

# New gTLD Program: IDN 3-Character Requirement

Feb 2010

## Introduction

Based on the work of the GNSO new gTLD policy development committee, the GNSO Reserved Names Working Group, and an independent IDN Implementation Team formed to consider this issue, it is recommended that the three-character requirement be relaxed in certain cases: in those cases two-character top-level domains should be allowed. It is also recommended that single-character registrations at the top-level should not occur pending additional policy discussion and consideration by ICANN's Supporting Organisations.

Previous drafts of the gTLD Applicant Guidebook have included a requirement that *"Applied-for strings must be composed of three or more visually distinct letters or characters in the script, as appropriate."* Over the course of the New gTLD Program implementation, ICANN has received a number of comments on this topic, many of which request that this requirement be modified or eliminated to allow gTLDs consisting of fewer than three characters.

A thread of the public comment is that the requirement for a minimum of three characters in a gTLD label will hobble the use of IDNs for many scripts because one- or two-character strings in certain scripts represent meaningful words. Without a mechanism for allowing single and two character labels at the top level, utility of IDN gTLDs will be significantly limited in some parts of the world.

This issue was discussed in the GNSO policy development process on the introduction of new gTLDs, which recommended allocation of single and two-character strings in certain scripts through case-by-case review. As implementation details are worked out, it has been difficult to identify a clear set of rules for determining which cases would be allowed or disallowed, based on review of proposed strings of less than 3 characters.

ICANN requested additional feedback from the community to enable the development of a set of rules that could be employed to allocate single and two-character strings and, at the same time, would not create any stability or security problems for the DNS. Further discussion of the concerns and issues involved is available in the Explanatory Memorandum at <http://www.icann.org/en/topics/new-gtlds/three-character-30may09-en.pdf>.

An Implementation Working Team was formed after discussions during the ICANN meetings in Mexico City and Sydney in 2009. The team included linguistic and technical experts from various language communities, and was co-chaired by two ICANN Board Directors who are experts in the fields of IDN and DNS. The scope of the working team included study and development of a potential approach to allow gTLD strings of fewer than three characters where appropriate.

## Proposed Terms for Discussion

The potential new language highlighted in this section is based on public comments (see analysis of public comments on draft Applicant Guidebook v3) and the work of the Implementation Working Team. The Implementation Working Team's report is at <http://www.icann.org/en/announcements/announcement-2-03dec09-en.htm>.

Under this proposal, the minimum string length for certain IDN gTLD strings is established to be two characters, subject to restrictions on two-character strings that would be likely to cause visual confusion in certain areas. This proposal does not allow for one-character TLD strings in any script, pending consideration by the GNSO and ccNSO.

Potential guidebook text based on this recommendation is included to help inform the discussion. The relevant section appears in Module 2 of the guidebook; see the full module at <http://icann.org/en/topics/new-gtlds/draft-evaluation-procedures-clean-04oct09-en.pdf>. Module 2 describes the various reviews that occur during the evaluation of an application, including review of each applied-for gTLD string to ensure that it complies with applicable rules and is not likely to have a negative impact on the stability of the DNS.

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.

### 2.1.1.3.2 String Requirements

ICANN will review each applied-for gTLD string to ensure that it complies with the requirements outlined in the following paragraphs.

If an applied-for gTLD string is found to violate any of these rules, the application will be denied. No further reviews are available.

**Part I -- Technical Requirements for all Labels (Strings)** – The technical requirements for top-level domain labels follow.

- 1.1 The ASCII label (i.e., the label as transmitted on the wire) must be valid as specified in technical standards *Domain Names: Implementation and Specification* (RFC 1035), and *Clarifications to the DNS Specification* (RFC 2181). This includes the following:
  - 1.1.1 The label must have no more than 63 characters.
  - 1.1.2 Upper and lower case characters are treated as identical.
- 1.2 The ASCII label must be a valid host name, as specified in the technical standards *DOD Internet Host Table Specification* (RFC 952), *Requirements for Internet Hosts — Application and Support* (RFC 1123), and *Application Techniques for Checking and Transformation of Names* (RFC 3696). This includes the following:
  - 1.2.1 The label must consist entirely of letters, digits and hyphens.
  - 1.2.2 The label must not start or end with a hyphen.
- 1.3 There must be no possibility for confusing an ASCII label for an IP address or other numerical identifier by application software. For example, representations such as “255”, “o377” (255 in octal) or “0xff” (255 in hexadecimal) as the top-level domain can be interpreted as IP addresses. As such, labels:
  - 1.3.1 Must not be wholly comprised of digits between “0” and “9”.
  - 1.3.2 Must not commence with “0x” or “x,” and have the remainder of the label wholly comprised of hexadecimal digits, “0” to “9” and “a” through “f.”
  - 1.3.3 Must not commence with “0o” or “o,” and have the remainder of the label wholly comprised of digits between “0” and “7”.
- 1.4 The ASCII label may only include hyphens in the third and fourth position if it represents a valid internationalized

