint Cerf <vint@google.com> 14 février 2010 21:03](#)
  - [12.8 Slim Amamou <slim@alixsys.com> 14 février 2010 22:50](#)
  - [12.9 Slim Amamou <slim@alixsys.com> 16 février 2010 10:21](#)
  - [12.10 Jefsey](#)

## ICANN position regarding IDNA2008 implementations

As far as IDNA2008 deployment and transition from IDNA2003, ICANN represents a potential of a few zone Management contracts (project FAST TRACK) over the billions of concerned existing and future DNS zone managers.

The work they achieve in their particular should be part of the IDNA2010 BCP on IDNA2008 deployment, transition and usage.

- The [ICANN implementation guidelines](#) reference page was last updated by a 2007 version which did not cover IDNA2008 issues.
- The [ICANN IDNA technical considerations](#) are based upon IDNA203 RFCs, [RFC 4690](#) calling for its review. It does not quote the existence of the IETF WG/IDNABIS, not its [Charter page](#), nor its deliverables, however they now are sent to the IESG.

**On 15 Dec 2009, Cary Karp indicated :**

"The reason that text has been unchanged for so long is that the authorship group has been waiting for the IDNA protocol revision to be concluded so its details can be reflected in the next version of the Guidelines."

"ICANN had been similarly holding off on finalizing the terms of the ccTLD Fast Track <<http://www.icann.org/en/topics/idn/fast-track/>> to ensure its conformity with the revised protocol. That goal was abandoned a while back and a reference group was appointed to advise about key pending implementation issues. The report produced by that group <<http://www.icann.org/en/announcements/announcement-2-03dec09-en.htm>> will likely be folded into the Guidelines to render them more applicable to the TLD space, without or without additional benefit of being able to address the outcome of the protocol revision. The report considers the matter of "character variant management" which should also be of relevance to the IDNA-UPDATE discussion of bundling.

**On 17 Dec 2009, Kary kindly gave the following indications:**

- In this endeavor, I am one of a group of TLD administrators who meet as peers and have volunteered to be their scribe. (Only two of us regard ourselves as native Anglophones and the other guy keeps forgetting to bring a pencil.)
- The group has been active since IDNA2003 was finalized, and its mandated concern has been to provide a basis for the responsible implementation of that protocol in the SLD space. Beyond that, we make every endeavor to formulate the Guidelines in a manner that will have commonsense appeal to the operators of all zones on all levels of the DNS when formulating their own IDN policies. We also make every effort for the Guidelines to reflect the best wisdom about IDN generated in other fora.
- Our work has been in abeyance for an uncomfortably long time while waiting for IDNA2008, and my proposal was intended to hasten progress on both fronts. The transition between IDNA2203 and IDNA2008 is a glaringly obviously concern for the TLDs that have been accepting registration under the former. Since there are no IDN-labeled TLDs, those that will be established are unlikely to have legacy difficulties in this regard.
- The next version of the Guidelines will of necessity contain transition recommendations. It therefore seems rather obvious for them to reflect the massive amount of work done by the IETF WG and the organizations that have participated in it. If the IETF nonetheless opts to establish a separate instrument in which corresponding recommendations are expressed, the gTLDs -- which are contractually bound to implementing the ICANN Guidelines -- can easily end up in the awkward position of needing to reject anything proposed by the IETF (or in any other context) that is at odds with the ICANN Guidelines.

# Microsoft IRT the French language orthotypography

Wed, 09 Dec 2009 14:05:51 -0800:

When asked to comment a WG/IDNABIS Members 25 vs. 4 feed back that Unicode/Google wants to disregard, Microsoft's representative, [\[Shawn Steele\]](#) wrote (Message-ID: <E14011F8737B524BB564B05FF748464A0446592E@TK5EX14MBXC139.redmond.corp.microsoft.com> In-Reply-To: <D3E448A8-20D2-4CFB-ACE0-8A165B297FC2@google.com> :

"A couple things concern me about the results here.

One is that, of course, "france@large" supports PVALID as it is closer to their case of separating Ecole and ecole. **To me, this is noise.**

The other votes I could almost group into 3 buckets:

- Those that think *names need to be distinct* and don't *like* mapping, or bundling, either.
- Those that think there's a linguistic difference and support is necessary, yet like mapping. This includes people who would bundle, so it's not clear to me if PVALID is voted to support the character, or because differentiation is necessary. *Bundling seems to be at odds with the need for differentiation.*
- Those who think the proper presentation is needed, *yet think* that some sort of enforced bundling, or mapping, is not harmful, and is important for compatibility.

These groups seem to also have different perspectives, and areas of expertise (**all valid**), so it concerns me that strictly counting numbers might ignore one perspective or area of expertise."

The main idea seems to be that a **25 vs. 4** feed-back addressing the French language requirements should actually be read as a **nil to 3** expertise opposing it.

---

NB: **france@large** does not take "Ecole" and "ecole" as a paradigm. **Project.FRA** uses to express the difficulty in explaining that two **.fra** domain names of which the orthotypography would clearly differentiate the meaning would however either be read as "State" or as "status". This would represent a semantically non-acceptable loss of meaning. This would therefore prevent the Internet DNS to be interoperable with the generalized intertechnology **Semantic Address System (SAS)**, that **Project.FRA** and other Internet multilinguistic projects are jointly exploring.

## Developer information

## [\[edit\]](#) 091104 14:31 Erik van der Poel

There are several different operations that you can perform on the labels of a domain name, and these operations occur at different times. Here are just a few examples:

- (1) single-label registration time
- (2) multi-label DNAME definition time
- (3) multi-label domain name lookup time
- (4) multi-label domain name display time

idnabis-protocol-17 focuses on (1) and (3). For (1), it says:

"If the proposed label contains any characters that are written from right to left it **MUST** meet the BIDI criteria [IDNA2008-BIDI]."

Note that the above is talking about a single label. For (3), the protocol says:

"Verification that the string is compliant with the requirements for right to left characters, specified in [IDNA2008-BIDI]."

Note that the above is talking about a "string", which presumably might contain more than one label. As far as I can tell, IDNAbis does not say much about operations (2) and (4) for bidi.

idnabis-bidi-06 says:

'A "BIDI domain name" is a domain name that contains at least one RTL label.' and

'The following rule, consisting of six conditions, applies to labels in BIDI domain names.'

Clearly, the above is talking about multi-label domain names. However, the rule itself tells you how to test a single label, so that part of the spec can be used at registration time (1).

Let's take an example. One of our favorite examples is 3com.com. At registration time, when we are registering the label "3com", there is no way of knowing that someone may, at some point in the future, define a DNAME that breaks the IDNAbis bidi rules. Since there is no way of knowing that, the registration is simply allowed.

Later, someone tries to define a DNAME, say, HEBREW.3com.com where HEBREW is a string of right-to-left Hebrew characters. At this point, the implementation might choose to check the IDNAbis bidi rules and either reject the DNAME or emit a warning about it if it breaks the rules.

Even later, someone tries to lookup HEBREW.3com.com. The implementation can check the entire domain name against the IDNAbis bidi rules. It does not have to check since the protocol says "SHOULD".

