

Tagged Name Reserved Names Report for RN-WG

Prepared by Chuck Gomes and Patrick Jones, 28 February 2007

1. Background

All existing ICANN registry agreements as of the date of this report contain the requirement for gTLD registries to reserve all labels with hyphens in the third and fourth positions (e.g., “xn--ndk061n”). This requirement comes directly from the approved technical standards for Internationalized Domain Names (IDNs). Note that this reservation requirement does not specify any domain name level, so it is assumed that it applies to all levels of names registered by a given gTLD registry.

Only ASCII characters are permitted in the Domain Name System (DNS) thereby limiting characters to the letters a-z, the numbers 0-9 and the hyphen-dash (-), the latter of which cannot be the first or last character of a domain name. Consequently, to be able to allow representation of domain names in non-ASCII characters, standards were developed in the Internet Engineering Task Force (IETF) that map international scripts to strings of ASCII characters. Those standards require that all ASCII representations of IDNs begin with a 4-character prefix with hyphens in the third and fourth positions.

The current prefix is “xn--”. To avoid confusion of IDNs with ASCII names having the same prefix, it is necessary to reserve the “xn--” prefix. Prior to the finalization of the IDN standards, other prefixes were used, the most recent of which was “bq--”. At that time, speculators started registering ASCII names with the “bq--” prefix. To avoid this possibility with future prefixes, it was decided to reserve all prefixes of this form.

It is also important to note that the current prefix might need to be changed in the future. If that happens, confusion will be avoided by the fact that all labels with hyphens in the third and fourth positions are reserved.

For further information regarding IDNs, please refer to the ICANN Internationalized Domain Names (IDN) information area: <http://www.icann.org/topics/idn/>.

2. Role of tagged name reservation requirement

The role of the tagged name reservation requirement is to be able to provide a way to easily identify an IDN label in the DNS and to avoid confusion of non-IDN ASCII labels. Implicit in this role is the need to reserve tagged names for future use in case the ASCII IDN prefix is changed.

3. WG Recommendations

- a. To avoid user confusion that might result in not being able to tell the difference between a legitimate IDN name and an illegitimate one and to provide maximum

flexibility in the unlikely case that the xn-- prefix should ever need to be changed, we make the recommendations shown in the following table.

Description of Current Reserved Name Requirement All labels with hyphens in the third and fourth character positions (e.g., "bq--1k2n4h4b" or "xn--ndk061n")			
Level	Type	More Work?	Recommendations
Top	ASCII	No	<ol style="list-style-type: none"> 1. Reserve all labels with hyphens in the third and fourth character positions (e.g., "bq--1k2n4h4b" or "xn--ndk061n") <i>without standardization activity and appropriate IANA registration.</i>¹ 2. For each IDN gTLD proposed, applicant must provide both the "ASCII compatible (ACE) form of an IDNA valid string" ("A-label") and in local script form (Unicode) of the top level domain ("U-label").²
Top	IDN	No	N/A
2 nd	ASCII	No	The current <u>reservation</u> requirement be reworded to say, "All labels with hyphens in the third and fourth character positions (e.g., "bq--1k2n4h4b" or "xn--ndk061n") <i>without standardization activity and appropriate IANA registration.</i> " ³ – added words in <i>italics</i> . (Note that names starting with "xn--" may only be used if the current ICANN IDN Guidelines are followed by a gTLD registry.)
2 nd	IDN	No	N/A
3 rd	ASCII	No	Same as for the 2 nd -level for any gTLDs for which registrations occur at the 3 rd -level
3 rd	IDN	No	N/A

Formatted: Indent: Left: 0.05",
Tabs: Not at 0.5"

Formatted: Bullets and Numbering

- b. The Tagged Name Subgroup relied exclusively on Ram Mohan, and Tina Dam as experts and did not believe that additional expert consultation was needed for the topic of tagged name reservations, but did recommend scheduling of a full WG consultation with Ram, Tina and Cary Karp to assist in the finalization of reports for other reserved name categories with regard to IDNs. That WG consultation occurred on 1 March 2007.

4. Consultation with Experts

¹ Considering that the current requirement in all 16 registry agreement reserves "All labels with hyphens in the third and fourth character positions (e.g., "bq--1k2n4h4b" or "xn--ndk061n")", this requirement reserves 1296 names (36x36).

² Internet Draft IDNAbis Issues: <http://www.ietf.org/internet-drafts/draft-klensin-idnabis-issues-01.txt> (J. Klensin), Section 3.1.1.1

³ Considering that the current requirement in all 16 registry agreement reserves "All labels with hyphens in the third and fourth character positions (e.g., "bq--1k2n4h4b" or "xn--ndk061n")", this requirement reserves 1296 names (36x36).

A large number of experts are possible for this topic. The various authors of the RFCs that are summarized in this document are all experts that could be asked to contribute to the RN-WG effort. But, it seems like this category of reserved names is relatively straight forward and has little if any controversy, so it is suggested that only minimal consultation with experts is necessary. The authors of this report consulted with Ram Mohan, Chair of the GNSO IDN Working Group and Tina Dam, ICANN IDN Program Director.

