



Examining the User Experience Implications of Active Variant TLDs

Interim Report

Oct. 2012

Status of this Document

As part of the ICANN IDN Variant Program, this *interim report* examines the user experience implications if variants of internationalized top-level domains are activated. This report will be published for public comment for thirty (30) days (plus 21-day reply period) and is to be followed by the final report, to be published within ninety (90) days after the closure of the public comment forum.

In order to better understand and address the challenges surrounding the activation of variant internationalized top-level domains (IDNs), this *interim report*: 1) summarizes and compares, from a user experience and registry management perspective, variant practices in several ccTLD registries, 2) proposes a set of guiding principles to define an acceptable user experience, and 3) identifies how various user communities (e.g., end users, system/network administrators, application developers, registrants, registrars and registries) will be impacted by active variant top-level domains. Based on community input, the final report will include guidelines and recommendations to address or mitigate the negative impacts identified in this report.

Table of Contents

Table of Contents	3
Executive Summary	5
1. Background and Introduction	7
1.1. Internationalized Domain Names	7
1.2. Study Objectives	8
1.3. Scope of Work	8
1.4. Balancing User Expectations with Consistent and Secure Implementations.....	9
2. Terminology, Assumptions and Limitations	11
3. Existing SLD and TLD Practices	12
3.1. Existing IDN Variants at Second Level.....	12
3.1.1. Arabic IDN Variant Management at Second Level by IDN ccTLDs.....	12
3.1.2. Existing Chinese IDN Variants at Second Level.....	13
3.1.3. Comparison of SLD practices across Arabic and Chinese Scripts	14
3.2. Synchronized IDN TLDs	15
4. Usability Principles for IDN Variants	16
5. User Roles.....	18
5.1. End Users	18
5.2. Users involved in the Registration Process.....	19
5.3. Technical Community.....	20
6. Challenges Related to Active Variant TLDs	20
6.1. Challenges with the Use of Variants.....	21
6.1.1. Users cannot find the complete set of variants for a primary label	21
6.1.2. Variants not intuitive.....	22
6.1.3. Variants defined inconsistently.....	22
6.1.4. Variants displayed inconsistently	22
6.1.5. Variants cannot be input by the user.....	23
6.1.6. Unable to distinguish specific variants	23
6.1.7. Identifier not bound to all variants.....	24
6.1.8. Accessibility and privacy impacted	24
6.1.9. Variants not searchable.....	25
6.1.10. Search rankings unpredictable	25
6.1.11. Search optimization affected by variants	25
6.1.12. Variants not part of URL/URI/IRI.....	26
6.1.13. Variants cause session re-establishment	26

- 6.2. Challenges in the Registration and Management of Variants..... 27
 - 6.2.1. Management across IDN TLDs inconsistent 27
 - 6.2.2. Registration for SLDs across TLDs inconsistent..... 28
 - 6.2.3. Inconsistent association of ASCII and IDN TLDs 28
 - 6.2.4. Software support inadequate..... 29
 - 6.2.5. Registration system not straightforward to localize..... 29
 - 6.2.6. Registration information inconsistent..... 30
 - 6.2.7. Trademark protection tracking complex..... 30
 - 6.2.8. Trademark protection dispute process complex 31
- 6.3. Challenges in the Configuration and Diagnostics of Variants..... 31
 - 6.3.1. Software configuration not supported..... 31
 - 6.3.2. Cannot associate variants for configuration..... 32
 - 6.3.3. Compounded certificate management..... 32
 - 6.3.4. DNSSEC validation inconsistent 33
 - 6.3.5. Log and history searching does not match..... 33
 - 6.3.6. Network traffic statistics incomplete 34
 - 6.3.7. Caching infrastructure inefficient 34
 - 6.3.8. Diagnostic and troubleshooting tools incompatible..... 34
 - 6.3.9. Forensics significantly more complicated 35
- 7. Future Work..... 36
- 8. Authors 37
- 9. Acknowledgements 37
- References..... 38
- Appendix A: Survey for System Administrators 42
- Appendix B: Survey for Registries and Registrars 44

Executive Summary

As part of the process of implementing internationalized top-level domains (IDN TLDs), ICANN is undertaking several projects to determine the requirements for activating their *variants*. This report is the first part of one study to examine potential challenges from a user experience perspective when variants of IDN TLDs are activated.

The identification and management of variants is determined by linguistic and technical factors and different communities may have opposing perspectives. Linguistic communities are primarily concerned with end users, and consequently may advocate for a maximal variant label set to enable a diverse linguistic expression and easier accessibility. The technical community, on the other hand, is primarily concerned with the security and stability of the Domain Name System (DNS), and therefore stipulates the minimal number of variant labels (if any) be added to the root zone.

To find the right balance between these two perspectives, within the context of usability, this report proposes seven guiding principles for activating the variants of IDN TLDs. They include: (i) *Minimality* (variants must introduce only the least changes necessary in the DNS), (ii) *Security* (variants must minimize the risks introduced by IDNs), (iii) *Predictability* (variants should behave and function as users expect in their language and script environments), (iv) *Equivalency* (variants must direct users to related content and be managed by the same entity), (v) *Consistency* (variants should behave similarly within and across TLDs and supporting technology), (vi) *Manageability* (variants should be straightforward to visualize and administer with supporting technology), (vii) *Ease of Use* (variants should be easy to use for new and existing Internet users).

To learn from existing deployment experiences, this report summarizes and compares, from a user experience and registry management perspective, variant practices in seven Arabic and Chinese script ccTLD registries. Analysis shows that these management practices have many similarities. In particular, although the label generation rules for variants differ between Arabic and Chinese TLDs, they all treat the variant label set as an atomic set for all aspects of the domain name life cycle, and they all limit the number of activated variants to ensure a positive user experience while blocking all other variants for security and stability purposes. However, although the Chinese script registries share the same tables and have cohesively defined variants for both the top and second level domains, Arabic script community exhibits some difference with Arabic language and across the languages using Arabic scripts.

Active variants of TLDs may introduce many challenges to various user communities. To identify these challenges, user communities are first grouped into three categories: (i) *End Users*—those who would use the variants, (ii) *Registrants, Registrars and Registries*—those who would manage registration of the variants, and (iii) *Technical Community*—those who would deal with usability, configuration and diagnostics of the variants. Based on these roles, this report identifies some of the challenges which would arise as the variants are activated, and, where possible, keeping them at high level and independent from specific implementation methodology. These challenges are summarized below and explained in more detail in the document.

- Challenges with the Use of Variants
 - User cannot find the complete set of variants for a primary label
 - Variants not intuitive
 - Variants defined inconsistently
 - Variants displayed inconsistently
 - Variants cannot be input by the user
 - Unable to distinguish specific variants
 - Identifier not bound to all variants
 - Accessibility and privacy impacted
 - Variants not searchable
 - Search rankings unpredictable
 - Search optimization affected by variants
 - Variants not part of URL/URI/IRI
 - Variants cause session re-establishment

- Challenges in the Registration Management of Variants
 - Management across IDN TLDs inconsistent
 - Registration for SLDs across TLDs inconsistent
 - Inconsistent association of ASCII and IDN TLDs
 - Technological support inadequate
 - Registration system not straightforward to localize
 - Registration information inconsistent
 - Trademark protection tracking complex
 - Trademark protection dispute process complex

- Challenges in the Configuration and Diagnostics of Variants
 - Software configuration not supported
 - Cannot associate variants for configuration
 - Compounded certificate management
 - DNSSEC validation inconsistent
 - Log and history searching does not match
 - Network traffic statistics incomplete
 - Caching infrastructure inefficient
 - Diagnostic and troubleshooting tools incompatible
 - Forensics significantly more complicated

It should be noted that the proposed guiding principles and the list of identified challenges are not considered comprehensive or final. By publishing this interim report, we seek to establish a dialog with the community to refine these principles and challenges, which will guide the second part of the study to formulate the guidelines and recommendations for activating variant TLDs.

1. Background and Introduction

This interim report is the first part of the study to examine the user experience implications of active variant internationalized top-level domains (IDN TLDs). Sections 1 and 2 provide the reader with a brief introduction to IDNs and a brief description of the objectives and scope of the study as well as terminology used in the report. It also explores the divergent interests and practices around variants to highlight the challenges for finding an acceptable solution for using variants.

Today, many country code top-level domain (ccTLD) registries already support variants of domain names at the second level. To learn from their experiences, Section 3 summarizes and compares variant practices across seven Arabic and Chinese script IDN ccTLD registries.

Section 4 proposes a set of principles to be used for determining the usability of variants TLDs. Sections 5 and 6 explore how various user communities may be impacted by active variant TLDs. In particular, Section 5 explains user roles studied, and Section 6 lists relevant impacts to these user communities if the variants of TLDs are introduced.

Using public feedback to this interim report, the study team will work on developing guidelines and recommendations to address and/or mitigate the negative impacts identified.

1.1. Internationalized Domain Names

There has been a long-term need to deploy domain names in local languages, also called Internationalized Domain Names (IDNs), for making the Internet accessible to populations using non-ASCII scripts. The Internet community has worked together for more than a decade to make this possible.

The Internet Engineering Task Force (IETF) first standardized a set of protocols [9,11,13] to handle Internationalized Domain Names in Applications (IDNA2003). These protocols have been revised based on community feedback [14], resulting in the IDNA 2008 [8,12,15,16,17].

Having some years of experience deploying IDNs at second and lower levels, recent work focuses on deploying IDNs in the root zone (at the top level). The ICANN IDN Fast Track Process [39], an experimental program that allowed certain ccTLDs to apply for IDNs, has enabled 31 IDN ccTLD strings be added to the root zone. Policy development process (PDP) is underway in the Country Code Names Supporting Organization (ccNSO) to replace the Fast Track process and allow for more ccTLDs to apply for IDNs. In addition, IDN TLDs will also be eligible as part of ICANN's new Generic Top-Level Domain Program (new gTLD).

One key issue to be resolved in both the ccNSO IDN policy development process and the new gTLD program is how to consider the inclusion of certain IDN labels called “variants” in the DNS root zone. To address this complex technology and policy issue, ICANN first engaged six script communities to produce a set of reports [1,2,3,5,6,7] that identify requirements for the respective script community. It then produced an integrated report [4] collating and synthesizing these issues associated with the possible inclusion in the DNS root zone of IDN variant TLDs.

Following the recommendations identified in [4] and subsequent project plan [41], this study was initiated to consider the user experience implications for active variant TLDs.

1.2. Study Objectives

The project considers the user experience implications when two or more IDN variant TLDs are activated (i.e., have resource records in the DNS). It seeks to answer the following set of questions:

- What are the components of an acceptable user experience for variant TLDs?
- How will various user roles be impacted if variant TLDs are activated?
- What are the necessary rules or guidelines a TLD should operate under in order to provide an acceptable user experience for variants?
- What are the policy/contractual considerations that will make these rules effective?
- How does the impact of variant TLDs on applications have on user experience?
- What other entities have a critical role to play in addressing these issues and what educational or consultative steps could be implemented to generate support and collaboration by these parties?

