New gTLD Program
Explanatory Memorandum

DNS Stability

Additional Technical Criteria, Including IDNs (Update from 22 Oct 2008 version)

Date of Publication: 17 February 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. Contributing to this policy work were ICANN's Governmental Advisory Committee (GAC), At-Large Advisory Committee (ALAC), Country Code Names Supporting Organization (ccNSO) and Security and Stability Advisory Committee (SSAC). 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 RFP 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 RFP. ICANN will release the final RFP in the first half of 2009. 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

- The ASCII representation (LDH or IDNA2008 A-label) of a DNS label may contain no more than 63 characters. It must consist entirely of letters, digits, and hyphens; must not start or end with a hyphen or a digit; must not be confusable with an IP address; and must not be a decimal, hexadecimal, or octal numeric string.
- All applied-for ASCII TLD strings must meet the technical requirements in *Names: Implementation and Specification* (RFC 1035) and *Clarifications to the DNS Specification* (RFC 2181).
- All applied-for non-ASCII TLD strings (Internationalized Domain Name TLDs) must meet the technical requirements in *Internationalizing Domain Names in Applications* (RFC 3490).
- The applied-for IDN TLD string also must meet the relevant criteria of the ICANN *Guidelines for the Implementation of Internationalized Domain Names*.
- The IDNA protocol used for IDN TLDs is undergoing revision through the Internet standardization process. Thus, additional requirements may be specified or requirements here may change or be removed as the protocol revision is being completed.

A Introduction

This document revises and supersedes the 22 October 2008 Explanatory Memorandum *Update to DNS Stability Paper—Additional Technical Criteria Requirements, Including IDNs*.
It supplements the description and analysis of DNS stability issues contained in *DNS Stability: The Effect of New Generic Top Level Domains on the Internet Domain Name System*, which was published by ICANN for Public Comments on 6 February 2008, by providing a concise summary of the technical criteria for new gTLD strings augmented by additional detail and clarification.

The criteria specified in this document proscribe strings that might cause technical instability in the DNS. It is important to distinguish between the stability of the DNS as a technical enterprise, in which computer systems, organizations, protocols, and other components manage the relationship between domain names and numeric IP addresses, and the stability of the DNS as a social and economic system, in which domain names are signifiers with a variety of non-technical meanings for human users. The effect of new TLDs on the social and economic stability of the DNS is beyond the scope of this paper.

As it covers only technical criteria related to DNS stability, this document does not represent a complete specification of all of the requirements that must be met by a proposed new gTLD, nor does it include disqualifications relating to reserved words (including reserved words for technical reasons, *e.g.* "localhost") or other policy–related

reasons.

Important Information Regarding IDN Requirements

The IDNA protocol used for internationalized labels is currently under revision through the Internet standardization process. As such, additional requirements may be specified or requirements specified here may change or be removed as the protocol revision is being completed. The current status of the protocol revision is documented at http://tools.ietf.org/wg/idnabis/ and additionally updated standards will be referenced at http://www.icann.org/en/topics/idn/rfcs.htm.

B Technical Requirements

1. Requirements for All Labels

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 may contain 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--).

1.1.2 Upper and lower case characters are considered to be syntactically and semantically identical.

1.2 The ASCII label must be a valid host name, as specified in *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 of 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. Therefore an ASCII label must not be:

1.3.1 a decimal number consisting entirely of the digits "0" through "9";

1.3.2 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

1.3.3 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 include hyphens in the third and fourth position only if it represents a valid internationalized domain name in its A-label form (ASCII encoding as described in Section 2).

1.5 The presentation format of the label (*i.e.*, either the ASCII label for LDH TLDs, or the Unicode label (U-label) for Internationalised Domain Names) must not begin or end with a digit.

