IDN Application Evaluation Facilities

Draft posted for public comments and discussions through 31 July 2007

Final amendment, 14 August 2007: Revision history on p. 15

Executive Summary
This document describes a plan for a publicly accessible facility that 1) provides for the insertion of temporary IDN TLDs (delegated A-labels) into the root zone and 2) enables end-users to evaluate the response of commonly used software applications to domain names with U-labels in all positions, corresponding to A-labels that have been included in TLD root zones of the DNS.

Table of Contents

1 General Overview: Evaluation of IDN TLDs in the root zone

This paper describes ICANN's plan to establish a public-benefit facility that will allow end-users, governments, incumbent and prospective top-level domain registries, and other parties interested in the global deployment of IDNs to evaluate the response of software applications to IDNs with localized labels on all levels, including the top-level. It also provides a channel for public feedback about the results of such evaluation from the IDN end-user perspective.

This plan is part of an overall program that focuses on enabling the routine introduction of TLDs with IDN labels. Parallel elements of this program include: the SSAC IDN study, the proposed IDNA protocol revision, and the development of policies for the introduction of IDN TLDs by the ccNSO, GNSO, and the GAC.

Consultation

There has been an on-going discussion on the topic of live tests, pre-deployment tests, technical laboratory tests, and application-level evaluations during the IDN workshops at the past years' ICANN meetings, during the meetings of the President's Advisory Committee on IDNs, during meetings held separately with the ICANN Board and liaisons to IETF, SSAC, and RSSAC, as well as during many ad-hoc meetings with individuals within the ICANN community who support ICANN's continued efforts at making internationalized top level labels available in the production root zone environment.

This report and evaluation plan considers all those contributions and input.

Overall Principles

The principles for the evaluation facilities described in this plan place the highest priority on preserving the stability and security of the DNS. Secondarily, but also of prime importance, the program seeks to avoid creating any perceived entitlement for long-term retention of any TLDs used for purposes of evaluation or in any time-limited regard. As such, the plan focuses primarily on:

- Ensuring that all evaluation takes place in the safe context of:
  - Ongoing DNS stability and security,
  - Creating names that are purely transient objects, with no expectation of a permanent delegation or registration.

- Ensuring that the scope and size of the evaluation remains manageable, and does not suddenly expand in a manner that jeopardizes project control or DNS stability.

- Ensuring all necessary domain name management requirements (for example string length), will be defined and remain in effect throughout the entire evaluation.

2   Success Criteria

The expectations to the result of the evaluations follow these success criteria (non-prioritized):

- Insertion and resolution of IDN domain names in the root zone will be done with negative impacts to the DNS (i.e., with less than the limits described in the *Tolerance specification*).

- Evaluation results will inform the Board discussion, resolve some of the technical concerns and contribute to the resolution of some of the technical discussion regarding the deployment of IDN TLDs.

- Documentation of user experiences with a core set of applications in which end-expect to use their internationalized domain names, (and whether those applications meet user expectations) will be reported in online tools to help understand potential issues at the application level.

- The application software review will be useful for any prospective applicant for an IDN TLD when considering issues they may encounter when IDNs are deployed in a production environment.

- The application software review is documented in sufficient detail for ICANN to communicate useful issues and reports to application developers.

3   Objectives and General Principles

Two separate evaluation facilities will be launched; both will include the entry of a limited number of resource records in the root zone of the DNS. There will be approximately 20 TLD labels (in 20 different languages selected to illustrate as many scripts as possible) representing the lexeme *test* in 20 different languages. (See Appendix B).

The selected labels are based on: an assessment of the number of language/script combinations that are frequent subjects of recent discussions of top-level internationalization; and the statements of interest that have already been made. If the announcement of the evaluation facility generates an increase in the number of languages requested for inclusion, some increase in the number of scripts can be accommodated. If there are many additional requests, it may be necessary to re-evaluate some aspects of the evaluation registry.

The initial deployment will be seeded with the labels described in Section 3.4, (and listed in Appendix B) with the expectation that a small additional number will be added. Requests made subsequent to the initial deployment based on that list will be queued for later release as is feasible given the status of the evaluation effort.

3.1 Evaluation Descriptions and Objectives

There are two separate evaluations planned that will be started in parallel.