Yet later again, someone tries to display HEBREW.3com.com. The implementation probably should check against the IDNabis bidi rules. If the domain name breaks the rules, the implementation can refuse to display it in Unicode form (choosing Punycode instead), or produce a warning of some kind.

So, the IDNabis drafts are in some sense incomplete, since they don't fully address DNAME time (2) and display time (4). But if the WG does start to discuss these, you can imagine what my position is going to be.

Erik

### **091104 15:38 Gihan Dias**

*the WG concluded that the registry or registrar had to be cognizant of this kind of anomaly and reject problematic registration requests.*

Unfortunately, it is impossible to make sure *\*all\** registrars (or even all registries) would uniformly handle problematic requests.

We need mandatory rules for registrations, which are checked by software in lookups.

### **091104 15:47 Andrew Sullivan**

*(2) multi-label DNAME definition time*

The problem there is that there is *\_no way\_* to have a rule about this. It's not that we never thought about it; it's that you can't possibly specify a rule that will solve this problem.

If there are zone cuts, it is *\_just impossible\_* to know what might end up lurking on the other side of the cut, not only at the time of registration but also later, since a DNAME could be introduced at a later date (indeed, that's actually one of the target use cases of DNAMEs).

### **091104 15:59 Vint Cerf**

There is no way to guarantee this uniformity; nor was the WG able to define every possible case in which problems would arise. The range of possibilities is too great given the huge number of new symbols introduced by IDNA. This was debated extensively during the last 2 years and that is the working group consensus as I see it.

### **901104 16:16 Gihan Dias**

I understand.

My point is that if the WG was unable to reach a consensus, then registrars would not be able to do so either, so what we'll have is an "anything goes" situation.



At the least, we (software developers, sys admins, website owners, advertising companies, users, etc.) need to be aware of the problems.

**091123 16:23 Erik van der Poel**

I was talking about a single DNAME definition, not about other DNAMEs that might enter the picture at some point. If somebody were to implement a tool that checks DNAMEs before inserting them into a zone, that tool could check a \*single\* DNAME against the IDNabis bidi rules.

Here is an example from <http://tools.ietf.org/html/rfc2672>

frobozz.example. DNAME frobozz-division.acme.example.

When the tool is about to insert this DNAME into the zone, it can apply the bidi rules to the full domain name (which consists of multiple labels, as you can see).

Even if some other DNAMEs were defined later that would allow the creation of domain names that break the bidi rules, the client can check the rules at lookup time and at display time.

**091104 16:32 Andrew Sullivan <ajs@shinkuro.com>**

*My point is that if the WG was unable to reach a consensus, then registrars would not be able to do so either, so what we'll have is an "anything goes" situation.*

Yes, but not by registrars. Registry operators need to make policy that makes sense given the characters they're willing to support. Note that the documents explicitly say, "Registries need to develop policy in this area." Will this vary from registry to registry? Yes: that's part of the goal. The idea is to make the protocol flexible enough that different operators can make different policies to match their different circumstances.

*At the least, we (software developers, sys admins, website owners, advertising companies, users, etc.) need to be aware of the problems.*

You bet. But how is that different from being aware of the linguistic environment in which your software runs, your systems work, or your advertising and websites are interpreted? For instance, if you're aiming to attract fans of 1970s Detroit automobiles, you'd be pretty foolish to name your site "underthebonnet.com". How is this different?

**091104 16:37 John C Klensin**

This would create more vulnerabilities than it would eliminate. The argument for "more run time checking" involves not trusting zone administrators. I think we have found a decent balance on that subject -- not going quite as far as we possibly could, but striking a balance with the other considerations, including different practices by different languages using the same script. If I were running a registry, I'd certainly consider it reasonable to apply some of

the tests that you suggest. But remember that DNAMEs can be created by anyone with appropriate dynamic update privileges and in zones anywhere in the tree pointing to anywhere else in the tree. If we say "the client MUST make these checks", then people will rely on their being made... and bad-guy clients simply won't.

I do believe that, at some point, the world is going to need better sets of "guidelines to registries and zone administrators for best practices" -- maybe globally (e.g., "you should think about these issues") but more likely on a per-language basis. But Rationale already does some of the former (personally, I'd welcome going a bit further, but WG consensus about what to say has been hard to achieve and it is even harder to know where to stop) and it is worth noting that there are already several per-script efforts for the latter (starting with the JET work that also brought us variants).

But trying to make a normative requirement would, I fear, just backfire.

### **091104 16:48 Andrew Sullivan**

*I was talking about a single DNAME definition, not about other DNAMEs that might enter the picture at some point.*

That sounds like operational guidance, not protocol. "Be careful" is not a useful thing to put in a protocol document, even if it's perfectly good operational advice.

I think you're right that it'd be nice to put together a set of guidelines for what people ought to do when operating IDNA-aware zones. There are several sets of considerations that need to be taken into account, and such advice ought indeed to be offered.

The IDNABIS WG is not the forum that should generate that advice: hardly anybody here is a DNS operator. I in fact previously offered (in a fit of insanity) to put together a -00 draft outlining such advice. I think it should probably be taken to DNSOP to see if there's any interest. But the response to me (one with which I have considerable sympathy) was that it'd be crazy to try to produce operational suggestions for a protocol that hasn't made it out of the IETF's process yet.

*If somebody were to implement a tool that checks DNAMEs before inserting them into a zone, that tool could check a \*single\* DNAME against the IDNabis bidi rules. Here is an example from <http://tools.ietf.org/html/rfc2672> frobozz.example. DNAME frobozz-division.acme.example.*

*When the tool is about to insert this DNAME into the zone, it can apply the bidi rules to the full domain name (which consists of multiple labels, as you can see).*

I think you don't fully understand how DNAMEs work. Your example does not alias "frobozz.example." to "frobozz-division.acme.example.", because DNAMEs don't alias the owner name where the DNAME appears. That DNAME does alias www.frobozz.example to www.frobozz-division.acme.example. It also aliases x1.frobozz.example to x1.frobozz-division.acme.example. In fact, it aliases

- .frobozz.example. The problem is to know whether there is in fact

the label  $\aleph_1$  in frobozz-division.acme.example, and without controlling both zones, the administrator of frobozz.example doesn't know (and can't know unless zone transfer is permitted or frobozz-division.acme.example is running DNSSEC with NSEC). This is why there is a problem. DNAMEs don't do what a lot of people think they do.

## Cross/inter label testing

### 091104 11:11 Vint Cerf

the question of inter-label or cross-label testing was extensively discussed on the WG list and rejected as overly complex at the protocol level. As with a number of cases, the WG concluded that the registry or registrar had to be cognizant of this kind of anomaly and reject problematic registration requests.

## Unicode TR46

### 20091027 Martin J. Dürst

*Subject: High-level comment on TR46*

I have been thinking a bit more about TR46 (<http://www.unicode.org/reports/tr46/>) and all the connections.

