

IDN Variant TLD Implementation: Motivation, Premises and Framework

25 January 2019

Contents

1	Background.....	2
2	Introduction and Motivation	2
2.1	Defining IDN Variant TLDs.....	4
2.2	Analyzing Variant TLD Management Mechanisms	5
3	Starting Premises for Implementing Variant TLDs.....	6
3.1	Variant labels are defined by the script community	6
3.2	An IDN variant TLD label is a TLD label.....	6
3.3	Root Zone Label Generation Rules are the single source of variant labels	6
3.4	IDN variant TLDs must not adversely impact the Domain Name System (DNS) and promote a good user experience	7
3.5	Active IDN variant TLDs should be minimized.....	7
3.6	Blocked variant TLDs should be maximized.....	7
3.7	Variant TLDs do not have a single and comprehensive technical implementation.....	8
4	Dimensions and Degree of Influence	8
4.1	Point A: ICANN's responsibility for the root requires a strong policy.....	10
4.2	Point C: ICANN's responsibility is limited to minimal administrative collision	10
4.3	Point D: Subordinate policies follow from variant status	11
4.4	Point F: Subordinate policies are as they are in other TLDs.....	12
4.5	Starting with an Intermediate Position	12

1 Background

The current report is part of the six documents finalized and published after the [public comment](#):

- A. IDN Variant TLD Implementation – Executive Summary
- B. IDN Variant TLD Implementation – Motivation, Premises and Framework
- C. IDN Variant TLD Implementation – Recommendations and Analysis
- D. IDN Variant TLD Implementation – Rationale for RZ-LGR
- E. IDN Variant TLD Implementation – Risks and their Mitigation
- F. IDN Variant TLD Implementation – Appendices (A: Glossary, B: Use ofROID, C: Limiting Allocated Variant TLDs)

2 Introduction and Motivation

Internationalized Domain Names (IDNs) enable people around the world to use domain names in local languages and scripts. While using IDNs, some script communities have identified that technically distinct IDN labels may be considered indistinguishable or interchangeable with other domain labels and therefore “same”, referred to as IDN variant labels. For example, the following may be considered variant labels by the script communities for different reasons:

- Chinese:
 - 澳門 and 澳门 (traditional and simplified Chinese)
- Arabic:
 - كتاب - کتاب (visually and logically same characters but different code points)
 - شبكة – شبکة (stylistically different orthography for same characters)
 - إمارات - امارات (spelling “simplification” by end-users but considered equivalent)
- Cross-Script:
 - Latin and Cyrillic: epic – epic (visually “same”)
 - Armenian and Greek: դրո – ηρο (visually “same”)
 - Kannada and Telugu: లఆగ్గ – లఆగ్గ (visually “same”)

As can be seen by the examples above, these IDN variant labels impact many different script communities globally.



The DNS was designed as an exact-match identifier system. The technical community, while stipulating how to use domain names in multiple scripts based on IDNs for Applications (IDNA) 2008 standard, also reflects on extra measures which are needed to manage the ensuing confusability introduced. In [RFC 5890](#)¹, it is stated that “DNS zone administrators may impose restrictions, beyond those imposed by DNS or IDNA, on the characters or strings that may be registered as labels in their zones. Because of the diversity of characters that can be used in a U-label [using Unicode standard] and the confusion they might cause, such restrictions are mandatory for IDN registries and zones even though the particular restrictions are not part of these specifications.” It is further explained that “DNS zone administrators may impose restrictions ... that try to minimize characters that have similar appearance or similar interpretations.” It is re-iterated in [RFC 5891](#)² that “Registries at all levels of the DNS, ... [including] the top level, are expected to establish policies about label registrations,” specifically pointing to the rationale in [RFC 5894](#)³ that “registries should develop and apply additional restrictions as needed to reduce confusion and other problems ... For many scripts, the use of variant techniques ... may be helpful in reducing problems that might be perceived by users. ... In general, users will benefit if registries only permit characters from scripts that are well-understood by the registry or its advisers,” suggesting some cases, e.g. “reduce opportunities for confusion by constructing policies that disallow characters used in historic writing systems or characters whose use is restricted to specialized, highly technical contexts”.