Pre-Deployment Evaluation Facility

Description: This initial evaluation is a straightforward replication of the queries run by Autonomica in their successful laboratory environment exercise. In this case (in the live root) the evaluation will be conducted using the *.test* TLDs in various scripts. RSSAC will be advised formally of this activity no later than 2 weeks prior to the insertion of the requisite strings.

Objective: Demonstrate that there is no deleterious effect to the stability and security of the DNS caused by the introduction of a limited number of IDN TLDs. (This result was initially demonstrated in the laboratory test conducted by Autonomica.)

While the test protocol will be based on the laboratory regimen, it will extend it to determine the response of root name servers, DNS resolvers, and also of the broader battery of software applications that need to be IDN aware. A notification containing the test protocol will be issued to top-level registries, governments, and any parties otherwise interested in running parallel evaluations. A complete definition of the evaluation criteria will be referenced in the open call for volunteer participation.

The root-server operators will be asked on a volunteer basis to participate in the evaluations by monitoring and measuring the traffic generated against the *.test* TLDs. Root-server participation is described in detail in Section 3.2.

The test protocol will be published openly, and anyone interested in running it in parallel to Autonomica will be encouraged to do so. A wiki page will be set up reporting of independent results in whatever language the participants prefer.

Application Software Evaluation Facility

Description: In this phase, the operator of the *.test* zones will insert a second-level domain in each of them with an A-label corresponding to the word *example* in the same language and script used for the top-level label. The origin of the *example* and *test* labels is described in detail in the text that established the label criteria for the successfully concluded laboratory test, testing the response of root name servers and iterative mode resolvers to A-labels of varying length. The test planned in the present document is the natural follow-on to that test. Relevant portions of the previous describing string selection are excerpted in Appendix A.

The *example.test* second-level domains will all be delegated to a single operator as a control to the evaluation. All registration related activity will be based on a single

platform, thus ensuring the statistical viability of the comparison of the results. The delegation of third-level domains is not an immediate component of this evaluation.[1]

If a plan for abetting registrar preparation of their own systems (outside that made available in TLD registry operators' testing environments) is determined to be necessary as a result of this testing or other impact, this may be framed as an additional evaluation phase upon the conclusion of the present plan.

Objective: The *example.test* domains are intended to allow anyone interested in assessing a domain name in one of the selected scripts to do so in the context of their actual daily work, research, or other self-defined scenario. Each of the *example.test* domains will resolve to an IP address at which a wiki will be operated in one or several of the languages using that script. The A-records that support this delegation will be parallel with MX-records associated with a facility for the short term display of e-mail sent to a special address in the evaluation domain. The wiki will also support the e-mail functionality and will provide users with: a target for both HTTP and SMTP calls, and a platform for describing the positive and negative aspects of that action.

These results will be publicly available for: consideration by prospective applicants for IDN TLDs intended to serve specific speech communities, support of the registrar assessment of market demographics, and consideration by software developers involved in the internationalization of their products.

3.2   Acceptable Limitations in Functionality (draft)

On 2 June 2007 ICANN posted a set of draft procedures describing how IANA will manage the insertion, administration and removal of internationalized top-level labels (as delegated A-records) in the DNS root zone. These procedures were developed to guide this evaluation plan. The announcement included the draft procedure itself (IANA Root Zone Procedures for Test IDN Deployment, and a draft paper (IDN TLD Root Server Performance / Tolerance) describing the negative impacts to the DNS (tolerance measure) that, if observed, would trigger an emergency revocation procedure. The revocation procedure provides for the rapid removal of IDNs from the root zone if these negative impacts rise to the level specified in the tolerance document. The Root Server System Advisory Committee (RSSAC) is also analyzing the tolerance measure and their expertise will be considered in finalizing these procedures.

The draft procedures were posted for public comment ending on June 22, 2007 and the final document is located at http://icann.org/announcements/announcement-06aug07.htm.

---

[1] It is, however, conceivable that a requirement for third level registrations under *example.test* will arise. This is unlikely to tax the underlying registries, but would pose need for specialized registration services. If necessary for the purposes of evaluating software applications, third level registrations will be treated separately from this test plan.

3.3 Life Time of the Evaluation Labels

It is planned that the *.test* labels will be kept in the DNS root zone and resolving with *example* positioned at the second level (i.e., translations of *example.test*) until registrations in a corresponding script are available in a production environment. Although it is anticipated that the evaluation facility will be of short-term utility the lifespan of the evaluation may be extended if it is demonstrated that target groups will derive continuing benefit from it.