2. Requirements for Internationalized Top-Level Labels

These requirements apply only to top-level domain labels that contain non-ASCII characters ("internationalized top-level labels"). Applicants for these internationalized top-level 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 (a) *Internationalizing Domain Names in Applications* (RFC 3490), until its replacement has completed its journey through the IETF standardization process, or (b) *Internationalized Domain Names in Applications* (currently an Internet Draft) thereafter. This includes the following, non-exhaustive, list of limitations: the label must

2.1.1 contain only Unicode code points that are defined as "Protocol Valid" or "Contextual Rule Required" in *The Unicode Codepoints and IDNA* (currently an Internet Draft), and that are accompanied, in the case of "Contextual Rule Required," by unambiguous contextual rules;

2.1.2 be fully compliant with Normalization Form C, as described in *Unicode Standard Annex #15: Unicode Normalization Forms* (see examples in http://unicode.org/faq/normalization.html); and

2.1.3 conform to *An Updated IDNA Criterion for Right-to-Left Scripts* (currently an Internet Draft).

2.2 The label must meet the relevant criteria of the ICANN *Guidelines for the Implementation of Internationalized Domain Names*. 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.

C Explanatory Rationales

1. Decimal, hexadecimal, and octal strings

Although IPv4 addresses are most often represented in the "dotted quad" format, in which the 4 bytes of the 32-bit address appear as four decimal numbers, each separated from its neighbor by a "dot" (recognized in an English-language context as a "full stop" or "period" character), they can also be represented in other ways as decimal, hexadecimal, and octal strings. TLD labels consisting entirely of decimal,

hexadecimal, or octal strings are proscribed in order to prevent confusion of domain names and IPv4 addresses in contexts in which either might appear (for example, in the user interface of web browsers and many other applications). Rules 1.3.2 and 1.3.3 specify a particular representation of hexadecimal and octal strings because those are the forms that are permitted in representations of IPv4 addresses.

IPv6 addresses are most often represented in a "colon-hex" format that uses the colon (:) character rather than the "dot" to delimit 8 groups of 4 hexadecimal digits. The distinction between the colon-hex representation of an IPv6 address and a domain name relies on the use of the colon in the former and the "dot" in the latter. IPv6 addresses may be represented in other forms, using "zero suppression" to shorten the written representation of an IPv6 address or a mix of colon-hex in the high-order 6 groups and dotted-quad in the low-order 2 groups (4 bytes) to facilitate working in mixed-stack environments, but at least one colon character appears in all cases. Note that it would not be practical simply to proscribe any string that could represent a hexadecimal number, other than in the form described in 1.3.3. The string "beef", for example, could be interpreted as a hexadecimal number; 1.3.3 proscribes it only in the form "0xbeef".

The least restrictive approach, of course, would be to proscribe only those all-digit strings that could possibly be mistaken for an IP address. One comment, for example, points to the fact that the decimal value of each byte in a "dotted quad" IP address cannot exceed 255, and suggests that a sequence of all-digit labels should be acceptable as a domain name as long as at least one of the labels in the sequence represented a decimal value greater than 255. However, some software will recognize a sequence of decimal numbers as an IP address even if the value of one or more of the numbers is greater than 255, because it looks only at the low-order 8 bits and ignores the rest.

2. Domain names and file extensions

If a domain name looks too much like a file name, there is a chance that applications (or even human users) could confuse them. A web browser encountering a string ending in ".mp3", for example, might not be able to tell from the context whether the user's intention was to enter a web site URL or a music file name. (Because it has become common practice for web browsers and other applications to accept incompletely specified identifiers in user interfaces, "filling in the blanks" by making assumptions about the user's intentions, it is impractical to suggest that user interfaces simply refuse to accept domain names or file names entered without the explicit "http://" or "file://" preamble.) Citing this risk of confusion, many people have asked ICANN to prohibit the use of "common file extensions" as TLD labels.

The difficulty with such a prohibition, of course, is that any list of strings that are, have been, or might some day be used as file extensions would be very long and could not be authoritative, as no standard for file extensions exists (or is anticipated). In the case of country codes, for example, ISO 3166 is authoritative for what is and is not a country code, and a well-recognized maintenance agency is responsible for dealing with changes to the standard. No such system exists for file extensions.

It has been suggested that ICANN could at least proscribe the "most common" file extensions, such as "exe", "pdf", and "jpg". However, it should be obvious that different people with different perspectives and interests would compile different lists of the "most common" file extensions; and some of those lists might include extensions that are

already in use as gTLDs, such as "com". ICANN has no authority (much less inclination) to assert that one file extension is "common"—and should therefore not be permitted as a TLD label—while another file extension is not.