domain name in its A-label form (ASCII encoding as described in Part II).

- 1.5 The presentation format of the domain (i.e., either the label for ASCII domains, or the U-label for internationalized domain names) must not begin or end with a digit.<sup>1</sup>

**Part II -- Requirements for Internationalized Domain Names** – These requirements apply only to prospective top-level domains that contain non-ASCII characters. Applicants for these internationalized top-level domain labels are expected to be familiar with the IETF IDNA standards, Unicode standards, and the terminology associated with Internationalized Domain Names.

- 2.1 The label must be a valid internationalized domain name, as specified in *Internationalizing Domain Names in Applications* (RFC 3490). This includes the following, non-exhaustive, list of limitations:
  - 2.1.1 Must only contain Unicode code points that are defined as “Valid” in The Unicode Codepoints and IDNA (<http://icann.org/en/topics/idn/rfcs.htm>) and be accompanied by unambiguous contextual rules where necessary.<sup>2</sup>
  - 2.1.2 Must be fully compliant with Normalization Form C, as described in *Unicode Standard Annex #15: Unicode Normalization Forms*. See also examples in <http://unicode.org/faq/normalization.html>.
  - 2.1.3 Must consist entirely of characters with the same directional property.
- 2.2 The label must meet the relevant criteria of the ICANN *Guidelines for the Implementation of Internationalised Domain Names*. See

---

<sup>1</sup> The primary concern relating to the use of leading- or trailing-numeric labels is due to issues raised by bi-directional scripts when used in conjunction with those labels. Experience has shown that presentation behavior of strings with leading or trailing numbers in bi-directional contexts can be unexpected and can lead to user confusion. As such, a conservative approach is to disallow numerals leading or trailing top-level domain labels.

This concern also applies to all-numeric strings; however, a larger concern with those strings is the risk of confusion and software incompatibilities due to the fact that a top-level domain of all numbers could result in a domain name that is indistinguishable from an IP address. That is, if (for example) the top-level domain .151 were to be delegated, it would be problematic to programmatically determine whether the string “10.0.0.151” was an IP address or a domain name.

<sup>2</sup> It is expected that the IDNA2008 protocol will be completed and conversion tools will be available before the Application Submission period begins, and that labels will be checked for validity under IDNA2008. In this case, labels valid under the previous version of the protocol (IDNA2003) but not under IDNA2008 will not meet this element of the requirements. Labels that are valid under both versions of the protocol will meet this element of the requirements. Labels valid under IDNA2008 but not under IDNA2003 may meet the requirements; however, applicants are strongly advised to note that the duration of the transition period between the two protocols cannot presently be estimated nor guaranteed in any specific timeframe. The development of support for IDNA2008 in the broader software applications environment will occur gradually. During that time, TLD labels that are valid under IDNA2008, but not under IDNA2003, will have limited functionality.

<http://www.icann.org/en/topics/idn/implementation-guidelines.htm>. This includes the following, non-exhaustive, list of limitations:

- 2.2.1 All code points in a single label must be taken from the same script as determined by the Unicode Standard Annex #24: Unicode Script Property.
- 2.2.2 Exceptions to 2.2.1 are permissible for languages with established orthographies and conventions that require the commingled use of multiple scripts. However, even with this exception, visually confusable characters from different scripts will not be allowed to co-exist in a single set of permissible code points unless a corresponding policy and character table are clearly defined.

### ***Part III - Policy Requirements for Generic Top-Level Domains –***

These requirements apply to all prospective top-level domains applied for as gTLDs.

