



New gTLD Program Explanatory Memorandum

Discussions about the 3-char String Requirement

Date of Publication:

30 May 2009

Background - New gTLD Program

Since ICANN was founded ten years ago as a not-for-profit, multi-stakeholder organization dedicated to coordinating the Internet's addressing system, one of its foundational principles, recognized by the United States and other governments, has been to promote competition in the domain-name marketplace while ensuring Internet security and stability. The expansion will allow for more innovation, choice and change to the Internet's addressing system, now constrained by only 21 generic top-level domain names. In a world with 1.5 billion Internet users—and growing—diversity, choice and competition are key to the continued success and reach of the global network.

The decision to launch these coming new gTLD application rounds followed a detailed and lengthy consultation process with all constituencies of the global Internet community. Representatives from a wide variety of stakeholders—governments, individuals, civil society, business and intellectual property constituencies, and the technology community—were engaged in discussions for more than 18 months. In October 2007, the Generic Names Supporting Organization (GNSO)—one of the groups that coordinate global Internet policy at ICANN—completed its policy development work on new gTLDs and approved a set of recommendations. The culmination of this policy development process was a decision by the ICANN Board of Directors to adopt the community-developed policy in June 2008 at the ICANN meeting in Paris. A thorough brief to the policy process and outcomes can be found at <http://gnso.icann.org/issues/new-gtlds/>.

This paper is part of a series of papers that will serve as explanatory memoranda published by ICANN to assist the Internet community to better understand the Request for Proposal (RFP), also known as *Applicant Guidebook*. A public comment period for the Applicant Guidebook will allow for detailed review and input to be made by the Internet community. Those comments will then be used to revise the documents in preparation of a final Applicant Guidebook. ICANN will release the final Applicant Guidebook and open the application process in the first half of 2010. For current information, timelines and activities related to the New gTLD Program, please go to <http://www.icann.org/en/topics/new-gtld-program.htm>.

Please note that this is a discussion draft only. Potential applicants should not rely on any of the proposed details of the new gTLD program as the program remains subject to further consultation and revision.

Summary of Key Points in this Paper

- Historically gTLDs have consisted of 3 or more characters. Whereas ccTLDs have been 2-character combinations based on the ISO3166-1 list.
- In the process of introducing new gTLDs, the Applicant Guidebook (version 2.0) states that *"Applied-for strings must be composed of three or more visually distinct letters or characters in the script, as appropriate"*.
- Several comments received express the need for allowing gTLDs consisting of less than 3 characters.
- The paper is outlining a proposed solution from the community that would allow for less than 3 characters in gTLD string in certain cases, and it details the pro's and con's for the proposed solution.

The gTLD Applicant Guidebook (version 2.0) states that *"Applied-for strings must be composed of three or more visually distinct letters or characters in the script, as appropriate"*. It also notes that ICANN has received a number of comments on this topic, of which some are request that gTLDs consisting of fewer than three characters should be allowed in some cases. The comment recommends that such cases would be in scripts featuring ideographs where single-characters or two-character combinations express a complete word (generic or geographical term).

It is asserted in the public comment that the requirement for a minimum of three characters in a gTLD domain name does not work for many scripts because one or two "syllables" in certain scripts can represent a meaningful word. For example, these concerns exist for Chinese, Japanese and Korean. Without single and two character URLs on the top level, introduction of IDN gTLD will be meaningless for Internet users in these scripts; a possible way forward is to make a consistent exception for some scripts such as Chinese, Japanese and Korean.

The issues encountered in the development of version 2.0 of the Guidebook with defining these requirements and providing an adequate method for implementation are discussed in the Public Comments Analysis at <http://www.icann.org/en/topics/newgtlds/agv1-analysis-public-comments-18feb09-en.pdf>.

As the Public Comment Analysis paper explains, the primary concern with allowing strings with less than 3 characters is that this could create user confusion with the existing ccTLDs, or any future ccTLD additions based on a possible expansion of the ISO-3166-1 list. However, it is at the same time clear that certain characters (in different scripts) would not cause confusion.

The GNSO Policy recommendation to ICANN was to allocate single and two-character strings in certain scripts through a case-by-case review. As implementation details were worked out, it was found difficult to identify a clear set of rules to decide which scripts would be allowed based upon ideographic nature. The rules would be used for a case-

by-case review of proposed strings of less than 3 characters. Hence, ICANN requested further 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, not creating any stability or security problems for the DNS.

This paper describes the comment for and against relaxing the requirement of a minimum of 3 characters in gTLD strings. The comment for relaxation in certain scripts where one and two characters often express complete words such as languages in the CJK scripts, also included a proposed test. ICANN is posting this material to generate more community discussion in preparations for the ICANN meeting in Sydney, Australia, 21-26 June 2009. In order for a solution to be acceptable it will need to set clearly understandable guidelines for when strings less than 3 characters long will be allowed as TLDs. The solution must be technical feasible and address valid concerns such as potential for user confusion.

Proposed Solution for allowing gTLD strings less than 3 characters long

The following comments and proposals were received through public comments and consultations at and following the ICANN meeting in Mexico City, March 2009.

“ICANN should follow the GNSO principle recommendations and lift the restriction on the length of an IDN TLD (i.e. more than two characters) or modify the clause to make it become script specific. Without change the restriction would be a significant deterrent for Chinese TLDs because most meaningful Chinese words are composed only by 2 Chinese characters. ”

The same sort of comment was received regarding Japanese and Korean characters.

The gTLD registry constituency stated that, “An exception should be allowed for Chinese, Japanese and Korean scripts to the 3 or more characters string requirement.”