The study findings are expected to support two outcomes:

- Recommended rules or guidelines a TLD should operate under to provide an acceptable user experience with regard to variant TLDs, including appropriate policy or contractual provisions to make these rules effective.
- Creation of a useful reference for educating application developers and others affected by these changes.

To answer these questions, this study considers the different user roles identified in the Integrated Issues Report [4] (e.g. end user, registrant, registrar, registry, system administrator, network operator), and the impacts to these users, should variant TLDs be activated.

1.3. Scope of Work

An A label is formed by letters (A-Z, a-z; case insensitive), digits (0-9) and hyphen-minus (LDH) with a string length limited to 63 octets (RFC 1035, RFC 1123, RFC 2181 and RFC 4343). TLD labels are formed using further constraints, which require the labels to be limited to letters and further limited in length (RFC 3696). The IDNA protocol allows labels to be formed with additional characters, beyond the LDH, based on the Unicode standard (as stipulated in IDNA 2008 [8,12,15,16,17], called the U-Label. This U-label is converted to an equivalent ASCII form or A-label following the LDH label constraints (prefixed by “xn--”). As the general TLD labels have a stricter criteria than LDH labels, the IDN TLDs also have additional constraints, as recommended by the technical community, limiting their character set to letters (with Unicode property Lx) without other characters (e.g., marks (with Unicode character property Mx) and CONTEXT O/J characters [42]) unless necessary and requiring the U-Label to be more than one character [43].

IDNs also introduce variants. According to the Integrated Issues Report [4], a variant is a

“term that has been used in multiple ways, to indicate some sort of relationship between two or more labels or names. It has been used variably to refer to, for example, a particular relationship between specific characters or code points in a particular script, or a set of alternate labels where some linkage relationship is articulated, or a desired procedure whereby names are registered in multiples, or a desired functionality causing shared behavior by some set of identifiers.”

For the purpose of this work, variants are considered at label level and not at code point/character level, though it is understood that they may be formed due to variants at code point level (see Section 2 on Terminology).

As discussed, TLDs can be ASCII labels or an Internationalized Domain Labels (IDL). For those TLDs that are IDLs, many do not have variants. The scope of current work is limited to the usability study of those IDN TLDs that *have* variants, as illustrated by the shaded region in Figure 1. The current work also examines the boundary cases for further understanding of usability related challenges, as indicated by the dotted line in the figure.

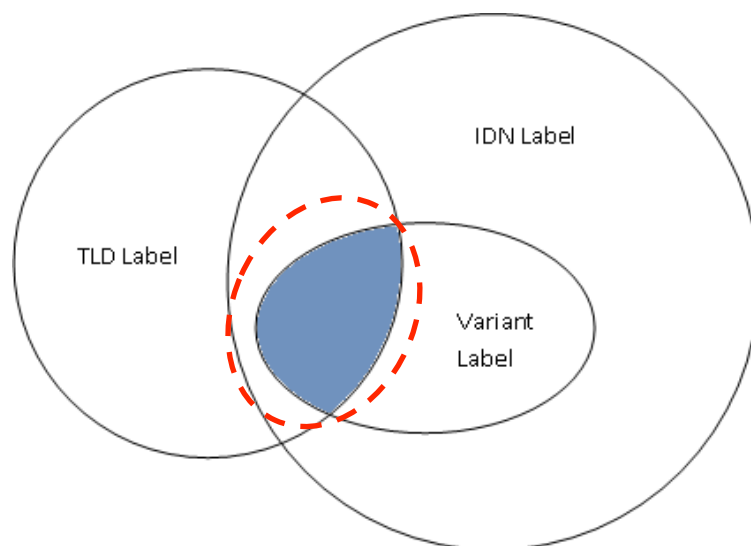


Figure 1. Scope of Study on IDN TLD Variants

1.4. Balancing User Expectations with Consistent and Secure Implementations

The expectations for the use of IDNs and their variants are very different across linguistic and technical communities. Linguistic communities are primarily concerned with serving end users, and require maximally distinct IDN labels from a character repertoire for capturing a diverse linguistic expression and additionally maximal variant label sets for accessibility. The technical community is primarily concerned with the security and stability of the DNS, and requires the minimal number additional labels in the root with minimal (if any) variants.

The current discussions for IDN TLDs and their variants at the root now involve the “script community,” which roughly represents the set of linguistic communities using a certain script. The motivation of defining this community comes from the fact that the root is shared by

everyone [4]. Therefore, though script distinctions are (arguably) possible, the language delineation cannot be represented at the root. This requires that the same set of rules should be applicable to all TLDs supporting the same script¹. This creates a third expectation, which has a more diverse “script” expression (containing a common character repertoire of many more characters than would be in an individual language) and more numerous and possibly a much larger sets of variants (as a larger script-level character repertoire causes more conflicts).

The challenge is to find the right balance between these competing linguistic, script and technical expectations.

Not only are there differing expectations, there can also be very different implementations of IDNs and variants if they are not consistently managed. This can be sub-classified into two aspects. First, what are the various states of a variant (see Section 5 of Integrated Issue Report [4]) and how a variant is assigned to a state (e.g. automatically through a label generation table, or arbitrarily chosen by the end user, etc.)? Second, how the variants which are active (if any), will be implemented. Differences in how variants are assigned different states, and how the activated variants are implemented would cause confusion in both the use and technical implementation. For example, if no common mechanisms are agreed upon, the activation and resolution of variants may result in different user experiences across different gTLDs. The experience across ccTLDs would also remain very distinct, as they serve different linguistic communities. Consistent guidelines can give users a more predictable experience across the TLDs. Further, the implementation may also change down the tree, from TLD to SLD and further below. This could add to further challenges in their configuration and use.

In addition to their implementation, the way variant sets are defined and managed could still vary greatly down the tree, e.g. between TLDs, SLDs because TLDs are defined at the script-level and SLDs may be defined at the language level. This will be one major source of variation in the way variants are enabled for users. This change in definition of variants across various levels could become a source of confusion and may have a significant impact on the usability of variants. If user confusion is to be minimized, this variation will need to be managed.

How the variants are enabled and expected to be used by a script community may also vary across script. For example, variants in Chinese script may be visually distinct but those in Arabic script may be visually similar. Some variants may be determined by language and some arbitrarily based on the implementations. For a better user experience, such script-level differences should at least be documented, and, whenever possible, avoided.

The variations in user experience in using variants of IDN TLDs may not only be caused by their implementation, but also due to the greatly differing interfaces being used. User systems are configured in a variety of ways, including different operating systems, different keyboards, different fonts, different personalized settings, different applications, different locales, etc. Therefore, even if other sources of variations are contained, the range of system configurations – along with varying levels of support for variants – will result in different user experiences.

¹ Single script loosely refers to TLDs defining their repertoire of characters from primarily one script (e.g. Cyrillic, Greek, Arabic; see <http://www.unicode.org/Public/UNIDATA/Blocks.txt>), with the understanding that there may be additional characters in it with other Unicode script properties (e.g. Common and Inherited). As per Unicode, “In text processing, the identities of all characters are normally known, but some characters may be shared across scripts or attached to any character, thus requiring special values for Common and Inherited” (Davis and Whistler (2012)).

The final source of variation in user experience is dependent upon users themselves. The competence of users with the script and technology they are using can greatly vary the user experience. Insufficient knowledge of the script or insufficient grasp of technology being used can significantly degrade the user experience. As the IDNs are rolled out, reaching the unreached, most of the new end-users will be technologically challenged.

The current work studies the different facets of expectations and variations to determine how and to what extent IDN TLD variants may be introduced to balance between these varied needs and constraints.

2. Terminology, Assumptions and Limitations

The terminology used in this document follows the conventions recommended by the community and as documented/defined by [4], particularly in “Appendix 2: Terminology.” Readers should understand the definitions for *Variant Label*, *Label Generation Rules*, *Fundamental Label*, *IDL*, *IDL Set*, *U-Label*, *A-Label* and other terms used in this document. Readers are also encouraged to consult RFC 6365: *Terminology Used in Internationalization in the IETF* for more general but relevant definitions.

For the purposes of this study, we use the term *variants* to strictly mean *alternate labels* or *variant labels*. These IDN labels or Internationalized Domain Labels (IDLs) are also cumulatively referred to as an IDL set (as defined by [4]). The specific characters or code points which cause these variants will be referred to as *variant code points* or *variant characters* but not as *variants*.

The notion of *primary label* is used to refer to one of the labels in an IDL set which is either preferred by a registrant or is pre-determined by a community to represent an IDL set. Labels are used at multiple levels: top-level domain label (TLD), second-level domain label (SLD or 2LD) and third-level domain label (3LD). Multiple level labels combine together to form a domain name; for example, with *www.icann.org* “icann” is SLD and “org” is TLD.

There are ongoing discussions on how variant labels will be implemented in the root zone. Though much of this document may remain relevant irrespective of the method used to implement variants, the final implementation could still have impact on this study, while possibly introducing additional issues not discussed in this document.

The integrated issues report [4], based on discussions in various case study teams (for Arabic, Chinese, Cyrillic, Devanagari, Greek and Latin [1,2,3,5,6,7]), identifies multiple states and transitions possible for variant labels in Section 5: “Discussion of Issues: Treatment of Variant Labels.” This life cycle for variants has to be finalized by the community, and the decisions eventually taken may also have impact on the findings of this report, because once the states have been defined, all stake holders will need to understand the states and transition process between the and then enable processes to manage them. For the purpose of this study, it is assumed that there will at least be two logical states, activated and not-activated.

3. Existing SLD and TLD Practices

Currently, IDN variants are already in practice at second and lower levels in the Domain Name System (DNS). Therefore, it would be valuable to learn from how the relevant end users are adopting them. Though the TLDs are different, the state of practice at lower levels can still provide some insight in understanding user expectations. Further, in the case of Chinese script, simplified and traditional Chinese IDN ccTLDs that are managed synchronously have been implemented to provide users with an experience similar to IDN TLD variants. Learning from their experience will be helpful in the context of this study. The current section gives an overview of these practices for the ccTLDs.

3.1. Existing IDN Variants at Second Level

3.1.1. Arabic IDN Variant Management at Second Level by IDN ccTLDs

Arabic script is used for a variety of languages globally. There has been a lot of work on analyzing variant labels as it relates to Arabic script and the languages that use this script. This has included work by the Arabic Script IDN Working Group [21], the Arab League, UN ESCWA, and other national and international efforts. RFC 5564: *Linguistic Guidelines for the Use of the Arabic Language in Internet Domains* [10], published in 2010, discusses in detail the source of user confusion in using Arabic script specifically for the Arabic language. A more comprehensive analysis of Arabic variants at the script level is presented in the *Arabic Script Issues Report* by the Arabic Case Study Team [3].

These guidelines have been largely adopted for the use of Arabic language IDNs by various ccTLD registries in Arabic-speaking countries. For example, dotEmarat (امارات) IDN ccTLD for United Arab Emirates has an extensive policy to manage variants [29] summarized as follows:

- Registrant applies for a domain name in Arabic language, which is then considered *primary*
- Variants are created due to digits which are semantically same and letters which are culturally confused due to local conventions
- Variant due to digits are automatically activated
- Variants due to writing conventions are blocked by default, but may be activated by the request of the registrant
- Blocked variants are not available for other registrants
- Variants can be activated at any time through the registrar
- A registrant may activate up to five variants at a time
- All variants are associated as a set to the *primary* and may not be separated

In addition, the variants are also considered a single group for registration data and a single registration data record is maintained for all the variants. When an active variant is looked up, the registration information of the *primary* label is returned. Thus, every time the status of a variant is updated, the registration data entry of the IDL set is also updated. Further, when preparing the reserved list of domain names, the registry also blocks/reserves all the variants of these domain names.

