



IDN ccTLD Fast Track Program Proposed Implementation Details Regarding:

Development and Use of IDN Tables and Character Variants for Second and Top Level Strings

Date of Publication:

18 February 2009

Background - IDN ccTLD Fast Track Process

One of the most significant innovations in the Internet since its inception will be the introduction of top level Internationalized Domain Names (IDNs). These will offer many new opportunities and benefits for Internet users around the world by allowing them to establish and use domains in their native languages and scripts.

The topic of IDNs has been discussed in the ICANN community for a number of years. Initially, development was focused on enabling the introduction of IDNs as registrations under existing TLDs, but focus has shifted to be on broadening the characters repertoire available for use in top level strings as well. The IDN ccTLD Fast Track Process is one process ICANN is working on that will enable such introduction. The process for implementation of new gTLDs will also support Internationalized Top Level domains as part of the new [gTLD program](#).

The initial steps for introduction of IDN ccTLDs were initiated by the ICANN Board at its meeting in Sao Paulo (December 2006). During consultations and discussions of the then joint GAC and ccNSO IDN working group, it became clear that a number of countries and territories have a pressing need for IDN ccTLDs. The IDN ccTLD Fast Track Process is specifically aiming at meeting this near-term demand and at gaining experience with the mechanisms for selection and authorization of such TLDs that can inform the ongoing long-term policy development process.

The implementation of the IDN ccTLD Fast Track Process is underway and it based on the [Final Report](#) of the [IDNC Working Group](#), recommending mechanisms to introduce a limited number of non contentious IDN ccTLDs, associated with the ISO 3166-1 two-letter codes. In the [initial Draft Implementation Plan](#) for the IDN ccTLD fast Track Process a number of open issues were identified that require further input from the community and need to be resolved, to complete the implementation.

This paper is part of a series of papers that will serve as proposed solutions on these open issues. The proposed solutions are based on received public comments and input received through meetings, such as those held during the ICANN meeting in Cairo, Egypt, November 3-7, 2008. The papers are being posted in conjunction with an updated Draft Implementation Plan to seek further community collaborations in particular before and during the ICANN Meeting in Mexico City, Mexico, March 1-6, 2009. A public comment period for these papers is made available to enable and document such community discussions. Received comments will then be used to revise the plan in preparation of a Final Implementation Plan.

Please note that this is a proposed discussion draft only. Potential IDN ccTLD requestors should not rely on any of the proposed included details as it remains subject to further consultation and revision.

A full overview of activities related to the IDN ccTLD Fast Track Process and implementation thereof can be viewed here: <http://www.icann.org/en/topics/idn/fast-track/>

Summary of Key Points in this Paper

- IDN Tables and associated character variants are developed to reduce the potentially increased confusion which may otherwise be posed to end users by the introduction of IDNs.
- Clarifications and proposed recommendations are made for the process of developing IDN Tables and in particular definition and use of variant characters, in both second and top level strings.
- It is strongly recommended that there be collaboration among communities sharing scripts, or where particular confusability exists between characters across the used languages, to develop IDN Tables and associated policies. This will ensure that all language communities are afforded equal opportunity for making their languages available for domain name registration.
- The proposed recommendations do not change how the development of IDN Tables has been developed previously. The responsibility for table development remains with the TLD registries.

I. Executive Summary

The topics of IDN Tables and variant characters were discussed in several sessions during the ICANN meeting in Cairo, Egypt, November 2008. As a result, some clarifying information was included in an update to the Draft Implementation Plan at <http://www.icann.org/en/announcements/announcement-26nov08-en.htm>

This paper provides additional information on IDN Tables, and why they are beneficial to TLD registries that are planning to introduce IDNs (either at the second or top level). The paper describes, in outline form how an IDN table can be developed, and a methodology for how ICANN will use the IDN Tables provided by registries for the TLD allocations and management.