There were several proposals made for testing whether single and two-character registrations should be allowed for certain scripts.

One recommendation stated:

If a majority of the Unicode characters of the writing system for a particular language possess a meaning on their own, then the restriction of 3 or more characters should not be applied. The applicant should specify the classification of the writing systems of the string they are applying for, namely Logographic, Syllabic, Alphabetic, Abugida, Abjad and Featural. The rule should apply on a per-string basis and not based on the language (e.g., Japanese hiragana string may still be restricted to 3 or more characters whereas a Japanese kanji string may be allowed on its own). J. Seng (13 April 2009).

Another:

The 3 character requirement should be lifted from strings whose writing system employ basic building blocks that have generally accepted semantic associations, where single and two-character sequences represent concepts in their own right without the need for abbreviation. These systems do not remotely resemble Latin so visual confusability will not be an issue (there’s string review for

that). The character repertoire for these scripts is orders of magnitude larger than that of alphabetic or syllabic scripts (e.g., 71,442 Han characters in Unicode version 3.2 versus 26 English alphabets). W. Tan (13 April 2009).

The most substantive test received:

For gTLD strings with less than 3 characters the following must be the case:

1. *Upper case alphabetic characters are not allowed - in the first rule, all upper case alphabetic characters are excluded. Any strings containing one or more character with the Unicode property "lu" will not be allowed as a TLD.*
2. *Lower case alphabetic characters are not allowed - in the second rule, all lower case alphabetic characters are excluded. Any strings containing one or more character with the Unicode property "ll" will not be allowed as a TLD.*

Rule 1 and 2 together eliminates almost all characters present in Unicode today that are visually confusable with the ASCII characters that are used in ccTLD strings. However, some corner cases are left for further evaluation.

3. *Strings that are not eliminated due to rule #1 or rule #2, but are less than 3 characters long need to go through a check for confusability with any of the basic Latin characters (a,b,...,z), to avoid conflict with any existing ccTLDs or any future additions to the ISO3166-1 list.*

The same rule can be expressed in different ways, such as for example:

1. *Strings containing characters that fall into either one of the below listed categories of Unicode Properties are considered string containing a "alphabetical writing system", and will be subject to 3-char limitation. Otherwise, it is a non-alphabetical writing system and therefore will not subject to a 3-char limitation.*
 - *General Category Values = L& (ie, Lu, Ll & Lt) [This will cover Latin based script]; or*
 - *PropertyList [<http://unicode.org/Public/UNIDATA/PropList.txt>] = Other_Alphabetic or Other_Lowercase; [This will cover non-Latin based, but alphabetic script, like Arabic, Hebrew, Sanskrit etc]*
2. *If a string contains characters with 'PropertyList [<http://unicode.org/Public/UNIDATA/PropList.txt>] = Ideographic', then it is not subjected to a 3-char limitation.*

This public comment suggests that limiting registrations to 3-characters per string or more would seriously hobble usage in these regions and the comment also offers concrete suggestions regarding a solution.

Opposing arguments that requires gTLDs to be a minimum of 3 characters

The opposing arguments fall into the following categories:

Fairness of treatment:

In addition to comments received from the CJK community, ICANN received arguments from the European region that certain single or two character combinations in European languages represent a word or a meaning and in some cases also geographic identifiers. These arguments were made to counter arguments for allowing less than 3 character strings per the proposal above. If less than 3 character strings are allowed for CJK based languages then they should also be allowed for other languages, for fairness of treatment.

Statements concerning the Chinese words and number of characters

Other arguments state that few Chinese characters are words, most Chinese words are two or more characters. If one separates out the specific phonetic implications (one character equals one phoneme), characterizing Chinese characters as syllables would be much more accurate. In addition, some opinions are that most Chinese words consist of more than one character.

Trial of a few gTLDs with less than 3 characters

Some suggestions have been made that ICANN perform a trial implementation of a certain small number of gTLDs that have less than 3 characters. This would then be used to inform the development of the process for allocating such strings more widely.

Translations of TLDs

Comments have been received that in relation to translation of existing TLDs, there has never been a model for "translating" TLDs. Therefore, the two-letter ISO codes or other TLDs that are "abbreviations" cannot be translated into IDN strings of less than 3 characters as a meaningful representation of that definition. They are not standardized abbreviations and abbreviations are not a standard concept across languages and cultures. The ccTLDs in particular are a standardized coding system, chosen as codes for a number of reasons including recognizability and distinctiveness of undecorated Latin character.

ICANN's ccTLD delegation function

Currently the IANA delegation function relies on the scarce availability of 2-character ASCII combinations and all of these (when entered in the ISO3166-1 list) are treated as ccTLDs. Discussions in the ICANN community as well as at ISO MA meetings in the past few years have focused on the feasibility of expanding the ISO3166-1 list to contain 2-character combinations of other scripts, representing country and territory names. This would require a multi-year ISO led development (which might occur after the ccNSO IDN PDP). The outcome of the Fast Track Process will inform the ongoing discussion about whether or how to expand the ISO3166-1 list and associated ICANN ccTLD delegation function, as well as the long-term ccNSO PDP for IDN ccTLDs.

Delegation of single and two-character labels now, might jeopardize the future shape of the ccTLD delegation mechanism.

Conclusion

The ICANN community continues to work on suggested solutions, balancing of the needs of the community for broad usage against the community needs for reserving potential uses by ISO standards. This should be an extensive and broad community discussion.

Discussion should continue in order to preserve the fundamental obligation to keep the Internet open for innovation.

ICANN invites the community to continue the dialogue and explore whether these or other potential solutions can be found to solve this problem.