The sunset procedure for an *example.test* domain will otherwise be initiated as part of the project plan for insertion of production TLDs in the DNS and registration in it is available to the community it serves.

3.4 Label Definitions and Requirements

The labels (*example.test* and the associated scripts) used for the evaluations have been chosen on the basis of the following main requirements:

  (i)   The string is not likely to be desirable in, nor under any circumstances available for, the production environment (further stressing the transient nature of this exercise).

  (ii)   The selected scripts are globally comprehensive to ensure correspondingly broad participation.

A detailed discussion of the requirements for the laboratory string selection was made available online at http://icann.org/announcements/announcement-05dec06.htm/idn-test-labels.pdf . For convenience, selected sections of this paper are provided in the first half of Appendix A. Selection criteria and background information for the present set of evaluation strings based on *example* and *test* is described in the second half of Appendix A. A proposed list of the evaluation strings is provided separately in Appendix B.

3.5 Constraints

Technical constraints apply to the repertoire of available characters so that all of the characters that may be desired by a specific community when making domain name registrations may not be available. A detailed discussion of these constraints is available online at http://icann.org/announcements/announcement-05dec06.htm/idn-test-labels.pdf and partly replicated for convenience in Appendix A.

A revision of the IDNA protocol is currently being considered and is aimed at making as many linguistically justified characters as reasonably possible available for domain name registration (and discusses a process by which additional characters can be included in the permissible IDN repertoire). The protocol is planned to incorporate an inclusion list of available characters. The latest version of the proposed revision is available at http://www.ietf.org/internet-drafts/draft-klensin-idnabis-issues-01.txt.

The present evaluation program will be in compliance with anticipated revisions to the protocol and hence only characters that are certain to be valid in the proposed revised IDNA protocol will be available for registration purposes in the evaluation facilities.

4   Appendices

Appendix A: Evaluation String Selection Criteria

Note: A detailed discussion of the requirements for the string selection was made available online at
http://icann.org/announcements/announcement-05dec06.htm/idn-test-labels.pdf
For convenience, selected sections of this paper are provided immediately below, with some modification to reflect terminology that has been generally adopted in the interim, and to correct minor factual errors.

Relating to the concluded first phase of the evaluation:

The minimum length of a string that is output by the ToASCII component of the Punycode algorithm (previously termed a "Punycode label", now termed an "A-label") is six characters, encoding a single non-ASCII character.(The impending revision to the IDNA protocol is likely to increase the minimum to seven characters.) If current policy constraints on one and two-character labels are taken to apply to their displayed forms (as produced by the ToUnicode component of the Punycode algorithm; previously termed a "Unicode label", now termed a "U-label"), the minimum length of an A-label will be eight characters (with the revision increasing that to nine). It should be noted, however, that these length restrictions were established when there was no difference between the stored and displayed forms of a label, and that some registries regard them as attaching to the stored form. It may be expected, nonetheless, that an IDN-aware revision of general policies can place the restriction on the number of characters in the U-label.

The longest TLD labels currently resolving in the root zone are the six character .museum and .travel. These are both stored and displayed in the same ASCII form but problems have been observed with the response of certain applications to them. However, since similar difficulty has been noted with shorter TLD labels, the problems are more likely to result from failure to recognize the labels as valid TLD designations, than simply from the length of the string. Assuming that requests will be made for new TLDs labeled with dictionary words with display lengths roughly equivalent to .museum and .travel, but written with non-ASCII characters, stored labels of twelve characters and longer will not be uncommon. If, as may also be expected, such things as the names of countries appear in full native representations, the lengths of the stored strings may be significantly greater. This suggests that stored strings of up to the maximum permitted length of 63 characters require evaluation. Even if there is no reason to expect that DNS resolvers will be taxed by the appearance of TLD labels of extreme length, one of the purposes of the technical testing is to identify unanticipated frailty. The response of other widely deployed applications requires testing, in any case.

In many scripts, the way a character is displayed depends both on its position in a string and on the specific characters adjacent to it. If these shaping properties are to be manifested in the test environment as they are likely to appear in actual registered names, a test string cannot

simply be a sequence of randomly selected characters. It can, however, be derived from a word taken from a dictionary of a language written with that script. If an online dictionary is available, its use will ease the determination of the requisite Unicode code points and avoid need for the manual transcription of unfamiliar scripts.