Based on the recommendations by the IDNA2008 standard, IDN variant labels must minimally be identified and managed to ensure that end-users are prevented from any security threats. Some of the IDN variant labels could possibly be activated to promote accessibility of the IDNs,

¹ Internationalized Domain Names for Applications (IDNA): Definitions and Document Framework

² Internationalized Domain Names in Applications (IDNA): Protocol

³ Internationalized Domain Names for Applications (IDNA): Background, Explanation, and Rationale

as different language communities using a script may use a different IDN variant label. Achieving these security and usability goals in a stable manner is the key challenges to be addressed.

Applicants for Internationalized Domain Name (IDN) country code Top-Level Domains (ccTLDs) and the new IDN generic Top-Level Domains (gTLDs) for some scripts (such as Chinese Han and Arabic) have indicated that the user community may consider different labels as variants of each other by identifying such candidate strings as part of their application. However, due to lack of a clear definition and a solution to implement them at the time, on 25 September 2010 ICANN Board [resolved](#) that “no variants of gTLDs will be delegated through the New gTLD Program until appropriate variant management solutions are developed.” Same has also been the case for variant labels for the IDN ccTLDs based on the Fast Track Process. However, for addressing the need from the community, the resolution further directed ICANN staff to develop “an issues report identifying what needs to be done with the evaluation, possible delegation, allocation and operation of gTLDs containing variant characters IDNs as part of the new gTLD process in order to facilitate the development of workable approaches to the deployment of gTLDs containing variant characters IDNs.”

To study these complex linguistic, technical and policy-related issues, ICANN organization undertook an initiative to engage experts from six script communities that analyzed issues in identifying variant labels for each of these scripts. This analysis of issues for [Arabic](#), [Chinese](#), [Cyrillic](#), [Devanagari](#), [Greek](#), and [Latin](#) scripts in 2011, collated in the [Integrated Issues Report, IIR](#) (2012), identified two challenges:

1. “in the DNS environment today, there is no accepted definition for what may constitute a variant relationship between top-level labels
2. “nor is there a ‘variant management’ mechanism for the top level, although such has often been proposed as a way to facilitate solutions to a particular problem.”

2.1 Defining IDN Variant TLDs

IIR outlined the follow-up work that might be undertaken. Based on it, to address the first problem noted in IIR, the community developed the [Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA Labels \(RZ-LGR Procedure\)](#). On the [direction](#) of the ICANN Board provided on 11 April 2013, ICANN undertook this procedure which requires a two-step process, requiring each community to develop individual script-based Label Generation Rules (LGR) proposals and an expert panel to review and integrate them into the Root Zone LGR (RZ-LGR). The RZ-LGR Procedure proposes that the IDN variant labels of a TLD will be determined using the RZ-LGR. Multiple script communities have [finalized proposals](#), from which Arabic, Ethiopic, Georgian, Khmer, Lao and Thai script proposals have been integrated into its second version, [RZ-LGR-2](#). Many other script communities are also [active](#) in defining their rules. Further, the community indicated that there needs to be a specification to encode these linguistic details into a formal machine-readable format. This format has also been developed and released through IETF as standards track [RFC 7940](#):

[Representing Label Generation Rulesets Using XML](#). A [LGR tool](#) has also been developed to create, use and manage the LGRs, and is available for the community online as well as for download with an open source license.

In summary, the label generation rules for the root zone developed using the using RZ-LGR Process define the set of IDN variant TLD labels, including those which are candidates for allocation.

2.2 Analyzing Variant TLD Management Mechanisms

To address the second part of the requirement stated in IIR, which is to develop a variant management mechanism for the top level, it is necessary for ICANN community to agree on a comprehensive set of procedures to cohesively govern allocation of IDN variant TLDs.

This document starts by identifying the fundamental premises of these mechanisms. These premises arise mostly from observations already made by the community in IIR and the analysis presented by the Security and Stability Advisory Committee (SSAC) in its advisories. Then the document presents a framework for the analysis and derives technical recommendations based on the framework. The recommendations have been designed to be conservative, with the view that the IDN variant TLDs are being implemented for the first time, and that the solution could accommodate implementation experience over time.

The analysis of how these recommendations could be implemented is also conducted. This is done on the current gTLDs and their application process as described in the [Applicant Guidebook](#), and on the current IDN ccTLDs and their application process based on the [Fast Track Process](#). The analysis encompasses discussion on relevant technical details for implementation of these recommendations, the impact on registry services and registry agreements, the possible effect on dispute resolution and trademark protection, and how this may influence the string similarity process. Moreover, the report adds suggestions on how such impact may be addressed, for further consideration of the community. Additional TLD related procedures, which may not be affected, are also reviewed.