- 3.1 Applied-for gTLD strings in ASCII (i.e., strings consisting exclusively of LDH characters) must be composed of three or more visually distinct characters. Two-character ASCII strings are not permitted, to avoid conflicting with current and future country codes based on the ISO 3166-1 standard.
- 3.2 Applied-for gTLD strings in IDN scripts (i.e., strings in which the U-label includes at least one non-LDH character) must be composed of two or more visually distinct characters in the script, as appropriate. Note, however, that a two-character IDN string will not be approved if:
  - 3.2.1 It is visually similar to any possible two-character ASCII combination; or
  - 3.2.2 It is visually similar to any one-character label (in any script).<sup>3</sup>

---

<sup>3</sup> The Implementation Working Team report included an additional requirement that “No TLD may be labeled with a string that is a meaningful identifier of a country or territory listed in ISO 3166-1 without the express consent of the government or properly constituted authority of the designated country or territory.” As is noted in the report, the requirement for government approval or non-objection in the case of an application for a country or territory or other geographical name is in place for *all* gTLD applications. As such, it is not included in this section as an IDN string requirement. However, as detailed in the draft Applicant Guidebook, an application submitted for a geographical name without the required documentation will not pass the geographical names review. See section 2.1.1.4 at <http://icann.org/en/topics/new-gtlds/draft-evaluation-procedures-clean-04oct09-en.pdf> for the definitions and requirements concerning geographical names.

### 2.1.1.3.2 *String Requirements* (Redlined to show changes from Guidebook v3)

ICANN will review each applied-for gTLD string to ensure that it complies with the requirements outlined in the following paragraphs.

If an applied-for gTLD string is found to violate any of these rules, the application will be denied. No further reviews are available.

***Part I -- Technical Requirements for all Labels (Strings)*** – The technical requirements for top-level domain labels follow.

- 1.1 The ASCII label (i.e., the label as transmitted on the wire) must be valid as specified in technical standards *Domain Names: Implementation and Specification* (RFC 1035), and *Clarifications to the DNS Specification* (RFC 2181). This includes the following:
  - 1.3.4 The label must have no more than 63 characters.
  - 1.3.5 Upper and lower case characters are treated as identical.
- 1.4 The ASCII label must be a valid host name, as specified in the technical standards *DOD Internet Host Table Specification* (RFC 952), *Requirements for Internet Hosts — Application and Support* (RFC 1123), and *Application Techniques for Checking and Transformation of Names* (RFC 3696). This includes the following:
  - 1.4.1 The label must consist entirely of letters, digits and hyphens.
  - 1.4.2 The label must not start or end with a hyphen.
- 1.5 There must be no possibility for confusing an ASCII label for an IP address or other numerical identifier by application software. For example, representations such as “255”, “o377” (255 in octal) or “0xff” (255 in hexadecimal) as the top-level domain can be interpreted as IP addresses. As such, labels:
  - 1.5.1 Must not be wholly comprised of digits between “0” and “9”.
  - 1.5.2 Must not commence with “0x” or “x,” and have the remainder of the label wholly comprised of hexadecimal digits, “0” to “9” and “a” through “f.”
  - 1.5.3 Must not commence with “0o” or “o,” and have the remainder of the label wholly comprised of digits between “0” and “7”.

- 1.6 The ASCII label may only include hyphens in the third and fourth position if it represents a valid internationalized domain name in its A-label form (ASCII encoding as described in Part II).
- 1.7 The presentation format of the domain (i.e., either the label for ASCII domains, or the U-label for internationalized domain names) must not begin or end with a digit.<sup>1</sup>

**Part II -- Requirements for Internationalized Domain Names** – These requirements apply only to prospective top-level domains that contain non-ASCII characters. Applicants for these internationalized top-level domain labels are expected to be familiar with the IETF IDNA standards, Unicode standards, and the terminology associated with Internationalized Domain Names.

- 3.3 The label must be a valid internationalized domain name, as specified in *Internationalizing Domain Names in Applications* (RFC 3490). This includes the following, non-exhaustive, list of limitations:
  - 3.3.1 Must only contain Unicode code points that are defined as “Valid” in The Unicode Codepoints and IDNA (<http://icann.org/en/topics/idn/rfcs.htm>~~<http://tools.ietf.org/wg/idnabis/>~~), and be accompanied by unambiguous contextual rules where necessary.<sup>2</sup>
  - 3.3.2 Must be fully compliant with Normalization Form C, as described in *Unicode Standard Annex #15: Unicode Normalization Forms*. See also examples in <http://unicode.org/faq/normalization.html>.
  - 3.3.3 Must consist entirely of characters with the same directional property.