A distinction is made between conditions that pertain to laboratory testing in transient namespaces, and those that attach to tests conducted in the public namespace. The former is served by the generation of A-labels of varying length, with display-side considerations relating solely to script. The latter, however, involves explicit linguistic considerations. Any label that is entered into the root zone of the DNS for the purposes of IDN testing will be categorically barred from subsequent delegation as a production domain. It is therefore also advisable to use a test term that would be unlikely to appear in that context, or is already restricted from such use. To reduce the potential for difficulty to an absolute minimum, a single word is therefore being recommended for all comparable test purposes.

The approach to creating test strings for use in private namespaces will be illustrated by deriving a non lexical sequence from the word "hippopotamus" (a term likely to be found both in bilingual desk dictionaries and in corresponding online resources). To obviate any conceivable residual concern about rendering it inviable for subsequent candidacy for encoding into a production TLD label, a numerical sequence will be embedded in it. This is taken from the abbreviation "i18n" for "internationalization", and is based (but not dependent) on the assumption that no TLD labels will have numerical components.

No harm is likely to be done if a resulting string is unrecognizable in the language from which it was derived. The purpose is to generate typographically plausible sequences of characters in a variety of scripts, with no further semantic value or linguistic correctness being necessary (or even desirable).

Relating to the present plan for the second phase of the evaluation:

A purely lexical alternative is needed for application in the public namespace. A convenient vocabulary is provided by the RFC 2606 list of "Reserved Top Level DNS Names", which explicitly lists four words that are restricted from autonomous delegation because:

> "There is a need for top-level domain (TLD) names that can be used for creating names which, without fear of conflicts with current or future actual TLD names in the global DNS, can be used for private testing of existing DNS related code, examples in documentation, DNS related experimentation, invalid DNS names, or other similar uses."

Of the four names then reserved,

> ".*test* is recommended for use in testing of current or new DNS related code"

and

"*.example* is recommended for use in documentation or as examples."

The name *example* is also reserved on the second level in the .com, .net, and .org TLDs.

It would be counter to the conditions of RFC 2606 for either *.example* or *.test* to resolve in the root, but no restrictions are placed on lexical equivalents to those terms in other languages. One obvious alternative would therefore be to generate test TLD labels from translations of the word *test* into at least one language using each of the scripts that are represented in the public test. The second level label in each such TLD could be similarly generated from the word *example*.

As noted above, any test label that is placed in the root zone will be unavailable for subsequent delegation. However, since the words *example* and *test* are already unavailable, similarly barring the equivalent words in any of the languages figuring in the public test (or perhaps generally in anticipation of future tests) would impose the smallest possible constraint on the production vocabulary. The translated *example* and *test* equivalents can therefore be used freely as IDN test strings in any situation where the intention is for them to be proper dictionary words. All requisite terms will also appear in any bilingual dictionary as discussed above.

Examples of strings determined in this manner are उदाहरण.परीका, 실례.테스트, and. παράδειγμα δοκιμή. If longer TLD labels are needed than those generated by a single instance of a translation of *test*, the word can be repeated as required, for example as, тест-тест-тест.

For differing reasons, all of the elements of a given writing system that might reasonably be requested for inclusion in a localized domain name may not be available. Some restrictions are inevitable consequences of the domain namespace never having been intended to serve as a vehicle for literary expression. Other limitations result from unanticipated problems and may be eliminated through protocol or policy revision.

Converse difficulties have also resulted from excessive latitude in the available repertoire, and some currently viable characters may become unavailable both as IDNA is refined, and as registries adopt more restrictive policies. Work is in underway in several venues that are intended to clarify and rectify these issues. On first consideration, this might appear to be on a level of detail that is irrelevant to technical trials of the type described above. The agencies conducting such activity may, nonetheless, become engaged in dialog with local communities about specific constraints placed on their languages. This will require some familiarity with the kinds of limitations that are still being addressed, and the ability to assess the degree of transiency of specific issues. Key pending details are therefore reviewed below.

Many symbols that are neither alphanumeric nor ideographic components of a written language, such as line drawings and pictographic dingbats, are currently permitted in IDNA but are likely to be blocked in the coming revision. Discussions are still being conducted about the extent to which other nonliteral and nonnumerical characters should be available for inclusion in alphanumeric strings, and about corresponding issues in the ideographic realm. In

one sense, this equates to the consideration of permitting punctuation marks in addition to the hyphen.