I think TR46 on a high level does try to do two things:

- Provide a well-defined mapping for domain names. I think the Unicode consortium is the best place to actually do this, and personally I can easily imagine that some IETF (or W3C) specs may refer to it.
- Try to "fix" some perceived problems in IDNA 2008 and between IDNA 2008 and IDNA 2003 (mainly sz, final sigma,...). While the intent here is very well-meant, and what's proposed is definitely one possible solution, from an IETF perspective, the Internet (and the use of IDNs) is far wider than browsers and search engines, and it is highly desirable for all parties involved if a solution to these problems comes directly from the IETF.

My proposal would therefore be to split the document, e.g. as follows:

- Keep TR46 with the uncontroversial mappings (essentially the extension of the IDNA 2003 mappings to the IDNA 2008 repertoire)
- Submit the proposal for how to deal with both IDNA 2003 and IDNA 2008 at the same time, and in particular how to deal with the special cases (called "deviations" in TR46) as an Internet-Draft (best with some co-authors from other affected communities, e.g. from DENIC,...). That's the best way to get the IETF to actually face the issues.

This is a 'refinement' from what we discussed with Larry about two weeks ago in Mountain View.

## 20091226 MARK DAVIS

My bandwidth is extremely limited until we get back to the states, so I will be brief. Please forgive me if by being brief, I am also overly brusque.

1. I have not been able to follow the 4 deviation character discussion, but it appears that there is agreement on some transition strategies that will work; a key approach appears to be to map on the client side if one is sure that the zone bundles, otherwise map.
2. Given that, I'd anticipate that the UTC would modify TR46 to be (a) support of symbols for some transitional period, and (b) a standard mapping. The rest of my comments are on the mapping issue.
3. One uniform mapping would be better than multiple, inconsistent mappings.
4. While one could argue either way, the advantage of the TR46 mapping is that it preserves compatibility with IDNA2003.
5. The current IDNA2008 mapping wouldn't maintain that compatibility, falls short in a number of cases for languages that don't have case/width issues, and has a number of formal problems.
6. We have major vendors that intend to implement the TR46 mappings; I don't know of any that have signed up to implement the current idna2008 spec.
7. The supposed argument from "harm" is specious.
8. First, there is a mixup below. If X is confusable with a PVALID Y, it is no problem to map X to Y; it would only be a (theoretical) problem if X were mapped to a PVALID Z.
9. Vastly more importantly, the argument from "harm" is faith-based, not data-based. I don't have access here, but I previously posted notes on the relative frequencies of spoofing techniques. From that data:
  1. Spoofing with confusable characters is FAR below spoofing with syntax (like <http://safe-amazon.com>) in frequency.
  2. There are essentially no letters that can be spoofed with the mapped characters that can't also be spoofed with other letters that are PVALID.
10. In sum, allowing the additional mappings makes \*no\* significant difference in the ability to spoof.
11. Best would be to incorporate the TR46 mappings into IDNA2008. Second best would be to reference them; third would be to remove the idna2008 mappings document, and fourth would be to leave them as is, and just deal with the muddle that results.

Mark

## "eszett" exception

### Maryhofer - NIC.AT 20091025

As you probably know, i'm working for the Austrian Domain Name Registry (nic.at). I've recently prepared a presentation to our board regarding the changes to expect from IDNAbis deployment, and I've been asked by our board to voice our concerns about the "szett" (U+00DF) exception in the current document set. I understand that the documents have progressed very far, and that we should have voiced our concerns earlier - however, i think that the information below is still valuable to the group.

Obviously, the DNS is an extremely important identity and naming system that is crucial to the operation of nearly all internet applications. Therefore, any changes to that structure are delicate operations. This is important for the creation of new portions of namespace, but particularly important when the semantics of a namespace (portion) are changed. The introduction of IDNA2003 was an extension of the namespace, at least from the application perspective (technically, it was changing the definition of an awkward-enough portion of the namespace, namely labels with "xn--").

Changing the semantics of a certain namespace is *\*really bad\**, and i agree to what Marcos said long time ago "Breaking backwards compatibility is to my eyes the big stigma of IDNA2008".

I understand and welcome the introduction of rigid rules in IDNAbis as the primary mechanism to identify copepoint classification and protocol validity. Independence from a certain Unicode revision ensures a stable specification, and should create few "surprises" (essentially, it shifts responsibility of character classification from the IETF to Unicode). I also understand and welcome the 1:1 relation on the protocol level between A-label and U-label.

However, the introduction of *\*exceptions\** that work around those rigid rules, and particularly changing the semantics of a part of a deployed, used namespace is *\*really really bad\** - particularly if the exception concerns such a "weird" character as the "szett" (Unicode folding-wise). Such changes generally have the potential to change the resolved destination for a certain domain name, which in turn creates *\*major\** security issues, and hurts interoperability badly, because unlike the introduction of IDN2003, where a label would either work or not, those exceptions now create a situation where such a label would resolve to either destination A (old application), destination B (new application).

I understand that the Rationale document proposes sensible approaches in Section 7.2 - however, i think the security issues could discuss the problems more explicitly, rather than just referring to the rationale document (which is informational anyways). I think that the sentence

"...a few characters that were mapped to others in the earlier version; zone administrators should be aware of the problems that might raise and take appropriate measures"

In the definitions document could easily be overlooked by implementors.

Another issue makes it even harder for zone administrator to deal with the problem: Actually *\*encouraging\** application developers to create their own fancy mapping definitions, beyond the mappings that were included in IDNA2003 allows for even more "variations", and are bound to hurt interoperability badly. One example of this is the Unicode TR46, particularly the proposal of "dual lookups" and "trusted registries" for "Deviations", which i believe to be a really really bad idea - but what are the other options?

Shifting the responsibility of mapping, and therefore allowing for creating a myriad of mapping options to application developers seems risky to me, particularly for the Exception codepoints for which protocol definitions have changed between the two versions. From my point of view, it makes such codepoints unusable - the "mapping du jour" of application X could be entirely different than that of application Y.

The Mapping draft says that it's "unusual" for the IETF to discuss user input processing steps - but on the other hand, Section 2.1 of [RFC 3761](#) (the ENUM base specification) clearly provides normative text about how user input should be prepared for a protocol (and i'm sure there are many other examples). So it seems the IETF *\*is\** concerned about how user input is mapped to protocol elements.

To sum up, we would have preferred the "szett" (U+00DF) to be kept "DISALLOWED", and to have the IETF describe the mapping procedures not just "Informational" (The contents of the mapping document itself is perfectly fine). We also hope that the IETF liases with application developers, particularly browser vendors, to establish one single "de facto" mapping procedure, so that at least the szett does not become a moving target.

### **Mark Davis, Unicode, 20101026**

The Unicode Consortium shares your concerns about the treatment of deviations, and the security and interoperability issues resulting from that and custom mappings. Unfortunately, while those points were raised consistently during the development of IDNA2008 (some would say too persistently), the working group decided on its current course.

