**New gTLD Program:  IDN Variants**

**Feb 2010**

## Introduction

Language communities that plan to use IDNs and variant characters are affected by the management and implementation of variants in new TLDs.  This is relevant for both gTLD and ccTLD implementations.  The goal is to develop a consistent set of procedures for the development and use of IDN tables, and for variant TLD allocation and management.

Definitions:

1.  An IDN table provides the list of characters available for registration in domain names according to registry policy, and contains any variant characters.
2.  Variant characters are characters with two or more representations (that may appear confusingly similar to each other).
3.  Variant TLDs are identical to one another except that variant characters are substituted for one another in the TLD string.

The management of IDN tables and variant characters for top-level strings has been discussed in the community in multiple contexts.  An independent Implementation Working Team was formed after discussions during the ICANN meetings in Mexico City and Sydney in 2009 to look at these issues.  Part of the scope of the working team was to study the topic of variant management in TLD strings and propose a potential solution.  The team included linguistic and technical experts from various language communities, and was co-chaired by two ICANN Board Directors who are well-versed in the fields of IDN and DNS.

The team recommended that variants not be delegated as TLDs at this time, but that mechanisms be tested to enable future use of variant TLDs.  It was specifically recommended that DNAME as a mechanism for variant delegation be systematically tested to formulate an appropriate solution for ensuring consistency when deploying variants in the DNS.  The team concluded that if variants are to be delegated in the future, certain conditions must be fulfilled, as specified in an arrangement (i.e., agreement) between ICANN and the applicant.  Subsequently a proposal for a new resource record (referred to as BNAME) has been made through Internet Drafts in the IETF as a solution for management of variant TLDs.  ICANN will test DNAME and BNAME, as well as other viable solutions, and compare the results.

The recommendations also suggested that a TLD applicant should identify any variant strings for the TLD as generated by the relevant IDN table, and if the applicant "desired" those variants, they would be allocated to the applicant in a pending status until such time as a mapping mechanism is completed.  It was also recommended that the applicant should identify "undesired" variants, which should be blocked as a preclusive measure to avoid applications being submitted for these strings.

The Implementation Working Team's report is at
http://www.icann.org/en/announcements/announcement-2-03dec09-en.htm.

See additional background on this topic at http://www.icann.org/en/topics/idn/fast-track/proposed-implementation-details-idn-tables-revision-1-clean-29may09-en.pdf.

## Proposed Terms for Discussion

To date, drafts of the gTLD Applicant Guidebook have not provided for the possibility of IDN variant strings at the top level. The proposed new text allows for the potential future delegation of variant TLDs pending the completion of a mechanism and a process to be developed.

Potential guidebook text based on the working team's recommendations is included below to help inform the discussion. The relevant section appears in Module 1 of the guidebook; see the full module at http://icann.org/en/topics/new-gtlds/intro-04oct09-en.pdf. Module 1 provides an introduction to the gTLD application and evaluation process, including information on requirements specific to applicants for IDN gTLDs.

Under the approach proposed here, ICANN collects lists of variant TLD strings from all new gTLD applicants. For the new gTLD application process (as in the IDN ccTLD Fast Track), IDN Tables are required to be submitted together with the application for the associated TLD string. Variant strings are generated from the applicant's IDN table for the relevant language.

At the time of the application, the applicant must identify which variants are desired (i.e., those strings the applicant might wish to have delegated as TLDs after a mechanism is developed), and which variants are undesired (i.e., those strings the applicant does not wish to make use of, but should be blocked). However, no variants are delegated into the root zone as a result of the application. Variants may only be delegated when there is a mechanism developed and tested, and the applicant has demonstrated that it can implement the mechanism in line with the accepted standards. This approach is consistent with that taken in the IDN ccTLD Fast Track. In the Fast Track, requests are to indicate the variant strings based on the IDN table and indicate those that are desired. As noted in the Requestors Manual for IDN ccTLD Fast Track Participants, this does not mean that the desired variant TLD will be delegated in the DNS root zone.

The implementation approach discussed here collects variant information from applicants and allows for additional detail on the procedures to be confirmed as work continues in this area.

ICANN encourages comment on the language provided here. This language is for discussion only, and has not yet been incorporated into the Applicant Guidebook. Comments will be considered for version 4 of the full draft Applicant Guidebook, scheduled to be published in June 2010.