The following questions were asked of Tina Dam and Ram Mohan:

- Would it be possible to only reserve a subset of the tagged names of the form character-character-dash-dash instead of all 1296 variations?
 - If so, how big a subset would be needed?
 - Would we need feedback from the technical community in this regard?
 - If so, who do you think we should contact in that regard?

Here is Ram's response:

"The IETF has defined "xn--" for IDNA, as you know. It is safe to say that questions of defining a subset of the available CCHH range should definitely be run by the IAB, with a note sent to the IAB Chair (Leslie).

"To your question regarding how big a subset would be "needed", the fact is that all CCHH names are restricted so that we don't have charlatans who sell unwitting customers some other CCHH name(s) that will absolutely not work with the existing technical protocols for resolving IDN names worldwide. Therefore, my sense is that it is much safer to restrict all CCHH combinations than to allow just a few, because the end-user is just not going to be able to tell the difference between a legitimate IDN name and an illegitimate one."

Here is Tina's response:

". . I agree with Ram. There is no reason currently to believe that the xn prefix will change but I still think it might be a good pre-caution to keep all labels with "--" in third and fourth place reserved.

One additional comment. The reservation of these kind of labels must include a process for allowing such reserved labels to be registered (at the time where internationalized top level labels are available for registration) and possible some reference to the Unicode version of that label (following the IDNA protocol) is reserved as well. The latter is to make sure that both the stored and displayed name is reserved together. More specific and clear terminology for the stored/displayed label will come for the protocol revision work. As soon as this is available I will send you another note for potential inclusion in the RN-WG work."

A lot of email exchanges occurred involving Tina and Ram to obtain clarity about Tina's suggestion regarding Unicode versions of labels. Rather than confusing the

issue by pasting all of the email messages, we will simply report that the basic suggestion is that for any IDN gTLDs that are proposed, the applicant should be required to provide the "ASCII compatible (ACE) form of an IDNA valid string" representations along with the corresponding Unicode representation to ensure that there is a one-to-one mapping between the "ASCII compatible (ACE) form of an IDNA valid string" and Unicode representations.

Tina also reported that clearer terminology is coming from the protocol revision group and suggests that all IDN related WGs incorporate this terminology. It is highly expected that the protocol revision soon to be released will recommend against the use of the term "punycode string" and instead recommend the use of "ASCII compatible (ACE) form of an IDNA valid string". She went on to clarify that "an IDNA valid string is a string that fulfills the requirements of the IDNA protocol" and noted that "the protocol document goes into further details of what this means". She suggested using the following term: "ASCII compatible (ACE) form of an IDNA protocol valid string". Finally, she stated that under the revised protocol, "Every ACE label will begin with the IDNA ACE prefix, 'xn--'."

5. Summary of Relevant Information Sources

a. ICANN Registry Agreement Requirements

All 16 existing gTLD registry agreements posted on ICANN's website as of 2 February 2007 (.aero, .asia, .biz, .cat, .com, .coop, .info, .jobs, .mobi, .museum, .name, .net, .org, .pro, .tel and .travel) contain the following requirement⁴

Except to the extent that ICANN otherwise expressly authorizes in writing, the Registry Operator shall reserve names formed with the following labels from initial (i.e. other than renewal) registration within the TLD:

C. Tagged Domain Names. All labels with hyphens in the third and fourth character positions (e.g., "bq--1k2n4h4b" or "xn--ndk061n").

ICANN also has ccTLD Sponsorship Agreements and MOUs in place with 12 ccTLD managers.⁵ Each of those agreements contain the following requirement on tagged names:

4. Tagged Domain Names. In addition, domain names in the Delegated ccTLD (excluding subdomain names under domains registered to third parties) having labels with hyphens in the third and fourth character positions (e.g., "rq--

⁴ See "Comparison of gTLD Registry Reserved Names" prepared for the RN-WG and ICANN Registry Agreements located at (<http://www.icann.org/registries/agreements.htm>).

⁵ ICANN ccTLD Agreements located at (<http://www.icann.org/cctlds/agreements.html>).

1k2n4h4b") are reserved from initial (i.e. other than renewal) registration, except as authorized by ICANN policy or by written exception from ICANN.⁶

b. RFC 3490, Internationalizing Domain Names in Applications (IDNA)⁷

The Introduction of RFC 3490 says:

“IDNA works by allowing applications to use certain ASCII name labels (beginning with a special prefix) to represent non-ASCII name labels.

“To allow internationalized labels to be handled by existing applications, IDNA uses an "ACE label" (ACE stands for ASCII Compatible Encoding). An ACE label is an internationalized label that can be rendered in ASCII and is equivalent to an internationalized label that cannot be rendered in ASCII . . . Every ACE label begins with the ACE prefix specified in section 5.”

Section 5 (ACE Prefix) reads:

“The ACE prefix, used in the conversion operations (section 4), is two alphanumeric ASCII characters followed by two hyphen-minuses. It cannot be any of the prefixes already used in earlier documents, which includes the following: "bl--", "bq--", "dq--", "lq--", "mq--", "ra--", "wq--" and "zq--". . . .