We have been consulting with browser and search engine vendors (many of whom are members of the consortium), and I would anticipate that most will not end up implementing IDNA2008 lookup as is because of the problems it has. TR46 is designed by those needing to implement IDNA lookup so as to provide a bridge specification, whereby implementations can maximize compatibility with IDNA2003 and IDNA2008 on the lookup side, and avoid these problems. On the "dual lookup" and "trusted registries" point that you mention: the text of TR46 is insufficiently clear. That section is discussing alternative approaches that were considered, but discarded (because they don't work well, as Marcos pointed out in detail). I'll make sure that that feedback is brought into the committee.

TR46 is not really aimed at the registry side. It is feasible for registries to implement IDNA2008 if they additionally DISALLOW the four deviations (including es-zett). This can be done while being conformant to IDNA2008, because registries can further limit the characters they support.

WG/IDNABIS Chair dialog with UNICODE before finalizing IDNA2008 text and sending it to IESG.

## Vint Cerf letter to Unicode

### Letter to Unicode

Ms. Lisa Moore  
Chairman, Unicode Technical Committee  
via email: lisam@us.ibm.com  
CC:  
Eric Muller  
Vice Chairman, Unicode Technical Committee  
via email: emuller@adobe.com  
Mark Davis  
President, Unicode Consortium  
via email: markdavis@google.com

28 November 2010

Dear Ms. Moore:

I am writing to you in my role as chairman of the IDNABIS working group, addressing this request to you as president of the Unicode Consortium. As you know, treatment of the two characters, Greek Small Letter Final Sigma (U+03C2) and Latin Small Letter Sharp S (U+00DF) have been the source of considerable discussion during the IDNABIS Working Group effort on specifying the IDNA2008 proposed replacement of the IDNA2003 standard for the use of Unicode in Internationalized Domain Names. Latin Capital Letter Sharp S (U+1E9E) was added in Unicode version 5.1.0 but recommended rules for its use were provided as shown below:

*Begin quote from Unicode Version 5.1.0*

#### ***Tailored Casing Operations***

*The Unicode Standard provides default casing operations. There are circumstances in which the default operations need to be tailored for specific locales or environments. Some of these tailorings have data that is in the standard, in the SpecialCasing.txt file, notable for the*

*Turkish dotted capital I and dotless small i. In other cases, more specialized tailored casing operations may be appropriate. These include:*

*Titlecasing of IJ at the start of words in Dutch  
Removal of accents when uppercasing letters in Greek  
Uppercasing U+00DF ( ) LATIN SMALL LETTER SHARP S to the new U+1E9E LATIN CAPITAL LETTER SHARP S*

*However, these tailorings may or may not be desired, depending on the implementation in question.*

*In particular, capital sharp s is intended for typographical representations of signage and uppercase titles, and other environments where users require the sharp s to be preserved in uppercase. Overall, such usage is rare. In contrast, standard German orthography uses the string "SS" as uppercase mapping for small sharp s. Thus, with the default Unicode casing operations, capital sharp s will lowercase to small sharp s, but not the reverse: small sharp s uppercases to "SS". In those instances where the reverse casing operation is needed, a tailored operation would be required.*

*End quote from Unicode Version 5.1.0*

In IDNA2003, Sharp S was mapped to "ss" by means of a casing operation that mapped lower case Sharp S to uppercase "SS" and then down to lowercase "ss". Registrations and lookups using the IDNA2003 rules applied this mechanism.

During the discussions in the IDNABIS Working Group on IDNA2008, a strong consensus developed around not mapping for example for registration purposes and also for preserving the property that the IDNA2008-defined A-Label and U-Label forms be fully symmetric (i.e., convertible into one another without change or loss).

During these same discussions, a consensus seemed to develop to permit (ie. make "PVALID" in IDNA2008 parlance) Latin Small Letter Sharp S (U+00DF) and Greek Small Letter Final Sigma (U+03C2). The recommended casing actions of Unicode (i.e. toCaseFold) on Sharp S and Final Sigma produce "ss" in the case of Sharp S and Greek Small Letter Sigma (U+03C3) in the case of Final Sigma.

To make the lowercase forms PVALID using the functional rules of IDNA2008, exceptions were required to overcome the recommended casing mechanics of Unicode (i.e. application of CaseFolding).

Note that IDNA2008 explicitly permits mapping for User Interface purposes:

- a) draft-ietf-idnabis-protocol-17#section-5.2
- c) draft-ietf-idnabis-rationale-14#section-4.4
- d) draft-ietf-idnabis-rationale-14#section-6
- e) draft-ietf-idnabis-rationale-14#section-7.3
- f) draft-ietf-idnabis-mappings-05



If Small Letter Sharp S and Small Letter Final Sigma were to be made DISALLOWED, these mapping provisions would permit these characters to be handled as a User Interface matter prior to lookup.

Because the practices of IDNA2003 are in conflict with the proposed practices of IDNA2008, and because the Last Call discussions have surfaced controversy over the incorporation of the two lowercase forms in question, I request an organizational recommendation from UTC as to the treatment of these characters. Taking into account the prohibition of mapping on registration, which I take to be firm, and the requirement that A-Label and U-Label forms must be unambiguously convertible into each other, would the UTC recommend to exclude the use of Small Letter Sharp S and Small Letter Final Sigma in IDNA2008 by removing their exceptions and making each DISALLOWED?

A prompt response would be much appreciated considering we have delayed reporting the results of the IETF LAST CALL to the Internet Engineering Steering Group while this matter is debated.

Sincerely,

Vinton Cerf  
Chairman, IDNABIS Working Group of the Internet Engineering Task Force

## **Unicode response**

Dear Vint,

The UTC appreciates the difficulty for users of IDNs, the registries, and all involved if lowercase sharp s (Latin Small Letter Sharp S (U+00DF)) and small final sigma (Greek Small Letter Final Sigma (U+03C2)), in particular, are mapped to other characters. As you know, our concerns are compatibility and potential security issues. However, based on the many ongoing discussions and much thought, the UTC would not be opposed to have lowercase sharp s, final sigma, and even joiner and non-joiner be valid and not mapped, as long as there can be policies in place for a transition period (of say 5 years) that will manage the expected compatibility issues.

The key for us is having policies for a well-managed transition with sufficient time for browser and other application upgrades. Without such policies in place, we would favor continuing the IDNA2003 treatment of the four above-mentioned characters.

Best regards,  
Lisa Moore  
Chair, Unicode Technical Committee

[Mark Davis, Unicode President](#)

## Problem

We would like to have the 4 deviation characters be valid, at some point. The key problem is that we don't want current URLs in web pages, etc. to go to two different locations depending on the browser, nor do we want *joe@fußball.com* to go sometimes to *joe@fußball.com* and sometimes to *joe@fussball.com*. Even once IDNA2008 is approved, for a long time a majority of the implementations will still be IDNA2003, so this also goes for new label registrations during the transition period.

## Proposal

IDNA2008 changes as follows:

The 4 deviation characters get the property PVALID\_AFTER\_2015

The requirements are:

On registration, PVALID\_AFTER\_2015 is equivalent to PVALID

On lookup, PVALID\_AFTER\_2015 is treated as DISALLOWED up until 2016 Jan 1, 00:00:00 GMT, and treated as PVALID thereafter.