Similar policies are adopted for other Arabic-language IDN ccTLDs, with some variation. For example, Qatar IDN ccTLD considers the characters *ﻯ* and *ﻯﺎ* as variants in addition to other characters considered variants by dotEmarat. Jordan IDN ccTLD allows up to three variants to be activated.

The IDN ccTLD for Saudi Arabia offers position-level variants for each character [22]. Though the technical solution is more complex, this enables them to offer many more labels to their potential registrants. The IDN policy [25] regulates the labels. The policy also states that it “may establish rules and procedures to resolve the problem of character variants with other Arabic script based languages (e.g., Persian, Urdu).” As explained in supporting documentation [26], this is to enable resolution of domain name variants generated from keyboards of other languages.

IDN ccTLD for Iran also offers up to five variant labels for (automatic) registration in addition to the primary domain name requested by the registrant in Farsi or Persian language. The variants are produced by a set of rules defined by the policy [24] for the ccTLD. These variants may also use Zero-Width-Non-Joiner in Farsi labels, which are not allowed in Arabic-language labels.

These TLDs are already offering variant registrations at SLD, except SaudiNIC, which intends to offer variant resolution if and when the variants are enabled at TLD. None of these Arabic-script based IDN ccTLDs offer IDN SLDs against the ASCII TLD.

3.1.2. Existing Chinese IDN Variants at Second Level

Chinese variants are defined as “characters with different visual forms but with the same pronunciations and with the same meanings as the corresponding official forms in the given language contexts [5].” Because of the ideographic nature of the Han script, and the evolution of the writing system, Chinese characters have many variants. The Chinese Domain Name (CDN) community has been working together for the past decade to solve this problem.

At the second level, the variant management system for CDN registries has three key components:

- *A Language Variant Table (LVT)* that defines the variants for each character that can be permitted as Chinese domain names. The CDN ccTLDs and gTLDs all use the language table developed by the Chinese Domain Name Consortium (CDNC).
- *Registration Policy* that is specified by RFC 3743 [18] and RFC 4713 [19]. RFC 3743 defines a set of IDN registration and administration guidelines for applying restrictions to CJK scripts and the zones that use these scripts. RFC 4713 describes how “.CN” and “.TW” apply the principles of RFC 3743 to manage Simplified Chinese and Traditional Chinese domain name equivalence.
- *Registry/Registrar Provisioning Systems* that perform functions such as reading the IDN Table, generating variant domain name, transcoding to punycode, and provisioning the zone file if synchronization is used.

In general, the Chinese Domain Name registries exhibit the following practices at the second level. The primary label and all the variants generated form an atomic package (called an Internationalized Domain Label package – or IDL package – in RFC 3743 [18]). Once an IDL package is created, the name holders can dynamically request the domain registry to activate and

deactivate some variant IDLs in the package. However, no IDLs can be inserted into or removed from the IDL package during its lifetime. When the IDL package is destroyed, due to either being unregistered or revoked, all IDLs in the package are available again to all name holders at the same time.

Because the number of variants for the Chinese domain can be extensive, the IDL package can expand and become quite large. For example the IDL package for the label 臺灣大學 (National Taiwan University, U+81FA U+7063 U+5927 U+5B78) contains 30 labels. Thus, if an IDN that contains 臺灣大學 as SLD as well as TLD could have 900 variants. This would create undue burden if all of them were activated. To reduce the complexity, RFC 4713 only allows a label with all simplified characters, a label with all traditional characters, and one additional label indicated by the user to be activated. Thus, 臺灣大學 IDN would only have 2-3 activated variants (as compared to 900 variants, most of which are not needed by the user community).

TWNIC has been offering Chinese SLDs (as a solution) under .tw since 2003. However, to carefully manage the DNAME mapping the registry requires the registrars to provide DNS hosting and management. As reported, they have managed around 200,000 Chinese .tw SLDs so far [27]. Due to its long history of offering domain names in Chinese under the .tw ASCII TLD, TWNIC states in its policy document [28] that it synchronizes the SLDs for not only the two (simplified and traditional) Chinese TLDs but also for the ASCII TLD.

3.1.3. Comparison of SLD practices across Arabic and Chinese Scripts

Looking at the practices between and across the Arabic and Chinese SLD practices, one finds similarities as well as differences.

- Both Arabic and Chinese script registries pre-define a fixed set of variants in an IDL set. This is atomic for all aspects of operation, including registration data information.
- As this IDL set may be large, the registries set limits on how many variants may be activated. However, the registries differ in the mechanisms and number of variants allowed. Chinese registries allow for all-simplified, all-traditional and one additional user-defined variant to be activated, limiting the total to three variants. Arabic registries allow the user to determine the primary and other registrations based on the language table, and do not pre-determine what will be activated. They limit the user to a fixed number of active variants as well (five for Emirates and three for Qatar IDN ccTLDs).
- Though the Chinese-script registries share the same table and have cohesively defined variants, the Arabic script community exhibits many differences within the Arabic language and across the languages using the Arabic script. For example, even though all Arabic-language registries use the same reference language table (RFC 5564), they implement it with slight variations. Jordan has extra variants for ﺝ which are not practiced by others. Furthermore, SaudiNIC intends to implement positional variants, whereas others are implementing character-level variants. Iran IDN ccTLD offers a different table as it is based on Farsi (and not Arabic) language.
- None of the Arabic ccTLDs offer IDNs with the ASCII TLD, whereas Taiwan has not only been offering them for a decade, but also synchronizes the IDN SLD across the IDN and ASCII TLDs.

- The IDN TLDs are using internal custom-built solutions to manage the registration process for IDNs. These systems differ by registry.
- The registries may also impose additional constraints to manage variants properly. For example TWNIC requires the registrars to manage DNS hosting and operations for the registrant. This solves some challenges which are also faced by Arabic SLD registrants in configuring their system. Some Arabic registries provide configuration support to their registrants in this context.

3.2. Simplified and Traditional Chinese IDN TLDs

In 2010, ICANN approved the proposed delegation of .中国/ 中國 top-level domains to China Internet Network Information Center (CNNIC) and the proposed delegation of .台灣/台湾 top-level domains to Taiwan Network Information Center (TWNIC). These pairs of TLDs are managed synchronously [37,38], in which the contents of the zones that the TLD operates are synchronized. In this section, we share some statistics and lessons learned.

Both .中国/ 中國 and .台灣/台湾 follow registration algorithms described in RFC 3743 and RFC 4713 using the master variant table produced by the Chinese Domain Name Consortium (CDNC). As there is currently no standard EPP extension to handle variants, they use proprietary extensions and supply client software to their registrars who wish to generate variants themselves [23]. Table 1 below lists registration statistics for these TLDs.

Table 1: Registration Statistics for synchronized IDN ccTLDs

	Number of Registrations (provided by registry)	% of Domains with variant forms (provided by registry)
.中国/.中國 (dot China)	320,000	77%
.台湾/.台灣 (dot Taiwan)	43,000	93%

In CNNIC, the delegated variant pairs (SC and TC) resolve to the same nameserver. They employ a practice called “parallel provisioning,” in which the SC and TC label is delegated separately, but the contents of the delegated zones are maintained together and from the same backend database. Updates to one also update the other. In TWNIC, the delegated variant pairs (SC and TC) also resolve to the same name server. DNAME is used. More information can be found in TWNIC’s policy document [27].

To better understand traffic patterns of IDN ccTLDs and IDN synchronized ccTLDs, a measurement exercise was carried out by ICANN. The exercise aimed to measure the relative incidence at L-Root of queries for names under two-character ASCII ccTLDs and for names under their corresponding IDN TLDs. The measurement was conducted from 2012-07-19 1410 UTC to 2012-09-21 1458 UTC and data was collected from all available pcap archives on all active (218) L-Root nodes.

The ratio of DNS queries seen for the two IDN synchronized TLDs are presented in Table 2 below. In particular for dot China, the percentage is calculated as the number of queries to Traditional Chinese TLD .中國/ Total number of queries to IDN ccTLD .中国/中國. For dot Taiwan, the percentage is calculated as the number of queries to Simplified Chinese TLD .台湾 / total number of queries to IDN ccTLD .台灣/台湾.

Table 2: Query Statistics for synchronized IDN ccTLDs

	% of DNS Queries for domains in synchronized TLDs (measured by ICANN on L-root servers from 2012-07-19 to 2012-09-21)
.中国/中國 (dot China)	12.2%
.台湾/台灣 (dot Taiwan)	18.4%

A few lessons can be drawn from synchronized IDN ccTLDs:

- The synchronized TLD query statistics show that Chinese IDN variants are being actively used. Given the fact that Chinese IDN ccTLDs tailor specifically to a linguistic community (e.g. Simplified Chinese community) and that IDN gTLDs tailor to a much more global audience, it is likely that variants for IDN gTLDs would receive more queries.
- RFC 4713 reduces the variant problem for IDN.IDNs from multiplicative to linear (see 3.1.2), thus making this a much simpler problem to manage for the registries.
- TLDs and SLDs variants are generated and managed in a cohesive, consistent, and predictable manner, from the perspective of the end user.

4. Usability Principles for IDN Variants

The following principles should guide policy and implementation of how IDN variants and primary labels are allocated, activated, and managed.

Principle 1: Minimality

Variants must be implemented with the least changes necessary in the DNS. Given the inherent complexity of supporting and using variant labels, a conservative approach to adding variants is advisable. Where the case is not convincing, the variant should not be activated. Even where need is evident, a variant may still not be activated if it poses a security challenge (see Security Principle). Appropriate criteria are needed to evaluate variants for activation.

For example, an Arabic IDN TLD may have more than a dozen variant labels. Activating all of these variants may not be in the best interest of the registry, registrants, or users. Active variants may be limited to those which are likely to be used, such as variants that can be

typed either by a Farsi keyboard or an Arabic keyboard and do not require toggling between the two keyboards to input a single label.

Principle 2: Security

Variants must minimize the risks introduced by IDNs. Variant labels are allocated to minimize security risks due to independent delegation of labels considered similar by a community based on visual and semantic reasons. However, the activation of variant labels should not pose additional security risks.

Principle 3: Equivalency

Variants must direct users to related content and be managed by the same entity. A variant label should have an equivalent relationship with the primary label. In other words, users expect variant labels to resolve to content that is the same or similar to the content linked with the primary label.

For example, a variant TLD label could take web users to the same web site as the primary label, or take users to content that has been localized based on the specific variant being used (e.g., a Traditional Chinese variant may direct users to a Traditional Chinese site whereas a Simplified Chinese variant directs users to a Simplified Chinese site). Furthermore, users would expect that this similar content be managed by the same entity, with similar expectations regarding security.

