Update to Module 2: String Requirements

30 May 2009

This section appears in Module 2; see the full module at http://www.icann.org/en/topics/new-gtlds/draft-evaluation-procedures-clean-18feb09-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.

The potential new language highlighted in this section is based on public comments (see analysis of public comments on draft Applicant Guidebook v2) and continuing development work by staff. Clarifications have been included to provide useful direction to applicants.

The requirement for gTLD strings to consist of at least three visually distinct characters remains under discussion. For proposals and additional details on the discussion, see the explanatory memorandum entitled "Discussions about the 3-Character String Requirement."

ICANN encourages comment on the interim 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 3 of the full draft Applicant Guidebook, scheduled to be published in September 2009.

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 the selection of 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 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 with an IP address or other numerical identifier. 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 Unicode label for Internationalized Domain Names) must not begin or end with a digit.

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 (Internet Draft "draft-faltstrom-idnabis-tables"), and be accompanied by unambiguous contextual rules where necessary.

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 Internationalized Domain Names. See 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.

The IDNA protocol used for internationalized labels is currently under revision through the Internet standardization process. As such, additional requirements may be specified that need to be adhered to as this revision is being completed. The current status of the protocol revision is documented at http://tools.ietf.org/wg/idnabis.

Regarding Leading-, Trailing-, or All Numeric-TLDs – 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.

Policy Requirements for Generic Top-Level Domains – Applied-for gTLD strings must be composed of three or more visually distinct letters or characters in the script, as appropriate.

2.1.1.3.2 *String Requirements* ***(Redlined to Show Changes from Guidebook v2)***

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 the selection of 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 ~~the~~ technical standards Domain Names: Implementation and Specification (RFC 1035), and Clarifications to the DNS Specification (RFC 2181). This includes the following:

  - ~~o~~ 1.1.1 The label must have no more than 63 characters. ~~In the case of Punycode (IDNA2008 A-label) representations of IDN labels (U-labels), this includes the four initial characters (xn--).~~

    - ~~o~~ 1.1.2 Upper and lower case characters are treated as~~considered to be syntactically and semantically~~ 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:

  - ~~o~~ 1.2.1 The label must consist entirely of letters, digits and hyphens.

  - ~~o~~ 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] ~~. Therefore an ASCII label must not be:~~

- ~~o~~ 1.3.1 Must not be wholly comprised of digits between~~A decimal number consisting entirely of the digits~~ "0" and~~through~~ "9".~~;~~

  ~~o~~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". ~~A hexadecimal number consisting of the digit "0" followed by the uppercase or lowercase letter "x||X" followed by a sequence of one or more characters all of which belong to the set of uppercase or lowercase letters "a||A" through "f||F" and the digits "0" through "9"; or~~

  ~~o~~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".~~An octal number consisting of the uppercase or lowercase letter "o||O" followed by a sequence of one or more characters all of which belong to the set of digits "0" through "7."~~

~~•~~1.4 The ASCII label may only include hyphens in the third and fourth position if it represents a valid ~~i~~Internationalized d~~D~~omain n~~N~~ame 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 ~~i~~Internationalized d~~D~~omain n~~N~~ames) must not begin or end with a digit.

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.

[1] ~~Refer to http://www.icann.org/en/topics/new-gtlds/update-dns-stability-18feb09-en.pdf for further background on octal and hexadecimal representations, and on the changes in this section.~~

- 2.1 The label must be a valid internationalized domain name, as specified in ~~the technical standard~~ Internationalizing Domain Names in Applications (RFC 3490)~~, or any revisions of this technical standard currently underway within the IETF. Due to this ongoing revision, the IDN-related technical requirements are subject to change. This includes, but is not limited to, the following constraints. Note that these are guidelines and are not a complete statement of the requirements of the IDNA specifications. The label~~ This includes the following, non-exhaustive, list of limitations:

    ~~o~~2.1.1 Must only contain Unicode code points that are defined as "~~Protocol~~ Valid" ~~or "Contextual Rule Required"~~ in The Unicode Codepoints and IDNA (Internet Draft "draft-faltstrom-idnabis-tables)~~http://www.ietf.org/internet-drafts/draft-ietf-idnabis-tables-05.txt)~~, and be~~that are accompanied, in the case of "Contextual Rule Required," by~~accompanied by unambiguous contextual rules where necessary.

    ~~o~~2.1.2 Must be fully compliant with Normalization Form C, as described in Unicode Standard Annex #15: Unicode Normalization Forms. S~~(s~~ee also examples in http://unicode.org/faq/normalization.html~~)~~.

    ~~o~~2.1.3 Must consist entirely of characters with the same directional property. ~~(Note that this requirement may change with the revision of the IDNA protocol to allow for characters with no directional property defined in Unicode to be available along with either a right-to-left or a left-to-right directionality.)~~

- 2.2 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:

    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. ~~Exceptions 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.~~

~~o~~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.

The IDNA protocol used for internationalized labels is currently under revision through the Internet standardization process. As such, additional requirements may be specified that need to be adhered to as this revision is being completed. The current status of the protocol revision is documented at http://tools.ietf.org/wg/idnabis.

Regarding Leading-, Trailing-, or All Numeric-TLDs – 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.

Policy Requirements for Generic Top-Level Domains – Applied-for gTLD strings must be composed of three or more visually distinct letters or characters in the script, as appropriate.[2]

[2] ICANN received a number of comments suggesting that gTLDs consisting of fewer than three characters should be allowed in some cases, for example, in scripts featuring ideographs. The issues with defining requirements for certain cases are discussed in further detail in the Public Comment Analysis at http://www.icann.org/en/topics/new-gtlds/agv1-analysis-public-comments-18feb09-en.pdf and ICANN invites further input on solutions.