Implementations must not map the characters after the switchover date.

Implementations that map the characters before that date, must map as in IDNA2003.

The goal is to

allow the 4 character to become valid, as soon as possible; avoid the 'nightmare' scenario of the same URL going to two different locations, as much as possible.

## Scenarios

Let's see what happens with fußball.xxx over time, where xxx is some registry (eg .de, .blogspot.com, or others). Background: essentially all browsers and other major implementations are planning to map for compatibility. We'll look at browsers, but this also applies to email, etc.

**Early 2010 (just as IDNA2008 is approved)**

**At this time the world browsers are 100% IDNA2003**

- browsers map fußball.xxx to fussball.xxx.
- registries can start accepting eszett, and should bundle with ss.
- fußball shows up as fussball in the address bar
  - note: it is only by convention that fussball is seen in the address bar in this case; a browser could also display fußball, as in UTS46.
- results:

- if the registry bundles, both fußball.xxx and fussball.xxx go to the same owner.
  - if the registry doesn't bundle, both fußball.xxx and fussball.xxx go to the same owner.
- The odd IDNA2008 browser that doesn't map just fails, because ß is not PVALID; it doesn't take fußball.xxx to a different location than the vast majority of browsers.

### **In 2013**

**At this time the world browsers are 50% IDNA2003, 50% IDNA2008**

same as above. No ambiguity in results.

### **In 2016 Feb**

**At this time the world browsers are 1% IDNA2003, 99% IDNA2008**

- 99% of browsers switch to not mapping fußball.xxx.
- Registries no longer need to bundle; they can have different owners for fußball.xxx and fussball.xxx.
- fußball shows up as fußball in the address bar
- results:
  - if the registry bundles, both fußball.xxx and fussball.xxx go to the same owner.
  - if the registry doesn't bundle, fußball.xxx and fussball.xxx go to different owners.
- The odd IDNA2003 browser that is left goes to the wrong location for the affected languages; people that use them need to upgrade.

## **Acknowledgment of the WG/IDNABIS Chair**

Dear Lisa,

Thank you and the UTC for its rapid response.

I believe that the discussions of the past week have confirmed a general consensus on the preference that Final Sigma and Sharp-S be PVALID. We did not poll for the joiner/non-joiner question because a consensus already existed, in my opinion, as chair, for these to be contextually valid (CONTEXTJ).

The method of introduction of IDNA2008 is important to all of us, to promote its utility. At the close of the day, I will review all of the comments received and attempt to synthesize what I believe is a plan around which consensus can be obtained.

Vint Cerf

## Pete Resnick's Recipe

*>> One slightly more solid question for browsers is, would it be entirely crazy to have different mapping algorithms for typed-in domain names than for links followed? There might be a locale-dependent mapping as well as a global mapping. (I assume that having every established locale mapping installed would be complete craziness.)*

*> IMO crazy. People type URLs into web browsers. Other people type URLs into HTML documents. We can't have the same letters in those two locations taking you to different places.*

Generally staying out of this discussion, but: If we were considering to do this correctly (IMO), typed-in domains with DISALLOWED characters would get mapped, and links followed would get "Bite Me!" in big bold letters and the equivalent of an NXDOMAIN. Anything that takes typed input that maps should dynamically "correct" to PVALID things, and anything that is taking data should scream bloody murder if the domain labels are not all PVALID.

Yes, I know I'm not convincing anyone of this. But it's the right thing to do.

Back to my hidey-hole.

pr

## IDNA 2008 in light of DYN DNS

### John Klensin 200910025

> Uh, yes. If dynamic update is configured to require that an > RRSET (from the viewpoint of IDNA, a label) is already present, > then one has a lookup situation. If it is configured for the > "RRSET does not exist" or "Name not in use" cases, then one has > a registration situation. That said, my personal recommendation > would be to use the more conservative Registration rules any > time one is going to start modifying DNS zones rather than > simply looking something up. But the WG has not discussed this > topic. If people are convinced that something must be said on > the subject, we will need to have that discussion.

### Bernard Adoba 20091026

I do think that something needs to be said about this, since the issue has come up in implementation. For example, based on the distinction above, a client handling a dynamic update on its own using TKEY would implement the lookup protocol, whereas a DHCP server handling a dynamic update on behalf of the client might implement the registration protocol.

### Vint Cerf 20091026

another alternative would be for you to issue an informational RFC about interpretation of IDNA 2008 in light of DYN DNS, would it not?

## RTL & Labels

### Abdulrahman I. ALGhadir 20091013

while I was reading draft-ietf-idnabis-bidi-06.txt I found this:

“4.3. Strings with numbers

*By requiring that the first or last character of a string be category R or AL, [RFC 3454](#) prohibited a string containing right-to-left characters from ending with a number.*

*Consider the strings ALEF 5 (HEBREW LETTER ALEF + DIGIT FIVE) and 5 ALEF. Displayed in an LTR context, the first one will be displayed from left to right as 5 ALEF (with the 5 being considered right-to-left because of the leading ALEF), while 5 ALEF will be displayed in exactly the same order (5 taking the direction from context). Clearly, only one of those should be permitted as a registered label, but barring them both seems unnecessary.”*

Why permitting done on the protocol level ? shouldn't this be done at registry level?

Ex. If someone wants to register 3COM (COM is a RTL word) registry will register 3COM for him/her and will lock COM3. At least this will give users the choice for picking not forcing them on one type?

### Vint Cerf 20091013

for the most part, the IDNA2008 specification does place a great deal of responsibility on the registry but for some cases, it was considered important to bar particularly confusing situations at the protocol level. This was debated substantially during the course of the development of IDNA2008.

### Abdulrahman I. ALGhadir 20091014

Well yes it is confusing and yes there are other confusing problems which the protocol simply assign it to registry what I say it is unfair to prevent all domains Which they start with digits to be as choice for user if this problem moved to registry level a registry can simply do:

- 1) Act same as what the protocol do now (preventing all domains which start with digits).
- 2) Register domains which start with digits and lock the same domain which end with digits and the opposite.

And any further problems which may result from using 2) can be solved as the registry wants. Plus if the protocol allowed this maybe bidi algorithm will support domains who knows.

## Full stop Mapping

### 20091011 Sarmad Hussain

In earlier discussions on U+06D4 (ARABIC FULL STOP), which is necessary for Urdu as a label separator (the reasons have been given on this list earlier), it was suggested that the various full stops will not be allowed and be mapped. It was subsequently requested to include the mapping reference in IDNA200x documents to ensure that the application providers incorporate it, but the request was not considered positively as it was perhaps suggested that such recommendations can not be made part of the protocol. However, the recent mapping document (<http://tools.ietf.org/html/draft-ietf-idnabis-mappings-04>) says on pg. 2:

4. [I-D.ietf-idnabis-protocol] is specified such that the protocol acts on the individual labels of the domain name. If an implementation of this mapping is also performing the step of separation of the parts of a domain name into labels by using the FULL STOP character (U+002E), the following character can be mapped to the FULL STOP before label separation occurs:

- IDEOGRAPHIC FULL STOP (U+3002)  
There are other characters that are used as "full stops" that one could consider mapping as label separators, but their use as such has not been investigated thoroughly.