Principle 4: Predictability

Variants should behave and function as users expect in their language and script environments. Based on experience, users have expectations for how their languages and scripts are supported by technology. The behavior of variant labels should be consistent with these general expectations.

For example, Chinese users may expect a domain in Simplified Chinese characters to be equivalent with a domain in Traditional Chinese characters, even though they are visually distinct, while Cyrillic users may not share the same expectations between visually distinct characters.

Principle 5: Manageability

Variants should be straightforward to visualize and administer with supporting technology. This principle applies to those who are required to register, administer, and manage primary and variant labels. The tools and processes available to these users should support all or any subset of (active or non-active) variant labels with equal ease. The relationship between primary and variant labels should be as transparent as necessary for the intended user group and scenario.

Registrants should have a clear understanding of the variants related to a registered domain. If the registry sets a limit on how many variant labels can resolve to a primary label, registrants should have the ability to understand and manage which labels will be active.

Software tools and services should be made available to enable the administration of IDL set for various monitoring and management functions.

Principle 6: Consistency

Variants should behave similarly within and across TLDs and supporting technology. Users view domain names holistically (not as individual labels but as a complete unit), and thus expect ASCII TLDs to behave consistently across labels and geographies. Users would expect similar consistency in IDN TLDs and variants. Therefore, they would expect variant delegation rules for an IDN ccTLD to be similar to the IDN gTLD delegation rules for a given script. They would also expect the second-level variants to be consistent with the variants at the top level. At a more general level, users and application developers will expect a consistent framework for handling IDN variants not just within scripts but also across scripts.

Principle 7: Ease-of-Use

Variants should be easy to understand and use for new and existing Internet users. Appropriate interface mechanisms are needed to view and use variants by different users. Active variant labels should require no additional configuration or software to function as well as primary labels. All active variant labels should only use characters that are well supported by technology, across existing software and operating systems.

For example, software interfaces should be able to distinguish visually same variants, and should be able to relate visually distinct variants, in such a way that users find it intuitive and easy to use.

5. User Roles

Different users interact with systems and use domain names in different contexts. Though there multiple ways the users can be identified, for this study they are grouped into three categories: (i) *End Users*—those who use the variants, (ii) *Registrants, Registrars and Registries*—those who manage registration of the variants, and (iii) the *Technical Community*—those who deal with usability, configuration and diagnostics of the variants.

5.1. End Users

End users use domain names for a variety of commonly understood functions. These include web browsing, emailing, desktop publishing, file transferring, etc. They access these functions using a variety of devices, user input/output methods software system configurations, software applications, and networks. This variety makes this user role very challenging to assess.

End users may or may not be familiar with the script being used, though the users who are not familiar with the script are less likely to use the IDNs or their variants. Also, end users may not be familiar or comfortable with the use of technology or commonly accepted usability practices, particularly those using the IDNs and variants, as they may be new to the Internet and DNS.

5.2. Users involved in the Registration Process

This group includes users involved in registering and managing the domain name life cycle. They include at least the following types of users:

Registrants

Registrants interact through the registrar interface to register, update, renew or delete a domain name. In addition, the registrants also update domain name registration data and may check for confusability of a domain name (e.g. for preventive registration to protect a trademark, etc.). Registrants may also use indirect mechanisms through proxy or privacy services (identified in the Registrar Accreditation Agreement) to perform these functions. Registrants may or may not be experienced in registration of ASCII or IDN domains, let alone variants.

Registrars

Registrars provide services to a variety of clients, including end users, registrants and registries. Services include registering domain names, billing for registrations, handling domain name disputes, and maintaining registrant information. Registrars also escrow registrant data, and interface with ICANN for compliance functions. Within the registrars, there may be different roles, including those involved in policy, marketing and sales and those involved in enabling the back-end systems. As the latter are the same as the technical community, the registrar role primarily focuses on the former set of users, for the purpose of this study. Registrars may have varied degree of expertise with IDNs and variants, and may or may not have personnel and/or tools to manage IDNs and their variants.

Registries

Registries interact with many different users providing a variety of services. They provide an EPP interface to registrars, host the DNS Server and Registration data server, and conduct performance and security analytics. The registries interface with ICANN for payments and performance. The introduction of variants at the TLD may impact one or more of the functions they perform. Registries include ccTLD, gTLD, IDN ccTLD and IDN gTLD. For the purposes of this study, only registries dealing with IDNs are relevant. The gTLDs and ccTLDs are different in multiple ways. For example, gTLDs are likely to offer script-level generic variants, while ccTLDs are likely to be more language specific (which would result in larger SLD to TLD level differences in what labels can be variants for ccTLDs). Further, policies on how to handle IDNs and variants are likely to be more similar across gTLDs, as will be defined by ICANN, while ccTLDs develop independent policies. Staff expertise with various scripts at gTLDs and ccTLDs may also differ. However, in terms of operations, the roles of the organizations may still significantly overlap. IDNTLDs may have staff performing various roles, including those working on policy, sales and marketing, registration review, etc., and those working on system development and testing, operation maintenance, technical support (for registrars/hosting companies), etc. Again, as the latter are the same as the technical community, for the purpose of this study, the registry role primarily focuses on the former set of users.

5.3. Technical Community

The technical community includes technology professionals engaged in developing, configuring and maintaining systems which handle variants. Their work may involve domain names across multiple scripts. Technical community members may be involved in providing front-end services, which require interfacing with end users or registrants, or back-end services. Those providing front-end services may have to troubleshoot domain names as U-labels and so would need to be more aware of relevant scripts, whereas those providing back-end services are more likely to use A-labels.

The technical community comprises many roles, all of which are critical in the functioning of the DNS and will continue to play a central role in the deployment of IDN variants. These include, at a minimum, system administrators, network managers, security managers, and application developers.

In many organizations, the boundary of system, network and security administrators may not be well defined. The following roles are defined for this study.

System Administrators

For the purpose of this study, a system administrator manages at least the following: configuring and provisioning computers, operating systems, and monitoring services; performance management; management of user accounts and identifiers; and management of certificates.

Network Managers

For the purpose of this study, a network manager accomplishes the following: configuring and provisioning routers, switches and network applications; monitoring network resources; and managing performance.

Security Managers

For the purpose of this study, a security administrator accomplishes the following: configuring and provisioning firewalls, VPNs and security policies on various network resources and servers, and monitoring security and logs. A law enforcement agency performs similar tasks albeit with a broader scope.

Application Developers

For the purpose of this study, we assign a broad meaning to application developers to include developers of operating systems, libraries, desktop applications, mobile applications, web applications, frameworks, etc. in any programming language. Application developers need to make applications aware of variants, where relevant, for a variety of users.

6. Challenges Related to Active Variant TLDs

This section details challenges that have been identified thus far. By no means, is this intended to be a complete list. The challenges have been grouped into three broad categories, focusing on (i) *use*, (ii) *registration management*, and (iii) *configuration and diagnostics* of variants.

Where possible, the challenges have been kept generic, applicable across various scenarios and implementations. As the list of challenges is presented, each challenge is first summarized. After the summary, a more detailed discussion is given for each challenge, followed by one or two illustrative examples. These examples are presented to help understand the challenge, but should not be interpreted as restricting its scope.

6.1. Challenges with the Use of Variants

The challenges presented in this section are primarily concerned with the end user. Some end users may be able to handle variants more proficiently. However, the current challenges are listed in the context of novice users, though familiar with the script. These challenges are equally relevant for application developers who try to make the interface friendly for end users. Variants of IDNs are a new concept and will surely challenge application developers to find mechanisms to make them easily and securely available to end users. Many of the challenges are generic enough to be applicable across many other users who will be interacting with variants, including registrants, sales and marketing staff with registrars and registries and even *power users*. For example, the first challenge, related to finding the complete set of variants, universally impacts all users.

The challenges can be logically sub-grouped into those related to the interface (input and output) and those related to their processing (web search, etc.).

6.1.1. Users cannot find the complete set of variants for a primary label

There is no single place from which users can determine the complete set of variants for a domain name, including TLDs.

While most end users may have little or no interest in understanding the full set of variants for a given primary labels, registration managers and technical community will need easy access to these sets. This is a challenging requirement to meet, as the variants may change depending on registry, level of domain label, and over time. Some users may also be interested in the history of the variants, in addition to their current status.

For example, users may need to view the list of variants to understand how variants may be generated in their script; registrants may need to know if and why two labels are equivalent, and view the list of variants to determine the possible registrations they may undertake. Software developers may need this information to enhance the capacity of the search engines. Trademark protection agents may need to have access to all possible variants for investigating possible dispute cases. Registrars may need to know this information for effective sales. Network configuration and security personnel may need this information for configuration, diagnostics, and forensic analysis.

Applicable Usability Principles:

- P2. Security
- P6. Manageability
- P7. Ease of Use

6.1.2. Variants not intuitive

Though the use of labels is generally interpreted by an end user within the context of a language, current discussions suggest that variants for TLDs will be defined in terms of scripts. This may impact user expectations.

Two code points which are considered distinct in a language may be considered variants for TLDs due to restrictions from other languages using the same script. For TLDs, and for many SLDs, to accommodate a more global audience, registries may base variants on common denominators of the script.

For example, in Arabic script U+06A9 and U+06AA are considered distinct in the Sindhi language, but may be considered as variants as these are stylistically different characters in other languages (e.g. Urdu and Farsi) (Arabic Case Study Team, 2011). This would mean that for Sindhi speakers, two distinct labels may be considered variants, which is not intuitive for users.

Applicable Usability Principles:

- P2. Security
- P3. Predictability
- P5. Consistency
- P7. Ease of Use

6.1.3. Variants defined inconsistently

End users may find variants to be inconsistent in how they are defined within a script across both TLDs and SLDs.

Users may assume that variants in a TLD remain the same when viewing an SLD, which may not be the case, leading to potential security issues. These differences could also arise in implementations of gTLDs and ccTLDs.

For example, for a Sindhi-language based SLD, the use of U+06A9 and U+06AA may result in two different labels, but as they are considered variants at the TLD, end users may also consider them as variants at the second level.

Applicable Usability Principles:

- P2. Security
- P3. Predictability
- P5. Consistency
- P7. Ease of Use

6.1.4. Variants displayed inconsistently

Variant TLD labels consist of different Unicode code points. These code points may not all be displayed properly by the technology at hand.

This inconsistency of support may be caused by many different factors, such as different software configurations (including legacy systems). This includes different operating systems, locale settings, fonts, rendering engines, and applications, to view the domain names. Further,

some of the recently encoded Unicode code points may not be supported by technology. Limitations in processing and storage capacity (e.g. on mobile platforms) may further aggravate the challenges.

For example, an Internet café in Iran (Farsi locale) may give a very different view of variants compared with an Internet café in Dubai (Arabic locale) because the computers may be configured with different operating systems, fonts and keyboards. So a user may not be able to properly view a URL embedded within a website.

Applicable Usability Principles:

- P3. Predictability
- P5. Consistency
- P6. Manageability
- P7. Ease of Use

6.1.5. Variants cannot be input by the user

As variant TLD labels may consist of different Unicode code points, it may not be possible to input a variant due to limitations/configurations of the technology at hand.