In summary:

1. An IDN Table is a tabular listing of all characters that a TLD registry is making available for domain name registration.
2. A TLD registry can have more than one such table, for example one per language. The table can be based on either: a language; set of language; or a script (per the IDN Guidelines).
3. Variant characters are two or more characters that have "the same meaning" when used in domain name registrations.
4. The IDN tables that define variant characters are useful because they reduce the potential for confusion that could result from typographic similarities.

5. Procedures for the development of IDN tables are proposed in this document. In these procedures applicants are strongly encouraged to collaborate when potential confusion might exist with languages of other countries and territories:
 - a. Languages/scripts are sometimes shared across geographic boundaries. In some cases this can cause confusion among the users of the corresponding language or script community.
 - b. Visual confusion can also exist in some instances between different scripts (for example, Greek, Cyrillic and Latin). An IDN Table with cross-dependencies of identified variant characters can limit this confusion in cases when several scripts are used under a TLD.
6. ICANN's limited role regarding the development of the IDN Tables will be to provide support to applicants when requested.
7. This paper proposes that ICANN will employ all submitted IDN tables when considering request for top-level strings. The tables will be used as a guide to determine if an applied for string would result in confusion with an existing string. Where user confusion would result from the use of a variant character the applied for string will not be delegated into the root zone.

By publishing this paper ICANN is actively soliciting your comments on this important subject. This feedback will play a key role in shaping final implementation plans, intended for presentation at the ICANN meeting in Sydney (June 2009).

II. IDN Table Definition

An IDN Table is a list of all those characters that a particular TLD registry supports beyond the twenty-six letters of the basic Latin alphabet (a-z), ten digits (0-9), and the hyphen (-). If any characters in a table are considered to be variants of each other (essentially meaning "the same as"), this is indicated next to each character in a variant group. The term "variant" designates orthographic equivalence on the character level, such as that between "œ" and "æ" in "encyclopœdia" and "encyclopaedia", but not in the broader sense that pertains to the variant spelling of words, as "encyclopaedia" vs. "encyclopedia" or "color" vs. "colour".

An IDN Table will typically contain characters that either represent a specific language, or are taken from a specific script without particular reference to any of the languages that are written with it. The term "IDN Table" as it is used here, corresponds to what in previous contexts was referred to as a "variant table", a "language variant table", a "language table", or a "script table".

Expertise in linguistics and orthography is required to determine whether a character should be considered a variant of another character, and the same elements of a given script may be regarded differently from language to language. (Referring again to the example of "œ" and "æ", in an English language table, the former would likely be treated as a variant form of the latter. In a Danish language table, the "œ" would be a separate letter of the alphabet.) The recommendations here do not change that approach.

III. The benefit of having IDN Tables

When the number of characters available for inclusion in domain names was expanded from the 37 characters noted earlier to about 100.000 characters from numerous scripts,

the potential for confusion resulting from typographic similarities increased dramatically. Even though a computer can, for example, easily recognize the difference between “a” (Latin), “α” (Greek), and “а” (Cyrillic), the human eye cannot. This difficulty is further increased by differences between fonts, the sizes at which they are displayed, and the time required to process and remember the character used.

To reduce this heightened level of potential confusability, (per the IDNC Final report recommendations) a TLD registry’s registration policy for IDNs must include the creation of IDN Table(s); so that a TLD registry’s IDN registration policy is based on a clearly defined set of characters. By using similarly structured IDN Tables TLD registries maintain a comparable basis for indicating the characters made available for registration, and the specific terms that apply to characters that are treated as variants of each other.

While the experience in this field is solely with reference to IDN registrations at second level under existing TLDs, as well as lower-level registrations, the basic concept is applicable to and becomes increasingly important with TLD strings. This ensures that we avoid having confusingly similar strings inserted in the root, in particular confusingly similar strings that are managed by different entities.

Historically IDN Tables have been developed by the TLD registries. And while IANA displays the tables online in a repository to provide a single source of information, ICANN’s IANA function does not validate the content of the tables. That said, the tables do need to fulfill the requirements articulated in the IDN Guidelines and the formatting rules from the IANA IDN Repository Procedure requirements, in order to be considered IDN Tables. The [IDN Guidelines](#) and [IANA IDN Repository Procedures](#) will, in turn, be adjusted in response to the outcome of the discussion of the present proposal, and its implementation.