If this is being explicitly done for U+3002, it could be done explicitly for ARABIC FULL STOP (U+06D4) as well. What is the reason for not including other such possible cases explicitly?

### 20091011 Vint Cerf

keep in mind that the Mappings document is NOT normative. It is intended to give some ideas for localization and pre-processing. The important point is that only U+002E will be recognized in protocol as a label separator. For purposes of exchanging IDNs, that's important. For local contexts, one might allow alternative full-stop inputs but these would need to be converted to the U+002E form prior to initiating a DNS query. It would probably be wise also to convert to U+002E for purposes of canonical exchange of domain names with other parties.

For Pete Resnick and Paul Hoffman:

this email might be interpreted as a request to add U+06D4 to the Mappings list of potential local mappings to U+002E. Have you an opinion whether this edit would be appropriate?

## **200910111703 Erik van der Poel**

In my opinion, it would be premature to include U+06D4 in the IDNAbis mapping draft (apart from the fact that it is rather late in the Last Calls process to be making such a change). U+3002 has a much longer history in IDNA and is much more firmly established. If U+06D4 would be mapped to U+002E at the data interchange level (think HTML), there would be a period where IDNA2003 implementations and new implementations would resolve domain names differently. Of course, the IDNAbis mapping draft explicitly states that it is intended to be used at the UI level (e.g. keyboard input), but, frankly, I don't think we have much experience with IDNA implementations that distinguish between the UI and data interchange levels.

## **20091011 At 17:24 int Cerf**

Thanks for this observations.

Clearly we would not advocate use of U+064D for interchange. Mappings was intended to focus on non-interchange UI treatments.

I agree that this is a rather substantive change; may we hear from others in IDNABIS WG please?

We are going to have to draw these last call discussions to a close soon.

vint

## **20091011 At 18:52 11/10/2009, Paul Hoffman**

It is not late at all. We are still in IETF Last Call, which is exactly the time we are supposed to be having these community-wide discussions.

*U+3002 has a much longer history in IDNA and is much more firmly established.*

Correct.

*If U+06D4 would be mapped to U+002E at the data interchange level (think HTML), there would be a period where IDNA2003 implementations and new implementations would resolve domain names differently.*

No one has suggested that we do that, of course. If someone wants to mis-implement draft-ietf-idnabis-mappings, there is nothing we can do to stop them.

*Of course, the IDNAbis mapping draft explicitly states that it is intended to be used at the UI level (e.g. keyboard input), but, frankly, I don't think we have much experience with IDNA implementations that distinguish between the UI and data interchange levels.*

So, what do you propose instead? If this is a proposal to abandon draft-ietf-idnabis-mappings, you need to be more explicit about it. If it is not such a proposal, then you need to say what parts of the draft need to be abandoned to deal with your concern.

**20091012 At 02:17 YAO Jiankang**

*In my opinion, it is not necessary to abandon idnabis-mappings.*

+1

*It is trying to solve a perceived problem at the UI level (which may be somewhat unusual for an IETF document). As long as it is not normative, it's fine with me. Maybe it should become an Experimental RFC instead of Informational.*

I think that at least it should be informational. Mapping is also very important. it is one way to make idna2008 be compatible with the idna2003 protocol. the suggested category is proposal standard.

**20091012 At 14:53 John C Klensin**

I see two issues with this idea. The second leads to a suggestion.

(1) While very local mapping of full stops makes perfectly good sense, any leakage is going to interfere with programs that need to parse dot-separated-label form into length-label pairs. Such programs may exist because they are not IDNA-aware or because they are trying to resolve names in private name spaces and follow the [RFC 2181](#) observation that the DNS can accommodate any string of octets. We know that things "leak", so these mappings need to be performed very carefully --even more carefully than mappings of characters within labels and any document should reflect that.

(2) One of the advantages of the list of characters that are now discussed in the mapping document is that the list is fairly stable. Because it is largely motivated by within-label IDNA2003 compatibility, there should be little or no need to expand the list as Unicode evolves. That is a good thing because we don't have comprehensive rules to generate the character and mapping list (even though much of it is either NFKC or CaseFold) -- the mappings document is essentially Unicode version-dependent.

By contrast, I think the general rule for candidate alternate full-stop characters is going to be:

- (i) The traditional, DNS-specified, label separator U+002E (ASCII dot) is unnatural or hard to type or render in the local environment.
- (ii) There is a character in the local environment and script that is in common use, that is an obvious logical substitute label separator, and possibly that users will expect it to be a substitute regardless of what we have to say on the subject.



Although I think one can make a slightly stronger argument for U+06D4 because of bidi implications, I don't personally see a very strong justification for a recommendation to map one of these characters and not other. Erik and others have made that point for some specific cases.

The list of alternate full stops (referred to on the IDNA list for a while as "dot-oids") second list initialized with the three East Asian full stops called out in [RFC 3490](#), i.e.,

U+3002 (ideographic full stop),  
U+FF0E (fullwidth full stop), and  
U+FF61 (halfwidth ideographic full stop)

plus U+06D4 (Arabic full stop) for the reasons identified in Sarmad's note.

But a search in the UnicodeData file shows 38 characters with names containing "FULL STOP" and 60 (22 more) containing "STOP". Some of these are clearly irrelevant compatibility characters (e.g., those in the 2488..249B range which encode numerals followed by full stop) as single code points or used for some completely different purpose (e.g., a series of Glottal Stop code points). We've also been told repeatedly that putting too much reliance on the assumption that Unicode names encode characteristic information is a bad idea -- there may well be characters out there that could be seen locally as reasonable label separators that are not identified as "Full Stop" in their names.

So building a comprehensive list would require character-by-character examinations and decisions based on local contexts and usage. While we clearly have the expertise available to make those judgments for East Asian scripts the Urdu use of Arabic script, we do not have that knowledge in the general case. The list is not close-ended either. As new scripts are added to Unicode, it is safe to assume that at least some of them will have their own full stop (or equivalent) characters that will need to be considered for addition to the list.

Whether publishing a list of recommended mappings to U+002E is a good idea or not depends on (1) above and causes me to again recall that only ASCII separators are permitted in IRI and URI syntax -- separators that raise many of the same issues as the label separator in domain names. But, assuming that there is rough consensus that doing so is appropriate, I suggest that the right way to proceed is to create yet another document that

- focuses on the label-separator mapping alone
- explains the issues and risks of doing such mappings
- creates an IANA registry, initially populated by the four characters identified above, and with a mechanism for adding addition characters to it as the need for them is identified.

Since that document would presumably be informational anyway, if Lisa is agreeable, it could be handled as an AD-sponsored individual submission, thereby separating it from the WG's schedule and task list. I would hope we would not go forward with it (and that Lisa would not sponsor it) unless there were rough consensus on this list that having such a document and list of characters would be wise. But decoupling it from the current Mapping

document seems to be appropriate, if only because of the expandability and open-endedness of the list of characters.