Due to variation in configuration of the user system, especially in the input method and/or keyboard available, the user may not be able to input all of the variants. Further, it may not be possible to input recently encoded Unicode code points, as they may not be visible and/or available through the input methods. Constraints on mobile technology may also restrict input methods due to processing and space limitations.

For example, a user who has an input method set for Simplified Chinese may not be able to input Traditional Chinese label, and vice versa. So a user may not be able to access relevant content.

Applicable Usability Principles:

- P2. Security
- P3. Predictability
- P6. Manageability
- P7. Ease of Use

6.1.6. Unable to distinguish specific variants

There may be instances where users may need to identify a specific variant, such as to log into a system. However, in many cases, the variants may be visually very similar or the same (when displayed as U-label) and A-labels may be mnemonically intractable. Thus, an end user will not be able to determine the specific variant needed in a specific case and will have problems in using such systems.

A user may need to identify a specific (e.g. primary) variant of a variant set. In many cases this may not be feasible because variants may be visually similar or exactly the same when displayed as U-labels. As the A-labels are not good mnemonics, the user will not be able to distinguish the variants from each other.

For example, a user may input an email address for signing in to an e-commerce website. The user would need to remember the specific variant to login.

Applicable Usability Principles:

- P2. Security
- P3. Predictability
- P6. Manageability
- P7. Ease of Use

6.1.7. Identifier not bound to all variants

Many websites ask users to identify themselves through their email addresses or another identifier which includes the domain name. If variants are introduced, users may be able to input different variants of the domain name (knowingly and, in most cases, unknowingly as these variants maybe visually identical). Thus, users may experience systems that do not work as expected.

For example, a person traveling in a different locale (and using a different keyboard) may not be able to sign in to an online email service which uses the complete email address (including the domain name) as the username.

Applicable Usability Principles:

- P2. Security
- P3. Predictability
- P6. Manageability
- P7. Ease of Use

6.1.8. Accessibility and privacy impacted

Many applications manage and log domain names to facilitate privacy and usability. These applications may not perform as expected with variants.

For example, the auto-complete functionality in web-browsers uses the history of the user to facilitate typing a domain name. History management interfaces allow the users to delete certain domain names they have accessed for privacy reasons. These applications may not be able to effectively collate the variants.

If a user deletes certain domains from history, the variants of the domain name accessed by the user (which may have been accessed without explicit user knowledge, e.g. by clicking on a visually same variant) may not be deleted. This will negatively impact the privacy of the user.

Applicable Usability Principles:

- P2. Security
- P3. Predictability
- P6. Manageability
- P7. Ease of Use

6.1.9. Variants not searchable

Search techniques may not consider variant domain labels as related, and thus do not find relevant web pages against a user query.

As variants introduce a new concept, they are not considered equivalent for search at this time. The situation is further complicated as the variant set for a domain name will require using multiple variant definitions at various levels concurrently. Thus it will be very difficult for applications to search variant domain names against a user query.

For example, if a user searches for variant3LD.variant2LD.variantTLD, the user could be searching for a particular variant combination (based on the system configuration of the user). However, the search engine is not able to divert the search towards the website the user is looking for as it does not know all the possible variants or is not configured to point to all of them.

Applicable Usability Principles:

- P2. Security
- P3. Predictability
- P4. Equivalency
- P6. Manageability
- P7. Ease of Use

6.1.10. Search rankings unpredictable

Users may not find the content they are looking for because search techniques may not rank variant domain labels equally.

Even if search engines are able to link two different web pages as being indicated by variants of the same TLD (with a variant combination of lower levels as well), they may be ranked very differently instead of being ranked “together” as users may expect.

For example, if a user searches for variant3LD.variant2LD.variantTLD, this user could be searching for a particular variant combination. However, even though the search engine locates the variants, the search results for pages indicated by the variants may be ranked very differently.

Applicable Usability Principles:

- P2. Security
- P3. Predictability
- P4. Equivalency
- P5. Consistency
- P7. Ease of Use

6.1.11. Search optimization affected by variants

Pointing multiple variants towards the same website may lower the website’s ranking due to search engine algorithms.

As most web frameworks do not support variants or multiple domains, software developers may use the HTTP “Move permanently” response to redirect multiple domains to the same web page. This may significantly increase the number of redirects, lowering the search engine ranking.

Applicable Usability Principles:

- P1. Minimality
- P3. Predictability
- P4. Equivalency
- P5. Consistency
- P7. Ease of Use

6.1.12. Variants not part of URL/URI/IRI

The concept of variants is only being implemented as part of domain name labels, but users may expect similar equivalence in resource identifiers. This will create inconsistent behavior for users.

Labels and domain names are re-used in many places in identifiers, not just in the domain name part. Users who use a label at multiple places, may expect similar behavior. This is a significant requirement creep (beyond variants in domain names) as it has significant implications on not just domains but on filing systems, etc.

For example, users may expect that if www.SLD.TLDvariant1 and www.SLD.TLDvariant2 work, then email addresses TLDvariant1@SLD.TLDvariant1 and TLDvariant2@SLD.TLDvariant2 should also work. However, TLDvariant1 and TLDvariant2 before the @ sign in the email may not be equivalent. Similarly, the following is not be considered equivalent in the URL <http://www.SLD.TLDvariant1/prg?t=TLDvariant2>. This will confuse the user who is not able to distinguish the two uses of the labels.

Applicable Usability Principles:

- P1. Minimality
- P3. Predictability
- P4. Equivalency
- P5. Consistency
- P6. Manageability
- P7. Ease of Use

6.1.13. Variants cause session re-establishment

Web sessions may need to be re-established if the same website is accessed through a different variant. This may confuse end users, especially when the variants are the same or similar looking.

Mechanisms for web session establishment, either authenticated or unauthenticated, e.g. through cookies, use domain names as the entry key. Since a variant is a different domain name, the session may be considered different by the mechanism and user will need to re-establish the session.

Users may consider variants as the same, especially if the variants look the same and/or they point to the same website. Thus, if users log into a website and save the login information, they would expect to be logged in automatically the next time they access it. If the website is accessed by clicking on a link (which is a variant), re-authentication will be needed and an additional (separate) copy of the information will be saved. Further data entered on the website may not be available or may be lost across these (redundant) sessions.

Applicable Usability Principles:

- P3. Predictability
- P4. Equivalency
- P7. Ease of Use

6.2. Challenges in the Registration and Management of Variants

Registration and management of variants is critical to their deployment, and relies on many user roles, though the registrant holds a central position in the process². Sales and marketing representatives from registrars and registries are involved in serving registrants. Trademark protection professionals are active in protecting the interests of the trademark holders, including registrants.

Registration management includes determining the right set of variants, presenting the choices to potential registrants, capturing the registrant preferences and information, making registrant information available for registered variants, activating these variants and then maintaining them over time (managing renewals and deletions). In some cases, the policy makers and community needs to be involved, to determine the sets of variants, how to manage information for these variants, and how to resolve registration disputes arising from them.

The following are challenges that have been identified.

6.2.1. Management across IDN TLDs inconsistent

Lack of consistent guidelines will make the registration process inconsistent across TLDs. The registration process of variants for TLDs and SLDs may require multiple aspects of the registration to be defined. These aspects and how they will be managed are not clearly identified and agreed upon by the community. Issues include at least the following: The relationship between SLDs and TLDs; primary vs. other variant labels; activation states and procedure to set and change these states; limits on activation of variants; and pricing of variants and changes. Without clear guidelines, the registration will be very different across TLDs.

For example, it is not clear if all variants could be activated. Further, how would TLDs decide which TLD variant should be primary, in case many can be activated? If a primary TLD variant is determined, can it be changed to non-primary at a later time? What will be the process? How would registrars handle arbitrary registry policies in this context, without confusing the registrants?

Applicable Usability Principles:

² Privacy and proxy services may be considered to be acting on behalf of the registrant.

- P3. Predictability
- P4. Equivalency
- P5. Consistency
- P6. Manageability
- P7. Ease of Use

6.2.2. Registration for SLDs across TLDs inconsistent

Differences in definition, arrangement, and activation of variants of a SLD across TLDs for a script (gTLDs and ccTLDs) may make the registration process difficult, especially if a registrant is managing similar domains across multiple TLDs.

The variant sets may be different for SLDs across TLDs. Further SLD variants may have different possible activation states for across TLDs. Additional complexity is involved as the SLD and the TLD labels may change their status (e.g. activation states and the choice of primary label at each level). Thus, the registration process may become much more complex and registry/registrar dependent.

For example, it may not be possible to have the same SLD variants to be activated in the same way across the various TLDs of the same script due to differing variant handling policies (activation status over time, primary label definition over time, pricing, etc.). Registrants may find the process for registration, renewal, de-activation and deletion of variant sets so confusing that they may be discouraged from using IDNs.

Applicable Usability Principles:

- P2. Security
- P3. Predictability
- P4. Equivalency
- P5. Consistency
- P6. Manageability
- P7. Ease of Use

6.2.3. Inconsistent association of ASCII and IDN TLDs

Some TLDs may associate ASCII and IDN TLDs for registrations, while for others this may not be possible or practical. This inconsistency may confuse users.

ASCII TLDs and IDN TLDs are not variants of one another. However, as a service to registrants and users, registries may package them together in the same way that variants are packaged. This may lead registrants to expect similar levels of support across other TLDs, and may result in additional confusion in the registration process.

For example, TWNIC is already offering registrants the ability to package IDNSLDs across ASCII and IDNTLDs. Similar mechanisms may also be done by gTLDs owned by same entities for more effective user facilitation and sales. However, in other cases, it may not be possible as the ASCII and IDNTLDs may be owned and operated by different entities.

Applicable Usability Principles:

- P3. Predictability
- P4. Equivalency
- P5. Consistency
- P6. Manageability
- P7. Ease of Use

6.2.4. Software support inadequate

Technology to support the registration currently does not handle variants, which may make variant registration challenging and difficult.

Registrants need to view all the variants for the domain name being registered, understand the differences between these variants, decide how to define which of the variants should be primary (if any), and which subset may be activated. Registrars normally offer automated services, through which domain names available through various registries are made available for sales, renewals and deletions. Currently, no well-defined automated services are available to support variants of the domain name, especially in the context of different variants for different registries for different levels. Further, there is no mechanism to update variant information over time, even if it is made available at one time for new and existing domain names. In addition, registrars also have back-end interfaces for communicating with registries, e.g. through EPP based technology.

For example, a user may not be able to access all variant TLDs during registration because the registrar interface is not enabled for variants. Alternatively, if a registrant has already registered an IDN, and variants are introduced afterwards, there is no mechanism defined to contact the user and update the variant activation. This issue is also relevant at other levels of the DNS tree.

Applicable Usability Principles:

- P2. Security
- P6. Manageability
- P7. Ease of Use

6.2.5. Registration system not straightforward to localize

As variants are defined at the script level, it is not possible to predict the language of the registrant for the localization of registration interface.

New registrants and users of IDNs and variants may be mono-lingual and may only understand a single non-Latin script. Registrars that are providing interfaces may not be able to predict the language of the registrant for a given TLD and thus may not be able to localize the interface. This may make it difficult for registrants to register IDNs and its variants, particularly given the inherent complexity of variants.