---

<sup>1</sup> The primary concern relating to the use of leading- or trailing-numeric labels is due to issues raised by bi-directional scripts when used in conjunction with those labels. Experience has shown that presentation behavior of strings with leading or trailing numbers in bi-directional contexts can be unexpected and can lead to user confusion. As such, a conservative approach is to disallow numerals leading or trailing top-level domain labels.

This concern also applies to all-numeric strings; however, a larger concern with those strings is the risk of confusion and software incompatibilities due to the fact that a top-level domain of all numbers could result in a domain name that is indistinguishable from an IP address. That is, if (for example) the top-level domain .151 were to be delegated, it would be problematic to programmatically determine whether the string “10.0.0.151” was an IP address or a domain name.

<sup>2</sup> It is expected that the IDNA2008 protocol will be completed and conversion tools will be available before the Application Submission period begins, and that labels will be checked for validity under IDNA2008. In this case, labels valid under the previous version of the protocol (IDNA2003) but not under IDNA2008 will not meet this element of the requirements. Labels that are valid under both versions of the protocol will meet this element of the requirements. Labels valid under IDNA2008 but not under IDNA2003 may meet the requirements; however, applicants are strongly advised to note that the duration of the transition period between the two protocols cannot presently be estimated nor guaranteed in any specific timeframe. The development of support for IDNA2008 in the broader software applications environment will occur gradually. During that time, TLD labels that are valid under IDNA2008, but not under IDNA2003, will have limited functionality.

- 3.4 The label must meet the relevant criteria of the ICANN *Guidelines for the Implementation of Internationalised Domain Names*. See <http://www.icann.org/en/topics/idn/implementation-guidelines.htm>. This includes the following, non-exhaustive, list of limitations:
- 3.4.1 All code points in a single label must be taken from the same script as determined by the Unicode Standard Annex #24: Unicode Script Property.
- 3.4.2 Exceptions to 2.2.1 are permissible for languages with established orthographies and conventions that require the commingled use of multiple scripts. However, even with this exception, visually confusable characters from different scripts will not be allowed to co-exist in a single set of permissible code points unless a corresponding policy and character table are clearly defined.

**-Part III -- Policy Requirements for Generic Top-Level Domains – These requirements apply to all prospective top-level domains applied for as gTLDs.**

3.1 Applied-for gTLD strings in ASCII (i.e., strings consisting exclusively of LDH characters) must be composed of three or more visually distinct letters or characters, in the script, as appropriate.<sup>3</sup> Two-character ASCII strings are not permitted, to avoid conflicting with current and future country codes based on the ISO 3166-1 standard.

3.2 Applied-for gTLD strings in IDN scripts (i.e., strings in which the U-label includes at least one non-LDH character) must be composed of two or more visually distinct characters in the script, as appropriate. Note, however, that a two-character IDN string will not be approved if:

3.2.1 It is visually similar to any possible two-character ASCII combination; or

3.2.2 It is visually similar to any one-character label (in any script).<sup>4</sup>

<sup>3</sup>The requirement for gTLD strings to consist of at least three visually distinct characters remains under discussion. An implementation support team of technical and linguistic experts is currently engaging in work on a proposed solution to enable gTLDs of fewer than three characters where appropriate. The proposed solutions will then be made available for public comment.

<sup>4</sup> The Implementation Working Team report included an additional requirement that “No TLD may be labeled with a string that is a meaningful identifier of a country or territory listed in ISO 3166-1 without the express consent of the government or properly constituted authority of the designated country or territory.” As is noted in the report, the requirement for government approval or non-objection in the case of an application for a country or territory or other geographical name is in place for all gTLD applications. As such, it is not included in this section as an IDN string requirement. However, as detailed in the draft Applicant Guidebook, an application submitted for a geographical name without the required documentation will not pass the geographical names review. See section 2.1.1.4 at <http://icann.org/en/topics/new-gtlds/draft-evaluation-procedures-clean-04oct09-en.pdf> for the definitions and requirements concerning geographical names.