## **20091012 At 15:22 Paul Hoffman**

*But decoupling it from the current Mapping document seems to be appropriate, if only because of the expandability and open-endedness of the list of characters.*

Sarmad's message did not seem to be a request to add all full stops to draft-ietf-idnabis-mappings, just U+06D4. (Sarmad, please correct me if I'm wrong.) The list that is given in the draft is clearly labelled as examples of full stops that a UI implementer might consider, not as the full list. The reason that the list exists at all is because of input from experts in a particular script; it seems reasonable to take input from experts on other scripts as well. If we get bombarded with experts from more than half a dozen languages before the end of IETF Last Call, I could see your view, but this is a single request from someone in a language community that has been part of the IDNabis process for quite some time.

Before I consider adding U+06D4, I would want to hear from additional experts, but other than that, I see no danger in adding another character to an optional list in an optional document.

## **20091012 Shawn Steele**

*I think that at least it should be informational. Mapping is also very important. it is one way to make idna2008 be compatible with the idna2003 protocol. the suggested category is proposal standard.*

Mapping isn't at all sufficient for IDNA2003 compatibility.

First of all, if there is going to be a list in the document, I think Sarmad's request (which has been supported by other analyses, by the way) is a completely reasonable one and that it should be added to the list. I am, of course, not an expert on either Urdu or Arabic script usage generally, but the consensus among those who are seems to be that adding this is at least as strongly justified as the East Asian characters we have and, at worst, harmless (whether you believe Sarmad, or me, about that or whether you seek other experts is up to you... and Vint).

One can generalize from the "harmless" part: it looks like all of the likely possibilities are going to be Po characters and hence DISALLOWED. As long as that relationship holds, and as long as they don't leak, they are all harmless. If they leak, we get into a whole new family of visual confusability issues and questions, but it seems to me that "won't leak" is a fairly basic assumption of the mapping document.

We discussed label-separator mappings (including, if I recall, U+06D4) much earlier in the life of the WG. What I think Sarmad's request points to was part of that earlier discussion: the East Asian character list is not a complete list of all reasonable "full stop" characters that

might be justified and recommended for mapping as label separators. That pointer is independent of the specifics of U+06D4, whether he intended it or not.

*Before I consider adding U+06D4, I would want to hear from additional experts, but other than that, I see no danger in adding another character to an optional list in an optional document.*

I don't either (see above about "harmless"), except that I think it would be very unfortunate to have to reopen and re-review this document in six months or a year when someone argues that one or two of U+0589, U+1632, U+166E (a special case because I'm sure Eric can advise us as to whether that request would be likely to arise and be plausible), U+1803, and so on should be added to the list for the same types of reasons as the ones Sarmad describes.

So I guess I'm making a recommendation about something I should have spotted and made the recommendation about long ago: create a registry for these things. The reason for that is precisely to prevent our needing to reopen the mapping document to add one or more of these characters, not to treat them with more or less authority. I think such additions are extremely probable, either as new scripts are added to Unicode or as communities that are now unrepresented in this WG and probably underrepresented on the Internet show up and explain their needs. Of course, YMMD on that likelihood assessment.

My suggestion about careful explanation and cautions is separate but, if we preserve the current model for separation of materials, if the registry is created by the mapping document, the explanation would need to be in Rationale anyway.

john

**At 17:06 12/10/2009, Paul Hoffman wrote:**

*At 11:13 AM -0400 10/12/09, John C Klensin wrote: I don't either (see above about "harmless"), except that I think it would be very unfortunate to have to reopen and re-review this document in six months or a year when someone argues that one or two of U+0589, U+1632, U+166E (a special case because I'm sure Eric can advise us as to whether that request would be likely to arise and be plausible), U+1803, and so on should be added to the list for the same types of reasons as the ones Sarmad describes.*

Fully agree. I can't speak for Pete, but I am not willing to re-open the document to add characters to a list of optional characters. In fact, I'm not even sure there will be a "we" around in six months: the WG should close down after we have fulfilled our charter.

*So I guess I'm making a recommendation about something I should have spotted and made the recommendation about long ago: create a registry for these things.*

And here we fully disagree. I think defining the rules for registries for any of the optional parts of the document will take at least another year, and we have already hit the exhaustion point. I see absolutely no upside for a registry of optional suggestions for an

Informational RFC, with a significant downside of taking a lot of review time that could be better spent elsewhere.

*The reason for that is precisely to prevent our needing to reopen the mapping document to add one or more of these characters, not to treat them with more or less authority.*

I see no "need" to reopen the document, ever. If someone wants to prepare a document with a different view, or with the same view but additional characters, that's fine, and quite easy in the RFC process.

*I think such additions are extremely probable, either as new scripts are added to Unicode or as communities that are now unrepresented in this WG and probably underrepresented on the Internet show up and explain their needs. Of course, YMMV on that likelihood assessment.*

We certainly agree on the likelihood of desired additions, but I think not on the "need" for a revision to the WG's document.

**At 17:16 12/10/2009, Vint Cerf wrote:**

Folks,

I think we might be able to come to a conclusion this way:

1. it is important that we agree that the canonical, exchange format for domain names contain ONLY U+002E as the label separator
2. UI contexts and practices will likely vary and reasonable choices for local use of "full-stop" characters will vary
3. The mappings document is essentially informational and notional, not prescriptive, so editing it as if it were normative is likely counterproductive
4. The idea of documenting UI practices seems useful, but, as I think Paul argues, this need not be the work of the IDNABIS working group.

If ICANN sees some value in (4) perhaps it can formulate a kind of "known mapping practices" registry (I hesitate to say "best practices") for various scripts (languages??) I don't see this as an IETF function, however.

comments?

vint

## **Bidi discussion**

**From: jefsey Sent: 13/Feb/2010 4:15 PM**

As you know I prepare an appeal to the IESG over the IDNA issue. In a nutshell this appeal concerns the way IESG/IAB handles IDNA2008 as a standalone architectural issue. On an usage point of view the architectural concepts approved at the occasion of IDNA2008 legitimate a much more innovative vision of the same Internet technology. However, that vision introduces new issues which should be addressed or at least announced at the same time in order to avoid confusion. The real matter therefore is most probably with the IAB. My evaluation is that IDNA2008 uses the Internet technology intrinsic capacity for multiplicity (what we are not used to) to address diversity.

The purpose of the appeal is therefore to obtain an (we think urgent) authoritative IESG and probably IAB position about:

- (1) the way they consider the IDNA2008 architectural insertion and impact
- (2) if such an impact exists who is to address its consequences. IETF, users, someone else, under which form.

In order to clarify the issues I consider three areas needing to be documented:

1. IDNA and the Internet system - this is what IDNA documents - where IETF can specify MUSTs
2. IDNA and the IDNs Internet peritem - this is what Mapping considers.- where IETF can specify SHOULDs
3. IDNA and the Users Internet exotem - this is the IDNA open use architecture by users' real world, which is not considered - where

IETF can specified MAYs

I am not really familiar with Bidi. My question therefore is: where in the above scheme do you locate your discussion?