A risk analysis has also been conducted and the corresponding mitigation strategy has been drawn, published separately. Based on the risks identified, additional work has been undertaken and included as separate reports in the appendix for the community to consider further. This subsequent work includes the following:

- Analysis of the use Repository Object ID (ROID) to determine same registrant for labels
- Possible mechanisms to reduce the allocation of IDN variant TLDs by further screening the allocatable IDN variant TLDs generated by the RZ-LGR

The remaining report presents the premises, the framework of analysis, the recommendations, their analysis, and further Consideration derived from the recommendations. The report concludes with a brief discussion on next steps.

3 Starting Premises for Implementing Variant TLDs

There are some underlying assumptions or starting premises, which need to be addressed while designing a solution for managing IDN variant TLDs. These are discussed below.

3.1 Variant labels are defined by the script community

As noted by ICANN's Security and Stability Advisory Committee (SSAC)⁴, using the conventions of IIR and the RZ-LGR Procedure, an IDN Variant Label is a "label that is considered the "same" in some measure by a given community of Internet users". This means that the definition of variant label is arbitrary and based on what is defined by the user-community, so cannot be derived algorithmically or technically.

Based on current analysis by the different script based user communities, the variant labels may or may not be visually distinguishable. For example, according to the Arabic script community, شبك would be a variant of the label شبكة even though the two are visually distinct, while the Khmer script community has suggested that strings like ភ្នំ (= ស ៊ ក) and ភ្នំ (= ស ៊ ដ) should be considered variant labels because they are visually identical, even though they are inherently different code point sequences.

3.2 An IDN variant TLD label is a TLD label

If a variant of a TLD label would be delegated in the DNS root zone as a TLD label, once delegated, its status is indistinguishable from any other TLD label in the root zone, including its other variant TLD labels. Therefore, each such variant label of a TLD is a potential TLD in its own right, once it is delegated.

This implies that by default the same policies and procedures are applicable to a variant TLD as to any other TLD, except in the cases where policies and procedures for variant TLDs have been additionally specified.

3.3 Root Zone Label Generation Rules are the single source of variant labels

Because the variant TLDs are defined arbitrarily by the community using a particular script, variant labels can only be determined once this definition has been provided by the community. These rules must exist for all the scripts to be supported for TLDs. Therefore, there is "Need for a comprehensive set of Label Generation Rules (LGR) for the root zone as a prerequisite for implementation of IDN variant TLDs" (IIR). This implies that variant TLDs must be defined using

⁴ See SAC060 report at <https://www.icann.org/en/system/files/files/sac-060-en.pdf>.

this Root Zone LGR (RZ-LGR). For scripts for which such rules are not defined, the variant labels cannot be determined.

Further, there cannot be additional mechanisms which can create different set of variant labels for the same TLD as that can cause confusion about what a variant label is. Therefore, the “root zone must use one and only one set of rules for the Root LGR procedure” ([SAC 060](#)).

3.4 IDN variant TLDs must not adversely impact the Domain Name System (DNS) and promote a good user experience

According to the Integrated Issues report ([IIR](#)), a variant management mechanism should achieve two goals. First, it should “support a workable implementation of variant TLDs” which means implementation must be done in a way that “actual operation and maintenance of the DNS ... not be adversely impacted by the introduction of IDN variant TLDs”.

Second, it should “promote a good user experience”, which means that it should “avoid including variant TLDs in a manner that would create user vulnerabilities or a probability of confusion”.

These goals, based on technical as well as usability needs, must be simultaneously achieved.

3.5 Active IDN variant TLDs should be minimized

It is stated in the Integrated Issues report ([IIR](#)) that “A cautious approach should be adopted; successively more liberal approaches may be adopted later ... The goal should be ... to minimize active entries ... to those where an explicit need has been established, the user experience implications have been fully studied, and no negative impacts to security or stability have been identified”.

Noting that the procedure for “a variant management mechanism could encompass both active use of labels in the DNS, and prevention of labels from use in the DNS” ([IIR](#)), a coherent policy should be developed which can identify and prioritize the needed labels from the allocatable pool of labels generated by the RZ-LGR.

If there are many allocatable variant TLDs available for possible delegation, the ICANN community should devise a process to select which should be activated.