## 1.3 Information for Internationalized Domain Name Applicants

Some applied-for gTLD strings are expected to be Internationalized Domain Names (IDNs) that require the insertion of IDN-encoded A-labels into the DNS root zone. IDNs are domain names including characters used in the local representation of languages not written with the basic Latin alphabet (a - z), European-Arabic digits (0 - 9), and the hyphen (-).

### 1.3.1 IDN-Specific Requirements

An applicant for an IDN string must provide accompanying information indicating compliance with the IDNA protocol and other requirements. The IDNA protocol is found at http://icann.org/en/topics/idn/rfcs.htm.

Applicants must provide applied-for gTLD strings in the form of both a **U-label** and an **A-label**.

An A-label is the ASCII form of an IDN label. Every A-label begins with the IDNA ACE prefix, "xn--", followed by a string that is a valid output of the Punycode algorithm, and hence is a maximum of 59 ASCII characters in length. The prefix and string together must conform to all requirements for a label that can be stored in the DNS including conformance to the LDH (host name) rule described in RFC 1034, RFC 1123, and elsewhere.

A U-label is the Unicode form of an IDN label, which a user expects to be displayed.

For example, using the current IDN test string in Cyrillic script, the U-label is <**испытание**> and the A-label is <**xn--80akhbyknj4f>.** An A-label must be capable of being produced by conversion from a U-label and a U-label must be capable of being produced by conversion from an A-label.

Applicants for IDN gTLDs will also be required to provide the following at the time of the application:

1. Short form of string (in English). The applicant will provide a short description of what the string would mean or represent in English.

2. Language of label (ISO 639-1). The applicant will specify the language of the applied-for TLD string, both according to the ISO's codes for the representation of names of languages, and in English.

3. Script of label (ISO 15924). The applicant will specify the script of the applied-for gTLD string, both according to the ISO codes for the representation of names of scripts, and in English.

4. Unicode code points. The applicant will list all the code points contained in the U-label according to its Unicode form.

5. IDN tables. An IDN table provides the list of characters eligible for registration in domain names according to registry policy. It will contain any multiple characters that can be considered "the same" for the purposes of registrations at the second level ("variant characters"). Once in use by an active TLD registry, tables will be lodged in the IANA Repository of IDN Practices. For additional information, see existing tables at http://iana.org/domains/idn-tables/, and submission guidelines at http://iana.org/procedures/idn-repository.html.

   IDN tables must be submitted for, at a minimum, the language or script for the applied-for gTLD string. IDN tables must be submitted for each language or script in which the applicant intends to offer IDN registrations at the second level.

   Applicants are urged to consider linguistic and writing system issues in their work of defining variant characters, and cooperate with other TLD operators that offer domain name registration with the same or visually similar characters. ICANN may also compare the IDN table submitted by the applicant with other tables that have been submitted for the same languages, and will inquire about any inconsistencies.

6. Applicants must further demonstrate that they have made reasonable efforts to ensure that the encoded IDN string does not cause any rendering or operational problems. For example, problems have been identified in strings with characters of mixed right-to-left and left-to-right directionality when numerals are adjacent to the path separator (i.e., a dot). If an applicant is applying for a string with known issues, it should document steps that will be taken to mitigate these issues in applications. While it is not possible to ensure that all rendering problems are avoided, it is important that as many as possible are identified early and that the potential registry operator is aware of these issues. Applicants can become familiar with these issues by understanding the IDNA protocol and in particular the proposed new version of the IDNA protocol (see http://www.icann.org/en/topics/idn/rfcs.htm), and by active participation in the IDN wiki (see http://idn.icann.org/) where some rendering problems are demonstrated.

7. **[Optional]** - Representation of label in phonetic alphabet. The applicant may choose to provide its applied-for gTLD string notated according to the International Phonetic Alphabet (http://www.langsci.ucl.ac.uk/ipa/). Note that this information will not be evaluated or scored. The information, if provided, will be used as a guide to ICANN in responding to inquiries or speaking of the application in public presentations.

## 1.3.2 IDN Variant TLDs

Each application contains one applied-for gTLD string. For the applied-for string, the applicant may also specify any existing variant strings. A variant string results from the substitution of one or more characters in the applied-for gTLD string with variant characters. Each variant string listed must also conform to the string requirements in section 2.1.1.3.2.

For each variant string listed, the applicant must specify whether it is:

- A Desired Variant. This indicates that the applicant wishes to make use of the string as a TLD at a later date.

or

- An Undesired Variant. This indicates that the applicant does not intend to make use of the string at any point.