For example, a potential Arabic script IDN TLD registrant may speak Urdu, Pashto, Arabic, Farsi, Sindhi, or any other of the many languages which use Arabic script. A gTLD may cater to many of these communities. In some cases, even a ccTLD may cater to multiple languages, e.g. the Arabic script IDN ccTLD for Pakistan and India. A registry or registrar may not know which language to offer within the registration interface, or even offer support such a language.

Applicable Usability Principles:

- P1. Minimality
- P2. Security
- P5. Consistency
- P6. Manageability
- P7. Ease of Use

6.2.6. Registration information inconsistent

Variants are not part of the registration data and services infrastructure at this time. Therefore, registrants will not be able to define and access data related to variants consistently across registries, creating usability and security challenges.

The variants assigned to the registrant should point to the information entered at the time of registration. However, it is not clearly defined how variants will point to this information and how this registration information will store the variants (and their properties). Arbitrary solutions maybe developed by registries to collect and distribute this information related to variants. Some registries may provide look up functionality for all variants whereas others may provide look up functionality for only active variants, and still others may provide look up functionality only through the primary label. This would confuse both the registrant in producing the data, and users in accessing and interpreting the data.

For example, users may consider the variants which do not show up in the registration data query to be unrelated, while in other registries all variants may be fully displayed.

Applicable Usability Principles:

- P6. Manageability
- P7. Ease of Use

6.2.7. Trademark protection tracking complex

Due to potentially large set of variants and their confusingly similar strings, protecting trademarks may become much more complex and costly.

A domain which may be seemingly unrelated may have a variant which can be exactly or similar to an existing but different trademark. This may be difficult to anticipate and track, especially due to variety of implementations across TLDs, due to complexity of variant management processes for both SLDs and TLDs, and due to lack of support of registration data and services for variants. With many more TLDs and potentially many more variants and strings similar to these variants, tracking and protecting trademarks will be an uphill task.

For example, if a trademark is similar to a string of an inactive variant of a TLD, it may be hard to identify and monitor (especially if inactive variants are not available through registration data services).

Applicable Usability Principles:

- P6. Manageability
- P7. Ease of Use

6.2.8. Trademark protection dispute process complex

Variants may introduce new scenarios into the dispute process, which are currently not addressed by the dispute policy. Determining such cases may be more complex.

Current disputes are limited to single strings and confusions arising from them. When more than one label is included to form a variant set and registered to a single registration request, it may interfere with other (seemingly) unrelated strings and lead to more potential disputes. However, as many variants will be automatically generated, and not requested by the registrants, it will be difficult to determine the intent behind the registrations. Further, this will cause more critical cases where variants are introduced at a later stage after the primary labels have already been delegated. A comprehensive policy on how to deal with such cases is still not defined, especially in cases where they have legal implications on the Label Generation Ruleset.

For example, two characters considered distinct in a language may need to be collapsed as they are considered equivalent in another language using the same script (e.g. U+06A9 and U+06AA; see 6.1.2 for more details). However, this may invoke a trademark issue in the context of the community speaking the first language.

Applicable Usability Principles:

- P1. Minimality
- P2. Security
- P6. Manageability
- P7. Ease of Use

6.3. Challenges in the Configuration and Diagnostics of Variants

Configuration is critical for making variants operational and diagnostics are necessary for their continued health. This is primarily undertaken by system support professionals, network managers, security managers and other users, who are expected to be competent in their understanding of how the DNS works.

6.3.1. Software configuration not supported

Tools are currently not available for configuring variants for the DNS, web servers, email servers, etc. System configuration personnel may find it difficult to manage many variant domains.

Variants are expected to support equivalent behavior for users. However, system configuration personnel may find that software may not permit entering more than one domain name or, in cases in which software allows multiple entries, it processes a variant as just another unrelated domain name. Thus, it would be hard to configure the systems to simulate the desired behavior or to make sure each variant entry is appropriately configured.

For example, a webmaster configures the web platform for all activated variants (tens of variants are possible). But because each variant may require a separate, unrelated configuration, the process is both tedious and error-prone.

Applicable Usability Principles:

- P2. Security
- P6. Manageability
- P7. Ease of Use

6.3.2. Cannot associate variants for configuration

System configuration software may display domain names only in ASCII format. This means that IDN labels will be displayed at A-labels (not as U-labels), making it challenging for system configuration personnel to manage them.

Many of the operating and configuration systems may require domain labels to be written in ASCII (A-label), as they may not fully support non-ASCII file formats and filenames. This means that even if variants can be supported, they may still need to be configured as A-labels. However, A-labels, due to the nature of the algorithm which produces them, generate intractable ASCII strings which cannot serve as mnemonics.

For example, the IDN ccTLD label for Pakistan has two variants, پاکستان (with U+0643 as the third character) and پاکستان (with U+06A9 as the third character); the latter string is the primary label which has been approved for delegation. Both U-labels generate exactly the same mnemonic visually. Punycode³ for the former is “xn--mgbai9a5eva00b” compared to “xn--mgbai9azgqp6j” for the latter string. So it is difficult to use either representation conveniently to configure systems.

Applicable Usability Principles:

- P2. Security
- P4. Equivalency
- P6. Manageability

6.3.3. Compounded certificate management

Certificates for authentication are bound with domain names. From a security perspective, a variant is a different domain name and would require a new certificate, if activated. As there may be many domain name variants, certificate management will be challenging and more costly.

For each variant combination of a domain name (which would include variants of TLDs and of labels at lower levels), as separate certificate would need to be purchased and managed. Different certificates may come from different vendors, at different dates and configured on different systems. Thus, for tens or hundreds of variants (which are possible as no. of variants at each level have a multiplicative effect on total variants), certificate management may become a difficult proposition. This will have a negative impact on the operational security of the relevant portions of the DNS.

For example, a website administrator would need a different TLS/SSL certificate for each variant. These certificates may have different pricing and renewal dates. This would need to be tracked for all activated variants. See challenge 6.1.13 “Variants cause session re-establishment” which covers the impact of certificates and session management.

³ Punycode generated from <https://www.centralnic.com/portfolio/domains/idn/converter>.

Applicable Usability Principles:

- P1 Minimality
- P2. Security
- P4. Equivalency
- P6. Manageability

6.3.4. DNSSEC validation inconsistent

Depending on the variant being used, users may not be consistently validated through DNSSEC, even though users may be expecting such validation.

Each variant needs to be independently signed for DNSSEC, and may have different expiration dates or different DNS records characteristics. Thus, all variants may not have a consistent DNSSEC validation status, causing challenges in secure access of relevant systems.

For example, a user may access one variant which is DNSSEC validated. Then the user accesses another variant which is not DNSSEC validated, resulting in an inconsistent and potentially less secure user experience.

DNSSEC is required and used in additional and upper-layer protocols such as the DNS-Based Authentication of Named Entities (DANE) Transport Layer Security (TLS) Protocol: TLSA. Therefore, the issues listed above are then inherited in those higher protocols, therefore further complicating the issue.

Applicable Usability Principles:

- P1 Minimality
- P2. Security
- P4. Equivalency
- P6. Manageability

6.3.5. Log and history searching does not match

Logs of domain names accessed by a system are stored by software for multiple purposes. The applications using these logs may need to associate the variants, but may not be able or configured to do so, negatively impacting usability.

Many applications maintain logs of domain names for facilitating use, management, diagnostics and security of the system. System configuration personnel may use the domain names to filter certain website content. Network managers may use the domain names to monitor traffic and law enforcement agencies and other security personnel may use them to follow up or prevent illegal activities online. In such cases, they would need to have tools and technology which can collate the variants into a single set. However, current systems do not allow these users to collapse the variants into a single set.

For example, law enforcement organization may need to track suspicious activity against a domain name. They would need to manually manage all the (tens or hundreds of) active domain name variants (SLD.TLD) manually and to map the behavior. Furthermore, this issue gets worse when the tactic is used at many levels of the tree at the same time. This may be more time consuming (and may not be easily possible) to accomplish.

Applicable Usability Principles:

- P2. Security
- P6. Manageability

6.3.6. Network traffic statistics incomplete

Network administrators look at network flows to manage their network, such as bandwidth allocation. Because variants may not be collated by applications providing these statistics, variants may make it difficult to manage the networks efficiently.

Based on access patterns of different domain names, network administrators may engineer how bandwidth is distributed. However, for effective management, statistics for variants which are accessing same web content may need to be conflated, to give realistic numbers. Network managers may not be able to detect patterns if the traffic is distributed across the tens (or hundreds) of variant domains (e.g. variants of 3LD.2LD.TLD). Even when the systems may be developed to collate such traffic, it may still be hard to find out which two domain names are variants as the variant generation rule sets are distributed across multiple entities.

For example, netflows for variant1 will be separate from variant2. However, the network admin may need to see both aggregated into a single netflow.

Applicable Usability Principles:

- P3. Predictability
- P6. Manageability

6.3.7. Caching infrastructure inefficient

Web pages are cached to reduce network traffic and provide better response time to end users against their requests. As caching matches domain names, it may not work effectively with variants.

Variants will have adverse impact on caching. First, as caching management software may consider variants to be unique, it will cache one copy of the same page for each variant. Storing multiple copies of the same web page would mean reducing the number of unique web pages which can be cached given the fixed space available. In addition, if a variant which is not cached is accessed, the cached webpage will not be used and a fresh copy will be retrieved, increasing the network traffic and the delay. As domain names may have many variants, this may have adverse impact on the response time.

Applicable Usability Principles:

- P2. Security
- P3. Predictability
- P6. Manageability

6.3.8. Diagnostic and troubleshooting tools incompatible

To troubleshoot network issues, low-level command-line tools are used. These tools generally use A-labels for IDNs and will not support variants. In addition, they typically associate a single

domain name against an IP address. These constraints may make network management more complex.

While doing investigation or troubleshooting, various low-level and command-line tools such as ping, dig, wireshark, curl, etc. are used. These tools use an ASCII interface and cannot process variants. Therefore, troubleshooting and investigation is more difficult and error-prone and takes more time. Further, many troubleshooting tools are based on the concept that a domain name does not have variants. IP addresses in log records or in traffic interception are translated to names by DNS PTR records. DNS PTR records typically point to a single domain name and do not relate the variants. PTR records are used for troubleshooting, logging and access-control lists.

For example, an investigator wants to test availability of all the websites related to domain name. A tailored ping command for this community would test all variants to make the investigation more comprehensive and complete. As another example, a network administrator sees an IP address in logs. The DNS PTR translation points to variant1; all other variants are unknown and can't be found.

Applicable Usability Principles:

- P2. Security
- P3. Predictability
- P6. Manageability

6.3.9. Forensics significantly more complicated

Variant domains should be considered related and equivalent for forensics and should be reported accordingly to aid in investigations.

Forensics software used by security professionals and law enforcement uses various heuristics and databases to detect issues and to provide reports for investigation. These heuristics may consider each variant unique and the analysis would need to be manually configured for variants. The configuration will be further obfuscated by the fact that URLs consist of multiple levels, and each level has different variant rules across different registries, and these rules are distributed across multiple entities located in multiple geographical locations. Further, the rules may change over time and active variant status may also change over time and the history of such changes may not be maintained. The varied, distributed and dynamic nature of variant definition and management makes forensic analysis very complicated.