IV. Development of IDN Tables

Depending on the number of characters in an IDN Table, and on which language or script it represents, varying degrees of difficulty will be involved in its development and in identifying the variants it may contain. For example, if a table holds characters from a single script that supports a single language, determining how that speech community regards similarity can be rather simple. However, if the characters in a script that is used to write many languages or if the TLD registry intends to support many languages, it may be more difficult to adequately consider the relevant linguistic elements of all those.

Fundamental differences among writing systems give rise to situations in which a given script element is used differently from language to language, which could confuse someone lacking a detailed understanding of variations in orthographic practices. This situation must be accepted in IDNs precisely as it is in other contexts where written language appears. Nonetheless, the user community will benefit from efforts to minimize the potential for confusion. The prototypical contribution to script-development-based policies serving multiple language communities has resulted in the Joint Engineering Team (JET) Guidelines for Internationalized Domain Names (IDN) Registration and Administration for Chinese, Japanese, and Korean, which can be seen at <http://www.ietf.org/rfc/rfc3743.txt>.

Similar initiatives are under way in language communities sharing other scripts, for example, the [Arabic Script IDN Working Group \(ASIWG\)](#).

The Arabic script is used widely for a number of languages originating in the Middle-East, Africa, and Asia. Each of these language communities will have its own perspectives on

the structuring of its IDN Table. The only way to ensure that the interests of every such community are reflected in the way their shared script is manifested in the IDN space, is for them all to take part in the coordinated development of the table(s), whether it is in development of one IDN Table for the script, or several IDN Tables for one or more languages. The alternative is to risk unintentional inconsistencies in the way a given element of a script is treated in different language tables in which it appears, to the disadvantage of all of the language communities sharing that script.

Another example of a similar initiative and a difficult situation is the more than 20 Indian languages that use about 13 scripts, and some of these languages are written with multiple scripts. Although the sizes of the respective language communities differ, no language within the country has a higher formal status than does any other. A common IDN Table, or several IDN Tables prepared in tandem, must consider the relevant linguistic elements from all languages sharing a script, or where visual confusability is a factor. This approach will ensure that all Indian languages can be supported on an equitable basis.

Regardless of the language or script basis, domain names do not always represent dictionary words, and nothing intrinsic to a label indicates the language or script it is intended to represent. Thus further attention must be given to the way a script is used for writing other languages that may be similarly reflected in IDNs (as the examples here illustrate). Without such action, the language-specific detail adopted by one registry could prove to be at odds with the policies of another registry supporting some other language also written the same script, possibly creating confusion within the broader Internet user community.

Usage of IDN Tables and variant characters in domain name registrations

There are a variety of ways to deal with variant characters in domain name registrations. Short descriptions of those that are most common follow. Which approach a TLD registry will take has historically been decided by the TLD registry alone. The recommendations in this paper do not change that approach and as such the following is provided for informational purposes only.

1. Bulk registrations – the characters that are variants will result in the registrant receiving two or more registrations (the variant domain names) for the same prices and automatically as one.
2. Blocked registrations – the characters that are variants will result in the blocking of the variant domain name(s). A block of a domain name means that it can never be registered.
3. Reserved registrations – the characters that are variants will result in a reservation of the variant domain name(s). A reservation most commonly means that only the registrant can release the reservation and register the domain name in question.

Proposed IDN Table procedure for SLD registration usage

The IDNC Final Report recommendations require that one or several IDN Tables are made available for any IDN ccTLD Fast Track applications. The IDN Guidelines makes the same observations for registries wishing to provide IDN support in domain name registrations.

The following proposed procedure is put forward to provide some additional clarifications around how IDN Tables can be developed. The proposal is for all TLD registries wishing to support IDNs at the second level.

Draft—for discussion only—please refer to the disclaimer on the title page of this document.

The primary goal of the following proposal is to ensure that all language communities have an equal opportunity for making their languages available for domain name registration.