Note that two variant characters may or may not be visually similar to one another. Thus, there is no requirement that variant strings be visually similar to another. This is dependent on the script involved.

Translations or transliterations of TLD strings are not considered variants. The GNSO recommended in their gTLD policy development work that rights that priority rights for new strings on the top-level should not derive from existing strings.

ICANN will record variant strings listed by the applicant for information, but will not perform evaluation steps on the variant strings at this stage.

If the application is successful, only the applied-for gTLD string will be delegated as a gTLD. Variant strings may be delegated only when a mechanism for managing variant TLDs is completed and has been tested by ICANN. At that time, applicants will be required to submit additional information including implementation details for the variant TLD management mechanism, and to participate in a subsequent evaluation process, with expected additional fees and review steps to be determined.

## 1.3 *Information for Internationalized Domain Name Applicants (Redlined to show changes from Guidebook v3)*

Some applied-for gTLD strings are expected to be Internationalized Domain Names (IDNs) that require the insertion of IDN-encoded A-labels into the DNS root zone. IDNs are domain names including characters used in the local representation of languages not written with the basic Latin alphabet (a - z), European-Arabic digits (0 - 9), and the hyphen (-).

### 1.3.1 *IDN-Specific Requirements*

An applicant for an IDN string must provide accompanying information indicating compliance with the IDNA protocol and other requirements. The IDNA protocol is currently under revision and its documentation can be found at http://icann.org/en/topics/idn/rfcs.htm http://tools.ietf.org/wg/idnabis/.

Applicants must provide applied-for gTLD strings in the form of both a **U-label** and an **A-label**.

An A-label is the ASCII form of an IDN label. Every A-label begins with the IDNA ACE prefix, "xn--", followed by a string that is a valid output of the Punycode algorithm, and hence is a maximum of 59 ASCII characters in length. The prefix and string together must conform to all requirements for a label that can be stored in the DNS including conformance to the LDH (host name) rule described in RFC 1034, RFC 1123, and elsewhere.

A U-label is the Unicode form of an IDN label, which a user expects to be displayed.

For example, using the current IDN test string in Cyrillic script, the U-label is <**испытание**> and the A-label is <**xn--80akhbyknj4f>.** An A-label must be capable of being produced by conversion from a U-label and a U-label must be capable of being produced by conversion from an A-label.

Applicants for IDN gTLDs will also be required to provide the following at the time of the application:

1. Short form of string (in English). The applicant will provide a short description of what the string would mean or represent in English.

2. Language of label (ISO 639-1). The applicant will specify the language of the applied-for TLD string, both according to the ISO's codes for the representation of names of languages, and in English.

3. Script of label (ISO 15924). The applicant will specify the script of the applied-for gTLD string, both according to the ISO codes for the representation of names of scripts, and in English.

4. Unicode code points. The applicant will list all the code points contained in the U-label according to its Unicode form.

5. IDN tables. An IDN table provides the list of characters eligible for registration in domain names according to registry policy. It will contain any multiple characters that can be considered "the same" for the purposes of registrations at the second level ("variant characters"). Once in use by an active TLD registry, tables will be lodged in the IANA Repository of IDN Practices. For additional information, see existing tables at http://iana.org/domains/idn-tables/, and submission guidelines at http://iana.org/procedures/idn-repository.html.

   6. IDN tables must be submitted for, at a minimum, the language or script for the applied-for gTLD string. IDN tables must be submitted for each language or script in which the applicant intends to offer IDN registrations at the second level.

   Applicants are urged to consider linguistic and writing system issues in their work of defining variant characters, and cooperate with other TLD operators that offer domain name registration with the same or visually similar characters. ICANN may also compare the IDN table submitted by the applicant with other tables that have been submitted for the same languages, and will inquire about any inconsistencies.

6. Applicants must further demonstrate that they have made reasonable efforts to ensure that the encoded IDN string does not cause any rendering or operational problems. For example, problems have been identified in strings with characters of mixed right-to-left and left-to-right directionality when numerals are adjacent to the path separator (i.e., a dot). If an applicant is applying for a string with known issues, it should document steps that will be taken to mitigate these issues in applications. While it is not possible to ensure that all rendering problems are avoided, it is important that as many as possible are identified early and that the potential registry operator is aware of these issues. Applicants can become familiar with these issues by understanding the IDNA protocol and in particular the proposed new version of the IDNA protocol (see http://www.icann.org/en/topics/idn/rfcs.htm), and by active participation in the IDN wiki (see http://idn.icann.org/) where some rendering problems are demonstrated.