For example, if an intruder accesses domain name variant1 and domain name variant2, the forensics software may not identify the relationship between the variants. Further, variants may significantly increase the botnet capability. A botnet could use all possible variants to register its nodes. Investigations to correlate all of these nodes would be more complex

Applicable Usability Principles:

- P2. Security
- P3. Predictability
- P6. Manageability

7. Future Work

The current interim report has identified the user roles and relevant challenges in the context of implementing variants of IDN TLDs. Some of these challenges may not fall directly in the ambit of ICANN and will need to be addressed by the larger technical community, such as developers of search engines and web clients. However, many of these challenges may be addressed by either the communities involved with ICANN or by ICANN directly. Subsequent work will use the usability principles identified for the variants to suggest guidelines and recommendations on how these challenges may be addressed.

8. Authors

This report is written by the following (in alphabetical order by last name):

Marc Blanchet
Sarmad Hussain
Steve Sheng (ICANN Staff)
John Yunker

9. Acknowledgements

We would like to acknowledge the support of the following individuals (in alphabetical order by last name) for providing feedback and/or supporting the process:

Greg Aaron	Baher Esmat	Christopher Mulligan
Abdulrahman Alzarooni	Patrik Fältström	Rod Rasmussen
Abdulaziz Al-Zoman	Robert Flaim	Peter Roman
Fahd Baynteh	Frédéric Gaudreau	Jonathan Shea
Marc Boivin	Nai-Wen Hsu	Terri Stumme
Sébastien Bourdon-Richard	Manal Ismail	Andrew Sullivan
Cary Carp	Gary Kibbey	Wang Wei
Edmon Chung	John Klensin	Hongbin Zhu
Mohammed El Bashir	Marc Moreau	

We would also like to thank the following organizations for their feedback and/or for responding to our surveys:

China Internet Network Information Center (CNNIC)
Hong Kong Internet Registration Corporation Limited (HKIRC)
National Information Technology Center Jordan (NITC)
Royal Canadian Mounted Police
South Asian Network Operators Group (SANOG)
Serious Organised Crime Agency (SOCA), UK
Sureté du Québec
Taiwan Network Information Center (TWNIC)
UAE Telecommunications Regulatory Authority (Operator of .ae IDN ccTLD)
US Department of Justice (US DoJ)
US Federal Bureau of Investigation (FBI)

References

Relevant ICANN Documents on Variants TLDs

1. Frakes, J, *et al.*, “Considerations in the use of the Latin script in variant internationalized top-level domains: Final report of the ICANN VIP Study Group for the Latin script”, (Marina del Rey, California: ICANN, October 2011). <http://archive.icann.org/en/topics/new-gtlds/latin-vip-issues-report-07oct11-en.pdf>
2. Govind, *et al.*, “Devanāgarī VIP Team Issues Report”, (Marina del Rey, California: ICANN, October 2011). <http://archive.icann.org/en/topics/new-gtlds/devanagari-vip-issues-report-03oct11-en.pdf>
3. Hussain, S, *et al.*, “Internationalized Domain Names Variant Issues Project Arabic Case Study Team Issues Report”, (Marina del Rey, California: ICANN, October 2011). <http://archive.icann.org/en/topics/new-gtlds/arabic-vip-issues-report-07oct11-en.pdf>.
4. Internet Corporation for Assigned Names and Numbers, “The IDN Variant Issues Project: A Study of Issues Related to Management of IDN Variant TLDs (Integrated Issues Report)”, (Marina del Rey, California: ICANN, February 2012). <http://www.icann.org/en/topics/idn/idn-vip-integrated-issues-final-clean-20feb12-en.pdf>.
5. Lee, X., *et al.*, “Report on Chinese Variants in Internationalized Top-Level Domains”, (Marina del Rey, California: ICANN, October 2011). <http://archive.icann.org/en/topics/new-gtlds/chinese-vip-issues-report-03oct11-en.pdf>.
6. Segredakis, V., *et al.*, “Study of the Issues Present in the Registration of IDN TLDs in Greek Characters”, (Marina del Rey, California: ICANN, October 2011). <http://archive.icann.org/en/topics/new-gtlds/greek-vip-issues-report-07oct11-en.pdf>
7. Sozonov, A., *et al.*, “IDN Variant TLDs – Cyrillic Script Issues”, (Marina del Rey, California: ICANN, October 2011). <http://archive.icann.org/en/topics/new-gtlds/cyrillic-vip-issues-report-06oct11-en.pdf>

Relevant IETF RFCs

8. Alvestrand, H., Ed., and C. Karp, "Right-to-Left Scripts for Internationalized Domain Names for Applications (IDNA)", RFC 5893, August 2010.
9. Costello, A., "Punycode: A Bootstring encoding of Unicode for Internationalized Domain Names in Applications (IDNA)", [RFC 3492](#), March 2003.
10. El-Sherbiny, A., Farah, M., Oueichek, I., and A. Al-Zoman, "Linguistic Guidelines for the Use of the Arabic Language in Internet Domains", [RFC 5564](#), February 2010.
11. Faltstrom, P., Hoffman, P., and A. Costello, "Internationalizing Domain Names in Applications (IDNA)", [RFC 3490](#), March 2003.

12. Faltstrom, P., Ed., "The Unicode Code Points and Internationalized Domain Names for Applications (IDNA)", RFC 5892, August 2010.
13. Hoffman, P. and M. Blanchet, "Nameprep: A Stringprep Profile for Internationalized Domain Names (IDN)", [RFC 3491](#), March 2003.
14. Klensin, J., Faltstrom, P., Karp, C., and IAB, "Review and Recommendations for Internationalized Domain Names (IDNs)", [RFC 4690](#), September 2006.
15. Klensin, J., "Internationalized Domain Names for Applications (IDNA): Definitions and Document Framework", RFC 5890, August 2010.
16. Klensin, J., "Internationalized Domain Names in Applications (IDNA): Protocol", RFC 5891, August 2010.
17. Klensin, J., "Internationalized Domain Names for Applications (IDNA): Background, Explanation, and Rationale", RFC 5894, August 2010.
18. Konishi, K., Huang, K., Qian, H., and Ko, Y., "Joint Engineering Team (JET) Guidelines for Internationalized Domain Names (IDN) Registration and Administration for Chinese, Japanese, and Korean", RFC 3743, April 2004.
19. Lee, X., Mao, W., Chen, E., Hsu, N., and J. Klensin, "Registration and Administration Recommendations for Chinese Domain Names", RFC 4713, October 2006.

Relevant Unicode Documents

20. The Unicode Consortium (2012a). "Unicode Technical Standard 46: Unicode IDNA Compatibility Processing", M. Davis, M. Suignard (Eds.). Retrieved from <http://www.unicode.org/reports/tr46/>.

IDN ccTLD Variant Practices

21. Arabic Script IDN Working Group (2012). Available at: http://asiwg.org/wiki/Main_Page
22. Al-Zoman, A. and Al-Fayze, R. (2008) "Proposed Solution for Writing Domain Names in Different Arabic Script Based Languages", ASIWG 3rd meeting, Cairo 2008. Retrieved from: <http://xn--ghb5rwd.com/docs/Resources/ASIWG-proposal-Nov08.pdf>
23. China Internet Network Information Center (CNNIC) (2010) "Implementation Plan of .中国 (xn--fiqs8S) and 中國 (xn--fiqz9S)", Retrieved from <http://www1.cnnic.cn/html/Dir/2010/06/12/5852.htm>
24. Iran Network Information Center (IRNIC) (2010) "Internationalized Domain Names". Retrieved from: http://www.nic.ir/Internationalized_Domain_Names
25. Saudi Network Information Center (SaudiNIC) (2010) "Guideline Rules for writing Arabic IDNs under the IDN ccTLD", Retrieved from: http://nic.sa/en/view/writing_arabic_idn_guideline

26. Saudi Network Information Center. (2010) “Arabic Domain Name Registration Regulation for Saudi IDN ccTLD”, Retrieved from:
http://www.nic.net.sa/docs/ADN/Arabic_Domain_Name_Registration_Regulation_For_Saudi_IDN_ccTLD_V1.pdf
27. Taiwan Network Information Center (TWNIC) (2009) “TWNIC’s Registration Policy and Technical Solution for the .台灣 and .台湾 IDN ccTLD”, Retrieved from:
<http://www.twnic.net/english/dn/pdf7a.pdf>
28. Taiwan Network Information Center (TWNIC) (2010) “Registration Policy and Rules for .台灣 Top-Level Domain Names in the Chinese Characters”, Retrieved from:
<http://www.twnic.net/english/dn/policya.pdf>
29. UAE Telecommunications Regulatory Authority. (2012) “Arabic (dotEmarat) امارات. Domain Name Policy”, Retrieved from:
[http://cp.nic.ae/NIC/uaenic/english/pdf/emaratdomain/Arabic\(dotEmarat\)DomainNamePolicy.pdf](http://cp.nic.ae/NIC/uaenic/english/pdf/emaratdomain/Arabic(dotEmarat)DomainNamePolicy.pdf).

Usability Resources

30. Grudin, J. (1992). “Utility and Usability: Research Issues and Development Contexts”, *Interacting with Computers*, 4(2): 209–217.
31. International Organization for Standardization (ISO) (1998). “Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11 : Guidance on usability”, ISO 9241-1998(E), International Standards Organization (ISO), Geneva, Switzerland.
32. Nielsen, J. (1993). *Usability Engineering*, Morgan Kaufmann, San Francisco, CA, USA.
33. Tullia, T. and Albert, W. (2008). “Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics”, Morgan Kaufmann, San Francisco, USA.