Many scripts use other symbols for purposes roughly parallel to the function of the hyphen in English orthography, but do not recognize the hyphen at all. Despite this, it is unlikely that there will be any general rule about one (or some other small number) of symbols being made available for every such script.

There are, however, situations with specific languages where nonliteral adjunct marks can be seen as essential elements of even a skeletal orthography. In addition to the prohibitive criterion of visual confusability with protocol elements, directional properties are also an important factor. IDNA currently requires that a string of characters in a script that is written right-to-left neither begins nor ends with a combining mark. (A string of left-to-right characters may not begin with a combining mark either, but it may end with one.) The clearest example of resulting difficulty that has thus far been noted is with Dhivehi, the official language of Maldives. This is written in the Thaana script (in the Unicode range U+0780...U+07BF), which requires the addition of a combining mark to every base character. A vowel following a consonant is indicated with a combining mark, and special combinations are used to indicate consonants and double vowels in syllable final position. Every Thaana string thus ends with a combining mark and will be rejected by Stringprep (as illustrated with the Dhivehi word for hippopotamus, ﷲﷲﷲﷲ).

There are reasonable IDN labels derived from other languages written with right-to-left scripts that will be similarly rejected because of final combining characters, and there are also cases with left-to-right scripts where the label-final character cannot be correctly represented. One example of this is the lowercase Greek final sigma "ς", which is normalized to the initial and medial form σ and therefore cannot appear at the end of a U-label. This prohibits the correct representation of many names, such as that of the country Cyprus, which can only be incorrectly represented in IDN as κυπροσ. The German Esszet "ß" is similarly irrecoverably normalized to "ss" in the encoding process. (Both the final sigma and Esszet are, however, acceptable input and may appear in the presentation form of a URL, or in offline publication.)

Appendix B: example.test strings

The following list is intended to seed a short-term discussion prior to the finalization of the list of evaluation strings. It includes explicit labels taken from the text in Appendix A (which had been vetted by members of the respective language communities during the preparation of that text), and lists additional languages and scripts that have been put forward in ICANN's IDN workshops. Further consideration has been taken of language demographics in a global perspective. Latitude has deliberately been left for the addition of a limited number of additional languages and/or scripts in response to public commentary. It must be noted, however, that this list is solely intended for purposes of time-limited technical evaluation, and neither does nor can provide an exhaustive representation of all languages and scripts that are expected to appear in the internationalized namespace.

Script	Language	SLD.TLD U-labels	SLD A-label	TLD A-label
Arabic	Arabic	مثال.إختبار	xn--mgbh0fb	xn--kgbechtv
Arabic	Persian	مثال.آزمایشی	xn--mgbh0fb	xn--hgbk6aj7f53bba
Chinese, simplified	Chinese	例子.测试	xn--fsqu00a	xn--0zwm56d
Chinese, traditional	Chinese	例子.測試	xn--fsqu00a	xn--g6w251d
Cyrillic	Russian	пример.испытание	xn--e1afmkfd	xn--80akhbyknj4f
Devanagari	Hindi	उदाहरण.परीक्षा	xn--p1b6ci4b4b3a	xn--11b5bs3a9aj6g
Greek	Greek	παράδειγμα.δοκιμή	xn--hxajbheg2az3al	xn--jxalpdlp
Hangul	Korean	실례.테스트	xn--9n2bp8q	xn--9t4b11yi5a
Hebrew	Yiddish	בײַשפּיל.טעסט	xn--fdbk5d8ap9b8a8d	xn--deba0ad
Kanji Hirigana, and Katakana	Japanese	例え.テスト	xn--r8jz45g	xn--zckzah
Tamil	Tamil	உதாரணம்.பரிட்சை	xn--zkc6cc5bi7f6e	xn--hlcj6aya9esc7a

Note: The names in Arabic and Hebrew script read fully from right to left, and the concatenated U-labels are thus in the apparent order TLD.SLD.

Revision history:

17 July 2007, Persian TLD label changed to clarify use of the Persian *yeh* (U+06CC)

14 August 2007, Appendix B replaced with URL to final version on p.6.

14 August 2007, Old Appendix C, now Appendix B.

14 August 2007 Rendering error corrected in one character in the Hindi TLD U-label.