3.6 Blocked variant TLDs should be maximized

The [RZ-LGR Procedure](#) suggests that, based on conservatism principle, “the output of this procedure should aim to maximize the number of blocked variants, and to minimize the number of allocatable variants.” This contributes towards reducing confusability between labels in the DNS. Security and Stability Advisory Committee notes that “Confusability cannot be considered

in isolation from other issues related to security. Phishing and other social engineering attacks based on domain name confusion are a security problem for end users” ([SAC 089](#)). When resolving a domain name, there are two failure modes ([SAC 60](#), [IIR](#)):

- a) **Denial of service:** the user attempts to visit <http://example.Y>, reading it as being the same as <http://example.X> ... but connection does not work because ... example.Y is not registered
- b) **Misconnection:** the user attempts to visit <http://example.Y>, reading it as being the same as <http://example.X> ... but arrives at a site controlled by a registrant different to that of example.X.

SSAC notes that misconnection causes worse results compared to denial of service because misconnection “presents issues of possible credential leakage, accidental disclosure of information, and user confusion and frustration” ([SAC 60](#)) and therefore should be avoided.

3.7 Variant TLDs do not have a single and comprehensive technical implementation

There is no single technical solution agreed by the technical community for implementing variant TLDs. In fact, there is no comprehensive technical solution at all which can address how variant TLDs may be used by the variety of applications. The possible solutions can bundle the variant TLD labels in different ways and to various degrees. It should also be noted that the requirements for the degree of bundling may vary based on the TLD.

Therefore, the technical implementation of variant TLDs may vary across different TLDs, if they are delegated.

4 Dimensions and Degree of Influence

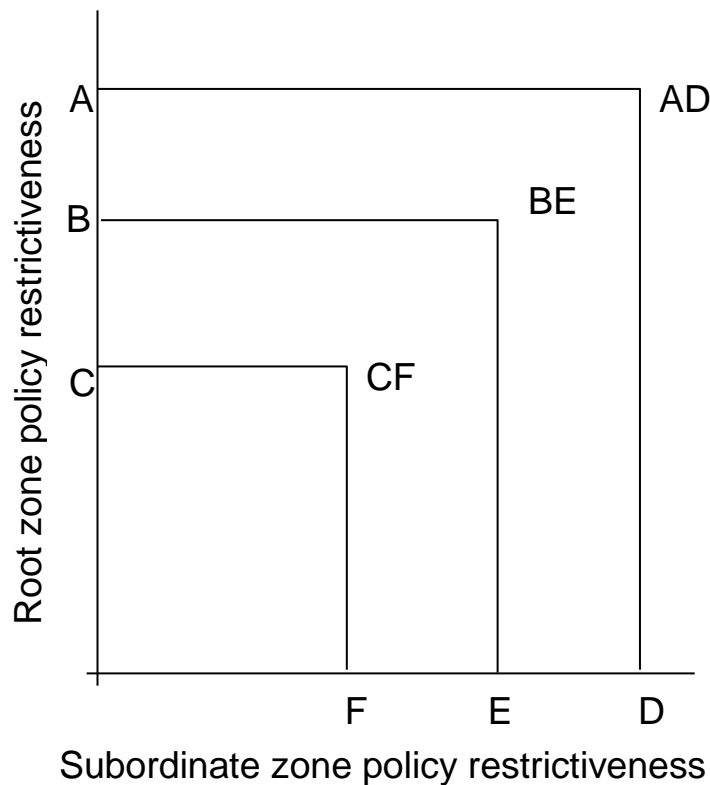
The implementation solution for IDN variant TLDs ICANN organization is proposing for the ICANN community to consider can be theoretically viewed from the perspective of how much influence on such implementation is warranted. The influence with respect to variant TLDs can be considered in two dimensions.

The first dimension concerns policies related to the root zone. In this dimension, the absolute minimum would be no policy at all. The absolute maximum would be full specification of exactly how variant TLDs would work in every way that might impinge upon any user experience.

The second dimension concerns policies related to contractual provisions for names delegated from the root regarding what the operators of those delegated names must undertake in the presence of variants. In this dimension a minimal policy would require no policy at all in

subordinate names. A maximal policy would require a complete specification, applied to all immediately-subordinate names, for how entries in those zones would be permitted.

In practice, neither the minimal nor maximal policies are practical. The minimal policy in both cases effectively says that variants are not an issue at all. IIR has already determined that there is a real user issue in variants, so the minimal stance is unrealistic. Moreover, [Examining the User Experience Implications of Active Variant TLD](#) (henceforth, “User Experience Study”) provided details about such issues for users (see especially Section 5). The maximal position in the first dimension specifies outcomes well beyond the zone operator’s (i.e. ICANN’s) ability to influence the world. And the maximal position in the second dimension would be an attempt to impose rules down the DNS tree from the root -- an attempt that is at once technically and practically infeasible and also beyond ICANN’s remit.