7. **[Optional]** - Representation of label in phonetic alphabet. The applicant may choose to provide its applied-for gTLD string notated according to the International Phonetic Alphabet (http://www.langsci.ucl.ac.uk/ipa/). Note that this information will not be evaluated or scored. The information, if provided,

will be used as a guide to ICANN in responding to inquiries or speaking of the application in public presentations.

## *1.3.2  IDN Variant TLDs*

Each application contains one applied-for gTLD string.  For the applied-for string, the applicant may also specify any existing variant strings.  A variant string results from the substitution of one or more characters in the applied-for gTLD string with variant characters.  Each variant string listed must also conform to the string requirements in section 2.1.1.3.2.

For each variant string listed, the applicant must specify whether it is:

- A Desired Variant.  This indicates that the applicant wishes to make use of the string as a TLD at a later date.

   or

- An Undesired Variant.  This indicates that the applicant does not intend to make use of the string at any point.

Note that two variant characters may or may not be visually similar to one another.  Thus, there is no requirement that variant strings be visually similar to one another.  This is dependent on the script involved.

Translations or transliterations of TLD strings are not considered variants. The GNSO recommended in their gTLD policy development work that priority rights for new strings on the top-level should not derive from existing strings.

ICANN will record variant strings listed by the applicant for information, but will not perform evaluation steps on the variant strings at this stage.

If the application is successful, only the applied-for gTLD string will be delegated as a gTLD.  Variant strings may be delegated only when a mechanism for managing variant TLDs is completed and has been tested by ICANN.  At that time, applicants will be required to submit additional information including implementation details for the variant TLD management mechanism, and to participate in a subsequent evaluation process, with expected additional fees and review steps to be determined.

*Note on Variants* -- Currently, the gTLD application process is established so that each application is for one string, whether ASCII or IDN. There has been comment that applications for IDN

strings should also accommodate variant strings. Discussions on possible methods of managing variants at the top level have indicated that restricting variants from being delegated in the DNS root zone might disenfranchise certain regions that otherwise would benefit greatly from the introduction of IDN TLDs.

Delegating variant TLDs in the root zone without a mechanism for ensuring that the TLDs are treated in a method that guarantees a good user experience is a stability concern related to confusability for end-users. This can be compared to the "companyname.com" situation, where two domain names (one with all Latin characters and the other with mixed Latin and Cyrillic) look identical, but were different technically. Users clicked on the "wrong" address leading to a site different than expected. This activity resulted in a change in the IDN Guidelines, requiring that scripts not be mixed in domain names unless there is a linguistic reason for doing so (e.g., in the case of Japanese that is represented by mixing of four scripts). This is also a requirement for TLDs, but does not solve the variant issue.

At the same time, disallowing or blocking variant TLDs means that some users will have a very difficult time using the IDN TLDs. In some cases it is not possible for the user to know which character he or she is typing. Some keyboards will offer one or another variant character but not both. In this way, without the variant TLDs in the root, communities may be getting error messages when attempting to reach, for example, a web address with a domain name under one of these IDN TLDs. This is not the intent of IDN deployment. Rather, the objective is to help all communities have equal access to the Internet.

Not all variants are visually confusing. To maximize benefit, ICANN has attempted to define variants in a narrow manner, only including variants that are visually confusing. The intent was to allow variant TLDs that are not visually confusable with others to be delegated in the DNS root zone while a stable solution was found to address the variants that are similar.

At this time it is an open question whether stability issues include variant TLDs that look different, and are typed differently, but are used interchangeably for the same term by the users.

Another open question is the content of an agreement between the IDN TLD operator and ICANN requiring that registrations under two variant TLDs be handled (say, in a bundled or aliased manner, following RFC 3747, or a different technical solution) in a certain manner.

Finally, there is the question of whether it is necessary to enforce rules required for the development of IDN Tables. IDN Tables hold information about the characters that should be treated as

variants. The TLD operators develop IDN tables. Presently, TLD operators are urged to consider linguistic and writing system issues in their work of defining variants, and cooperate with other TLD operators that offer the same or very similar looking characters. This is not always practically possible, and there are currently no rules about defining variants. There also are no defined dispute mechanisms in cases where communities may disagree on a variant definition.

An implementation support team of technical and linguistic experts is examining this set of issues and expects to publish a proposed solution for managing variants at the top level. The proposed solution would then be available for public comment.