1. does that affect IDNA2008 as the interface between the DNS and IDNs?
2. does that relates to IDNs conversion from/to A-labels/U-labels
3. does that belongs to the User experience

My question is how Bidi should be best inserted in the appeal I consider to be of best use?

**Slim Amamou <slim@alixsys.com> 14 février 2010 04:57**

hi,

It belongs probably to User experience. Although the issue is broader than that, and deals with the question of the universality of the representation of the domain name. In other words : will a domain name look the same in all countries, for all cultures and on every media (screen, print, etc...) or not?

**Abdulrahman I. ALGhadir <aghadir@citc.gov.sa> 14 février 2010 05:39**

*1. does that affect IDNA2008 as the interface between the DNS and IDNs?*

No, it doesn't.

*2. does that relates to IDNs conversion from/to A-labels/U-labels*

No, it doesn't.

*3. does that belongs to the User experience*

Well, it doesn't give the user him/herself any problem in reading domain names (because they are rendered in correct way) the issue is that the logical order sometimes doesn't have the same order as network order (which is against some rules in IDNA rfc and some other rfcs).

*My question is how Bidi should be best inserted in the appeal I consider to be of best use?*

Well I don't see the IDNA has major impact on this problem (it is rendering problem) so I am not sure in this part.

**Slim Amamou <slim@alixsys.com> 14 février 2010 11:23**

*does that belongs to the User experience -- Well, it doesn't give the user him/herself any problem in reading domain names (because they are rendered in correct way)*

Actually Abdulrahman and I are not agreeing here. I started the thread because I consider rendering L1.R2.R3.L4 as L1.R3.R2.L4 obviously not correct. And I still don't understand how this could be acceptable for anyone. Note that this will be a very common usage, since most domain names begin with LTR www. and end with LTR TLD.

**Vint Cerf <vint@google.com> 14 février 2010 15:36**

unless there are strong indicators that a string IS a domain name, this is going to be an unsolvable problem I think. Won't it also depend on the actual label contents and the appearance of numerics in the labels, to add more complexity?

vint

**Slim Amamou <slim@alixsys.com> 14 février 2010 20:41**

*unless there are strong indicators that a string IS a domain name, this is going to be an unsolvable problem I think.*

I think dealing with cases where a domain name can not be identified as such is out of the scope. It always was, there is no way to tell that example. is a domain name. But our problem could be narrowed to displayed URLs (or should we say IRLs?) and email addresses. In the two contexts the domain name is obvious.

*Won't it also depend on the actual label contents and the appearance of numerics in the labels, to add more complexity?*

No, because

1- the structure of the domain name will prevail. DNS was not meant to be used like del.icio.us or to write sentences using the period in place of the space. Some components of the internet depend on the implied hierarchy of authority. like HTML5 same origin policy for example, and in general whenever the word "subdomain" is used in a spec.

2- If someone really can't help using a domain name for writing a sentence like www.this.is.an8.label.domain.name.com in arabic, it won't matter to him to input it in reverse order : name.domain.label.an8.is.this.www during domain registration process. what will matter to him is that the domain will *\*always\** be displayed the same whether it is LTR or RTL context. (if this is not clear, please ask, I will develop more)

I'm aware that it would be confusing for someone writing RTL to input a domain name including periods during the registration process, just to have it inverted on display. But this is due to the limitations of the current deployed domain registration platforms, and could be mitigated by simple GUI improvements (GUI improvements have to be made anyway to make them BIDI capable).

**Vint Cerf <[vint@google.com](mailto:vint@google.com)> 14 février 2010 21:03**

Slim,

one doesn't register the whole domain name in one place - one registers a label with each zone manager.

so I am not sure we should take your example below as literally as you might have meant it?

vint

**Slim Amamou <[slim@alixsys.com](mailto:slim@alixsys.com)> 14 février 2010 22:50**

*one doesn't register the whole domain name in one place - one registers a label with each zone manager.*

But current registrar web interfaces allow for authority zone management. And from what I recall, they unfortunately allow periods in subdomains. If that is not the case, that's even better because enforcing network order for label ordering won't introduce any usability problem.

**Slim Amamou <[slim@alixsys.com](mailto:slim@alixsys.com)> 16 février 2010 10:21**

*(...) The way I approach semantic addressing is to locate a notion through its coordinated concepts. So, this is a genitive chain: abc of cde of ghi of ijk which can also be ijk's ghi's cde's abc. When it is a domain name it will use the first format, in the case it is a C structure it will be the second format. In that perspective I presuppose that order within labels is to be purely conventional respecting the punycode order respecting the user entered/registered order.*

I agree with this, and I chose the network order out of convenience. But what is a C structure?

*(...) To do that I just consider that the DNS uses alphanumerical digits (i.e. from 0 to Z, what fits with the way it relates to uppercases). I then consider that a label may be made of several sublabels united by "-" "subjoiner/subseparator", and that labels are tied into a LTR or RTL name sequences in using "." "joiner/separator" (the xn-- header is therefore a sublabel separated from the other sublabels, by an empty sublabel).*

So you are introducing a new separator class, namely the "-" within the label. Thus structuring the label in its punycode form (A-label) for the sake of meaningfulness. Right?

*In such a syntax only URN is reordered if it is a domain name or a structure, not sublabels. This means that this is transparent to punycode.*

For that matter, I support the non-reordering of URN components and subcomponents.

*As far as I understand Bidi is only treated on "UDN" (user entered domain name, that can still be more complex than the IDN U-label). This is why I am quite concerned by your mention that Bidi registries could "rigidify" the label order in entering "." that would be some kind of intermediary between what I call a subjoiner and a joiner. This could be addressed as a special joiner, but would complexify the whole syntax analysis. The same as a "to be hidden" subsubjoiner/separator, could indicate sequences indifferent to Bidi?*

Yes. If it's hidden, it's not a problem. In the layer you are working on, only ASCII is allowed, so BIDI and i18n altogether is not a concern.

*I wish to underline that in some way or another, a semantic address will have to indicate it is a semantic address and to indicate if it uses the DNS or the C structure format. The idea I currently have (to be script and language transparent) is to use something prefix like "|+|" and "|+|" or "+++" that are universal. Such a prefix/posfix to be removed before using the URL could also indicate that a string is a domain name (out of the habit in semantic addresses) ?*

I have no problem with that, it's just out of the scope of the issue I indicated.

**Jefsey**

*have no problem with that, it's just out of the scope of the issue I indicated.*

Yes.



This is totally out of the WG Charter. Yet, IMHO, the WG Charter if correctly completed (as it has been, except that the Mapping document is not approved yet) can only raise the question "is its continuation on the user side in the scope of the whole IETF, if yes who does take care of it, if not who should be the leader because IETF has to interface with them". This is why I consider that this is in the IAB scope/area of decision and that IESG should have not approved the IDNA2008 document set without making it extremely clear. So, no one respecting the IETF positions, like ICANN and responsible lead users do, attempts to use IDNA2008 in an operational (test or not) conditions.