ICANN could nevertheless adopt reasonable positions along these two dimensions. There are practical maxima and minima illustrated above. We can think of policies in the root as being either expansive or minimal (A and C respectively). Similarly, we can think of policies covering subordinate zones as being either more or less restrictive (D and F respectively). These possible positions are outlined below, followed by a specific recommendation.

4.1 Point A: ICANN's responsibility for the root requires a strong policy

From the perspective of Point A, ICANN is responsible (in a policy sense) for the delegations in the root zone. As such, it must take a strong position on inclusion of variant labels in the root zone, and that requires specifying how they must work. This is consonant with (but not implied by) [RFC 6912, Principles for Unicode Code Point Inclusion in Labels in the DNS](#). Moreover, the discussion in Section 5 of IIR makes clear that variant domain names are unlikely to work reliably in a number of cases, and also that their failure modes are unpredictable.

Therefore, under this scenario any acceptable policy for the root zone would be required to enforce the following:

1. Predictable allocation and delegation. Only the LGR for the zone (in this case the root zone) is the source of variant labels: other names could be related in some administrative way, but they are not “variants” for the purposes of policy. It follows from this that, if two strings are not variants under an LGR, then they are not variant labels even if they were previously allocated under a claim that they are variant labels. They must in that case be treated as two separate delegations.
2. Minimal allocation and delegation. This is a principle of conservatism, and requires that ICANN agree to the very smallest set of TLD allocations practically possible. For many TLD labels, this will mean “no delegated variants”. The consequence of this in the short term is probable failed lookups under some entry methods, but this view asserts that the gradual improvement of local mappings in user agents is less destabilizing and better for the Internet than even small numbers of violated user expectations in resolution over the long term.
3. Maximal blocking. According to IIR and the principle of conservatism, most generated variants from the LGR ought to be blocked. If they are not, then as a matter of policy ICANN would (under this approach) allocate no more than some arbitrary and small number of variants.
4. Strict rules on how subordinate delegation must be handled. Every delegation from the root to any variant requires some form of administrative mirroring beneath it (not necessarily technical mirroring – which does not have an available, complete, and deployed mechanism as of this writing). See points C and D below.

4.2 Point C: ICANN's responsibility is limited to minimal administrative collision

Under this view, ICANN is formally responsible for delegation from the root, and it has existing policies for that. The policies are intended to ensure basic operational competence beneath a TLD delegation, to minimize failures, and to be consistent with the principles outlined in RFC 6912. The discussion in Section 5 of IIR makes clear that variants are unlikely to work reliably in

a number of cases, and also that their failure modes are unpredictable. Given the unpredictability of the user experience of variants, it is wisest to allow different mechanisms to be explored by different TLD operators, and to expect that some approaches that some operators use will not always work⁵ in the face of user pressure (including the decisions of software makers, system administrators, and both direct and indirect users of various domain names on the Internet). Therefore, this view claims that the only requirement is to ensure that any allocation of any member of an IDL set⁶ must be to the same entity, and that there are no further requirements.

Therefore, under this scenario, any variant delegation from the root must conform to the following rules, but may follow other rules as the operator sees fit:

1. Every allocation of labels from the root zone must be acceptable under the root zone LGR.
2. In addition, every allocation of any variant TLD must be to the operator of any other variant TLD of the string under the same Root Zone LGR. (This is the “same entity” restriction; see below.) There is no requirement that any allocated label be delegated.
3. Every delegation is treated as a separate delegation according to prevailing delegation rules, but it follows from 1 that allocations must travel as a group, so all redelegations entail new allocations of every variant name where one of the set is redelegated, in case that redelegation is part of a new allocation.
4. Allocation and delegation are subject only to the limits of root zone allocation and delegation in general.
5. Any blocked label under the LGR must not be allocated.
6. ICANN shall take no position on subordinate delegation except for those policies already in place about subordinate delegation.