Other References

34. ICANN Generic Name Supporting Organization (GNSO) (2007a). “Final Report: Introduction of New Generic Top-Level Domains”, Retrieved from:
<http://gns0.icann.org/en/issues/new-gtlds/pdp-dec05-fr-parta-08aug07.htm>.
35. ICANN Generic Name Supporting Organization (gNSO) (2007b). “Outcomes Report of The gNSO Internationalized Domain Names Working Group (IDN WG)”, Retrieved from
<http://gns0.icann.org/en/drafts/idn-wg-fr-22mar07.htm>.
36. Internet Corporation for Assigned Names and Numbers (ICANN). (2009d). “Registrar Accreditation Agreement”, Marina Del Rey, CA: ICANN. Retrieved from
<http://www.icann.org/en/registrars/ra-agreement-21may09-en.htm#2>
37. Internet Corporation for Assigned Names and Numbers (ICANN). (2010a) “Board resolution (2010.04.22.10) CNNIC Request for (Simplified and Traditional Chinese) IDN ccTLDs”,

- Marina Del Rey, CA: ICANN. Retrieved from <http://www.icann.org/en/groups/board/documents/resolutions-22apr10-en.htm#synchronized>
38. Internet Corporation for Assigned Names and Numbers (ICANN). (2010b) “Board resolution (2010.04.22.11) TWNIC Request for (Simplified and Traditional Chinese) IDN ccTLDs”, Marina Del Rey, CA: ICANN. Retrieved from <http://www.icann.org/en/groups/board/documents/resolutions-22apr10-en.htm#synchronized>
 39. Internet Corporation for Assigned Names and Numbers (2012a). “IDN ccTLD Fast Track Process”, Retrieved from: <http://www.icann.org/en/resources/idn/fast-track>
 40. Internet Corporation for Assigned Names and Numbers (2012b). “Updated Final Implementation Plan for IDN ccTLD Fast Track Process”, Retrieved from <http://www.icann.org/en/resources/idn/fast-track>
 41. Internet Corporation for Assigned Names and Numbers (2012c). “IDN Variant TLD Program – Revised Program Plan”, Retrieved from <http://www.icann.org/en/resources/idn/idn-variant-tld-updated-plan-23aug12-en.pdf>
 42. The Internet Architecture Board. (2012). Response to ICANN questions concerning “The interpretation of rules in the ICANN gTLD Application Guidebook”, Retrieved from: <http://www.iab.org/documents/correspondence-reports-documents/2012-2/response-to-icann-questions-concerning-the-interpretation-of-rules-in-the-icann-gtld-applicant-guidebook/>
 43. ICANN Security and Stability Advisory Committee (SSAC). (2012) “SSAC Advisory of the Delegation of Single-Character IDN Top-Level Domains”, (SSAC publication No. 052).
 44. ICANN Generic Names Supporting Organization (GNSO). (2012) “Draft Final Report of Internationalized Registration Data Working Group”, Marina Del Rey, CA: ICANN. Retrieved May 17, 2012, from <http://gnso.icann.org/issues/ird/ird-draft-final-report-03oct11-en.pdf>

Appendix A: Survey for System Administrators

There has been significant discussion in the global community on what may be the impact of introducing IDNs and variants, especially in the context of country code and generic Top Level Domains (TLDs). The current survey aims to bring out specific challenges which may be encountered in the context, from the perspective of the role of System Administration and Support.

For System Administration and Support multiple servers need to be configured and managed. These include a Domain Name Server, Web Server, Mail Server, Proxy Server, FTP Server and others. The current document is divided into separate sections for some of these servers, with the aim to understand what issues may arise as IDN variant TLDs are introduced, and how may these be addressed.

The questions below are being provided to guide the discussion, though you should feel free to add additional information you think may be relevant in the context even if it is not directly asked for in the sections below.

Thank you for your time and effort for filling in this form.

1. Web Server

- a. Variants of IDN TLDs can be numerous (even up to 10 or more variants), and some of them can be visually confusing. Further, their A-labels⁴ will be lengthy and contain a seemingly random ASCII sequence. This variation will create challenges to configure the Web Server.
 - i. Given the configuration tools, policies and processes used, list the issues which may be faced at particular steps in the configuration process?
 - ii. How may the tools, policies and processes be enhanced to address these challenges?
- b. Given multiple variants of an IDN TLDs, they could be directed to the same or different website directory
 - i. In which cases should they point to the same directory? Why?
 - ii. In which cases should they point to different directories? Why?
 - iii. Thinking through the steps involved in configuring the systems, what are any potential technical challenges in setting up these systems as in (i)?
 - iv. Thinking through the steps involved in configuring the systems, what are any potential technical challenges in setting up these systems as in (ii)?
 - v. Is there an advantage of using the scheme in (i) or (ii) over the other?
- c. Same website directory can be accessed by multiple variants by re-directing variants to one directory or through virtual hosting.
 - i. What are the advantages and challenges of re-directing variants?
 - ii. What are the advantages and challenges of using virtual hosting?
 - iii. What additional tools are needed to manage these effectively?
- d. How would TLS-enabled sites be specially impacted by variant IDN TLDs?
 - i. Web server configuration (virtual hosting)?
 - ii. TLS certificate management?
- e. Are there any other issues related to Web Servers which may be impacted by the introduction of variants in IDNs?

2. Mail Server

⁴The ASCII-compatible encoded (ACE) representation of an internationalized domain name, i.e. how it is transmitted internally within the DNS protocol. A-labels always commence with the prefix "xn--" (source: <http://www.iana.org/glossary>). Also see RFC 3492 and RFC 5890 for further details.

- a. End user would like to receive the mails sent to variant email addresses in a single mailbox. Mails sent using variant email addresses will look up MX record in the DNS server and use that information to select the mail server.
 - i. How could a single mailbox be configured to receive mails from different variant email addresses (e.g. through forwarding? Aliasing? Other mechanisms?)?
 - ii. What could be the configuration and operational challenges in pointing variant mail addresses to the same mailbox?
 - iii. How may the different email services be impacted by variants:
 1. SMTP
 2. IMAP
 3. POP
 4. Malware and SPAM filtering
 - a. Could variants cause additional false positives?
 - b. Could variants cause additional false negatives?
 - iv. Will there be a need for updating DB settings for handling variants in a mailbox or can it be handled automatically? Will there be any other implications on managing the DBMS?
3. Domain Name Server
 - a. Fundamental question regarding IDN variant TLDs is whether there should be single or multiple zones. If a single zone is used, then there has to be a mechanism in the root to direct the multiple variants to same zone, similar to the mechanisms of CNAME and DNAME. For multiple zones, a synchronizing mechanism has to be employed.
 - i. Can the current technology map IDN variant TLDs to a single zone? If yes, how? What would be the challenges and advantages of using this approach?
 - ii. Can the current technology synchronize multiple zone files, one zone file per IDN variant TLD? What would be the challenges and advantages of using this approach?
 - b. What are general issues related to managing records in the DNS zone if IDN variant TLDs are introduced?
 - c. Which resource records in the DNS zone may bear impact due to IDN Variant TLDs? What will be the challenges?
 - i. MX
 - ii. A
 - iii. NS
 - iv. TXT
 - v. PTR
 - vi. CNAME
 - vii. SRV
 - viii. NAPTR
 - d. Will there be any issues in synchronizing master-slave DNS configuration if variants are activated?
4. Proxy Server
 - a. Domain name re-writing/re-directing is done manually in Proxy servers. Will there be impact on this due to variants? Will there be any other issues introduced?
5. Please share any other comments you may feel are relevant.

Thank you for your feedback!

Appendix B: Survey for Registries and Registrars

This survey is designed to understand (i) what are the existing Registry/Registrar practices for variant management below the root, and (ii) how introducing variants in the root may impact the registry/registrar operations.

The survey is intended to gather data from multiple aspects, including capacity, technology, process and policy dimensions.

Existing Variants below the Root

IDN Variants are already in practice at second and lower levels at many registries.

Understanding how they are managed would contribute to planning how variants of TLDs should be managed, if and when they are introduced in the root.

IDN in general can be represented either as U-Label or A-Label. This duality is changing the way (ASCII LDH) names have been managed traditionally. IDN variants further complicate this management as now multiple labels would need to be managed as a logical set, and this set may be potentially large (with hundreds of variants for a label).

Registration Process

1. Briefly describe the registration process for a IDN domain names and their variants, listing the roles of staff members involved, the tools/utilities used and data gathered:

Capacity Building of Staff

2. What are the different people/roles in the Registry/Registrar who have to understand and deal with variants (e.g. policy makers, management (viewing billing reports, performance and security reports, etc.), marketing and sales staff, operations staff (e.g. DNS services, WHOIS services, EPP transactions, Billing, Security, etc.), etc.). Give the designation/roles of your organization and briefly state how they are involved with variants.

No.	Role (of Staff)	Brief Description of how the Staff is Involved with IDNs and Variants
1		
2		
3		
4		
		... add rows

3. What have been the training requirements for the staff for handling IDNs and Variants (e.g. informal description, formal training involving technical details, etc.)? How long have these training been (one hour, half day, one day, one week, etc.)?

No.	Role (of Staff)	Training Needs for IDNs (topics discussed) and Duration	Additional Training Needs (topics discussed) for IDN Variants and Duration
1			
2			
3			
4			
			... add rows

Technology Support

4. For each the role above, list the **desired** tools, utilities and other software (e.g. for DNS, WHOIS, EPP (for registration, renewal, updation, deletion, etc.), Billing, Security, etc.) to be used by the person performing the role (if any) to handle IDNs and additionally any changes to handle IDN variants.

No.	Tool/Utility	Relevant functions/use within the Tool/Utility	Role(s) of Staff using the Tool/Utility	Changes needed for the management of IDNs	Additional changes in software for managing IDN variants	Status: Current/Future
1						
2						
3						
4						
						... add rows

Process Changes

5. Which processes (e.g. registration, registration data collection, billing, dispute resolution, sunrise, gathering analytics, etc.) have changed due to the introduction of IDNs and variants? How have these operating procedures changed? Are there any procedures which may change (but not updated yet) based on the operational experience so far?

No.	Process/Operating Procedure	Brief Description of change due to IDNs	Additional Change due to IDN Variants
1			
2			
3			
4			
			... add rows

Policy Update

6. What policy (e.g. registration (and limits of variants), billing, sunrise, dispute, etc.) has been developed specifically to deal with IDNs and additionally for IDN variants?

No.	Policy	Brief Description of change due to IDNs	Additional Change due to IDN Variants
1			
2			
3			
4			
			... add rows

Variant TLDs at the Root

There are currently discussions in the community about the need and methods to introduced TLD variants in the root zone. If such variants are eventually introduced, they will have an impact on Registry/Registrar operations as well. The following questions pertain to impact on the Registry/Registrar in this context.

Capacity Building of Staff

7. What are the different people/roles in the Registry/Registrar who will have to understand and deal with variants TLDs (e.g. policy makers, management (viewing billing reports, performance and security reports, etc.), marketing and sales staff, operations staff (maintaining DNS services, WHOIS services, EPP transactions, Billing, Security, etc.), etc.)? Give the designation/roles of your organization and briefly state how they may be involved with variants. What would be the training requirements for the staff for handling Variants of IDN TLDs(e.g. informal description, formal training involving technical details, etc.)? How long would this training need to be (one hour, half day, one day, one week, etc.)?

No.	Role (of Staff)	Brief Description of how the Staff may be Involved with IDN TLD Variants	Training Needs for introducing Variants of IDN TLDs(topics to be discussed) and Duration
1			
2			
3			
4			
		... add rows	

Technology Support

8. For each the role above, list the desired tools, utilities and other software (e.g. for DNS, WHOIS, EPP (for registration, renewal, updation, deletion, etc.), Billing, Security, etc.) to be used by the person performing the role (if any) to handle IDN TLD Variants.

No.	Tool/Utility	Relevant functions/use within the Tool/Utility	Role(s) of Staff using the Tool/Utility	Changes needed for the management of IDNs	Additional changes in software for managing IDN variants	Status: Enhancement Available/Desired but not available
1						
2						
3						
4						
						... add rows

Process Changes

9. Which processes (e.g. registration, registration data collection, billing, dispute resolution, sunrise, gathering analytics, etc.) would need to be updated due to the introduction of IDN TLD Variants?

No.	Process/Operating Procedure	Brief Description of change due to IDN TLD Variants
1		
2		
3		
4		

Policy Update

10. What policy (e.g. registration (and limits of variants), billing, revised-sunrise of existing names, dispute, etc.) would need to be developed to deal with IDN TLD Variants?

No.	Policy	Brief Description of change due to IDN TLD Variants
1		
2		
3		
4		

Concerns, Comments, Suggestions, etc.