1. The IDN ccTLD fast track requestor decides the characters that will be available for inclusion in SLD labels, seeking at its own discretion the advice and comment from governmental agencies, and its target community.
2. The IDN ccTLD fast track requestor assesses the extent to which the characters on the resulting list can also be expected to appear in IDN ccTLD requests submitted by other countries or.
 - a. If there is no such likelihood, the requestor will decide if any characters should be listed as variants in its development of the associated IDN Table(s). (It is still recommended that advice be sought from expert linguists that are thoroughly familiar with the language or script).
 - b. If the characters are likely to appear in other requests, the requestor should coordinate the development of the IDN Tables(s) and the listing of variant characters with the corresponding action in other countries or territories. This collaboration should decide whether a single character table can be shared or if separate tables are required. This joint effort is the only means to ensure that inadvertent confusion is avoided, and to prepare a narrative explanation for the general user community, of the reasons for any unavoidable ambiguity.
3. As the requestor must be able to determine in Item 2) which other countries or territories to collaborate with, as part of the Fast Track process, ICANN will facilitate bringing requestors into contact with bodies having relevant linguistic expertise, if such assistance is needed.

Proposed IDN Table usage for TLD Registrations

The IDNC Final Report recommendations require that one or several IDN Tables are made available for any IDN ccTLD Fast Track applications. The IDN Guidelines makes the same observations for registries wishing to provide IDN support in domain name registrations.

The characters and variants presented in an IDN Table for SLD registration will also be applied to the top level. ICANN will use these IDN Tables when reviewing requests, and requestors are encouraged to consider this carefully when preparing their IDN Tables and selecting their TLD labels.

There will be situations in which an IDN ccTLD requestor may have reasonable grounds for wishing to have more than one label for the requested domain, which differ either in a detail of encoding that is not readily visible when displayed, or in some more obvious orthographic regard (called "variant strings"). There is, however, currently no standard or mechanism by which such aliasing can be implemented at the root level and the Fast Track Process does not provide for the delegation of multiple labels in the same language and script for a single IDN ccTLD.

ICANN proposes that variant strings be either allocated or blocked for registration, following the logical arguments and requirements here:

- a. Variant strings must fulfill the same requirements from the fast-track process as the requested string(s) in order to be allocated.

- b. While the IDNC Final Report on the Fast Track process recommended “one string per territory per official language” it was mute on the concept of variant strings.
- c. The concept of the number of strings should be expanded to allow various countries and territories to have their variant string(s) allocated. Otherwise the Fast Track Process objectives of meeting community demand would not be met, and it would most likely create unnecessary confusion among certain populations if variant strings were not allowed.
- d. The variant strings will be allocated only if it is agreed that they be treated as aliased functions of the requested string.
- e. The variant strings will be inserted as separate delegations in the DNS root zone.
- f. Since there is no known technical standard or mechanism by which aliasing can be successfully implemented at the root level, requestors must include in their IDN TLD implementations a mechanism for ensuring that aliasing is enforced between the requested string and the identified variant strings.

Variant strings fulfilling these requirements also must be requested by the IDN ccTLD applicants, with a specific focus on:

- In their IDN TLD requestors must provide a statement of support from an authority in the country or territory having recognized expertise in the orthography of the language in which the TLD label is represented. This expert also must be familiar enough with the writing systems of other languages using the same script to be able to attest that the TLD label uses the script in a manner that would not conflict with its use in another TLD label representing another language, or to indicate the extent of any potential ambiguity.

Further confusability prevention mechanisms at the root level are discussed in the Module 7 to the revised Draft Implementation Plan that has been released together with this paper.

Variant strings that do not fulfill the above requirements be blocked for allocation in the DNS. This would be in line with practices currently used by TLD managers for IDN second level registrations. Blocked strings will be considered as “existing strings” when incoming applications are checked for conflicts with existing TLDs. Therefore, any later application for the same string will be denied.

As mentioned in the beginning of this paper, ICANN is actively soliciting your comments on this important subject. This feedback will play a key role in shaping final implementation plans, intended for presentation at the ICANN meeting in Sydney (June 2009).