4.3 Point D: Subordinate policies follow from variant status

Users often expect that two variant labels somehow represent “the same name” (see User Experience Study). Therefore, under this view, names beneath variant TLDs must follow similar principles of allocation and delegation as the TLD (i.e. that variant generation, allocation, and delegation rules are the same as they would be for such labels in the root zone). ICANN does not wish to delegate any names that violate the principle of least surprise, so it will adopt and enforce rules in an effort to minimize such violations. In particular, under this scenario it would be required of delegations beneath the set of variant TLDs that they implement the following policies:

⁵ It is possible to raise the objection that, if variants will not work reliably in any technical sense for users, then variants should not be permitted at all. But under such a regime, the “variants” would all either be withheld (in which case users will still see failures) or else they would simply be independent names from one another (in which case users might experience even greater surprise).

⁶ See glossary.

1. Every TLD operator responsible for a group of variant labels must apply coordinated allocation to labels beneath any variant TLD. Operation affecting allocation of any one such label must automatically affect every other variant.
2. Registries must have enforceable delegation policies among variants such that variants do not yield surprising results to users. This does not require a technical mirroring mechanism, but does require consistent results. Except for temporary differences due to the eventual consistency of the DNS, if the same user resolves a name *s1.t1*⁷ and resolves a name *s1v1.t1*, the experience should be consistent between them. “Consistent experience” is dependent on the use: a web page might (or might not, depending on the target audience) offer identical content, whereas an email address would need to end up at the same mailbox.
3. Any policy that produces a result for one Internationalized Domain Label (IDL) set in a given script must produce similar results for other IDL sets in the same script. For example, suppose *s1.t1*, *s1.t1v1*, *s1v1.t1*, and *s1v1.t1v1* are all variants of one another in script S, and that they produce allocatable variants, and that the label *s2* follows the same patterns. In that case, *s2.t1*, *s2.t1v1*, *s2v1.t1*, and *s2v1.t1v1* must all be allocatable variants too, and in the same way.
4. As an elaboration of (3), the rules under (3) are waived for any case where some variant in the set does not work at all (i.e. that results in additional blocking).
5. Policies outlining this, and that are testable, are a necessary condition for the delegation of any names in a root-LGR-generated set.

4.4 Point F: Subordinate policies are as they are in other TLDs

As “Integrated Issues Report” makes clear in Section 5, it is not possible to produce predictable behavior using the notion of variants. Therefore, under this view, ICANN shall take no position on the variant-handling policies beneath delegations it makes, except to require that any label allocated in one variant of a TLD IDL set must be allocated to the same entity in every other label in the IDL set, and that all allocations and delegations in such cases conform to prevailing relevant ICANN policies (or the non-conforming label should be blocked). That is, under this policy, if *s1.t1* is allocated, then *s1.t1v1* must be allocated to the same entity if it is allocated at all. The registry for *t1* may make additional policies about variants (e.g., of *s1*) under *t1*, but those are not required as part of the delegation of *t1* and its variants (e.g. *t1v1*, *t1v2*, ...). For the purposes of this document, this policy is called the “same label beneath variants rule”.

4.5 Starting with an Intermediate Position

The point “BE” as shown in the figure is an intermediate level of specification, a conservative stance from the minimal positions described by the point “CF”, but not the very restrictive positions indicated by point “AD” in the figure above. This position suggests some restrictions at the TLD level and at the second level, balancing innovation needed for this new area of implementation at the top-level with the conservatism in deploying the new technology due to

⁷ See the Glossary below for the meaning of this convention.

the limited previous experience. The solution may be reviewed over time, as more experience is gained by the community in the implementation of IDN variant TLDs. This report recommends the positions described by point “BE”.

The reasons follow from IIR:

- (a) It is not, in general, possible to provide a technical mechanism that will make variants behave reliably throughout the DNS. In particular, in the absence of a general-purpose algorithmic definition of variants available both for the purposes of lookup and for zone operators, both user and server software would need to implement complicated algorithms that, in a distributed environment, might easily drift;
- (b) Users are adaptable to technology as long as it behaves consistently (see IIR section 1.1 item 4), so it should be possible to permit flexibility to accommodate that fact. Under the recommended position, the same entity is responsible for variant names. Moreover, *ex hypothesi*, there are users who will treat the domain names as being “the same”. Therefore, the entity in question has an interest in ensuring reliable and unsurprising user experience, and should be free to undertake that in whatever way works for the situation;
- (c) ICANN is not a regulator and is not in a position to make unwarranted policies for subordinate zones; and
- (d) The minimal safe action for any variant label is to ensure that it is controlled by the same entity: in that case, the entity in question is in a position to ensure that user expectations are not violated. If user expectations *are* violated, they are violated by the entity in question, which is consistent with the way user expectations are handled in the absence of variant labels.