“The ACE prefix for IDNA is "xn--" or any capitalization thereof. This means that an ACE label might be "xn--de-jg4avhby1noc0d", where "de-jg4avhby1noc0d" is the part of the ACE label that is generated by the encoding steps in [PUNYCODE].

“While all ACE labels begin with the ACE prefix, not all labels beginning with the ACE prefix are necessarily ACE labels. **Non-ACE labels that begin with the ACE prefix will confuse users and SHOULD NOT be allowed in DNS zones.**” (Bold font added – this is the primary reason for reserving the ACE prefix.)

c. RFC 3492, Punycode: A Bootstring encoding of Unicode for Internationalized Domain Names in Applications (IDNA), March 2003⁸

The Introduction of this RFC says the following:

“[IDNA] describes an architecture for supporting internationalized domain names. Labels containing non-ASCII characters can be represented by ACE labels, which begin with a special ACE prefix and contain only ASCII characters. The

⁶ .AU ccTLD Sponsorship Agreement, Attachment F, <http://www.icann.org/cctlds/au/sponsorship-agmt-attf-25oct01.htm>. The identical provision appears in the other 11 ccTLD agreements.

⁷ <http://www.ietf.org/rfc/rfc3490.txt?number=3490> (P. Faltstrom and P. Hoffman)

⁸ <http://www.ietf.org/rfc/rfc3492.txt?number=3492> (A. Costello)

remainder of the label after the prefix is a Punycode encoding of a Unicode string satisfying certain constraints. For the details of the prefix and constraints, see [IDNA] and [NAMEPREP].”

- d. GNSO Preliminary Issues Report Policy Issues relating to IDN at the top-level, 28 May 2006⁹

An introduction of PUNYCODE is provided in this document:

“Punycode is a bootstring encoding that will convert the local characters in a domain name into the limited character set that is supported by the DNS. The encoding is applied to each component of a domain name and a prefix 'xn--' is added to the translated Punycode string. For example, the first component of the domain name rødgrødmedfløde.dk becomes 'xn--rdgrdmedflde-vjbdg', and the domain will be represented as xn--rdgrdmedflde-vjbdg.dk. This kind of encoding would apply for top-level labels with characters from non-Latin scripts.”

- e. Informational RFC 4690, Review and Recommendations for Internationalized Domain Names (IDNs), September 2006¹⁰

The following excerpt relates to the possibility of the need to change the Punycode prefix:

“It is worth noting that sufficiently extreme changes to IDNA would require a new Punycode prefix, probably with long-term support for both the old prefix and the new one in both registration arrangements and applications. An alternative, which is almost certainly impractical, would be some sort of "flag day", i.e., a date on which the old rules are simultaneously abandoned by everyone and the new ones adopted. However, preliminary analysis indicates that few, if any, of the changes recommended for consideration elsewhere in this document would require this type of version change. For example, suppose additional restrictions, such as those implied above, are imposed on what can be registered. Those restrictions might require policy decisions about how labels are to be disposed of if they conformed to the earlier rules but not to the new ones. But they would not inherently require changes in the protocol or prefix.”

- f. Internet Draft, Proposed Issues and Changes for IDNA - An Overview, October 16, 2006¹¹

Section 5, The Question of Prefix Changes, says the following:

⁹ <http://gns0.icann.org/issues/idn-tlds/issues-report-28may06.htm>

¹⁰ <http://www.ietf.org/rfc/rfc4690.txt?number=4690> (J. Klensin, P. Faltstrom, C. Karp)

¹¹ <http://www.ietf.org/internet-drafts/draft-klensin-idnabis-issues-00.txt> (J. Klensin)

“The conditions that would require a change in the IDNA "prefix" ("xn--" for the version of IDNA specified in [RFC3490]) have been a great concern to the community. A prefix change would clearly be necessary if the algorithms were modified in a manner that would create serious ambiguities during subsequent transition in registrations. This section summarizes our conclusions about the conditions under which changes in prefix would be necessary.

“5.1. Conditions requiring a prefix change

“An IDN prefix change is needed if a given string would resolve or otherwise be interpreted differently depending on the version of the protocol or tables being used. Consequently, work to update IDNs would require a prefix change if, and only if, one of the following four conditions were met:

1. The conversion of a Punycode string to Unicode yields one string under IDNA2003 (RFC3490) and a different string under IDNA200x.
2. An input string that is valid under IDNA2003 and also valid under IDNA200x yields two different Punycode strings with the different versions. This condition is believed to be essentially equivalent to the one above.

Note, however, that if the input string is valid under one version and not valid under the other, this condition does not apply. See the first item in Section 5.2, below.

3. A fundamental change is made to the semantics of the string that is inserted in the DNS, e.g., if a decision were made to try to include language or specific script information in that string, rather than having it be just a string of characters.
5. Sufficient characters are added to Unicode that the Punycode mechanism for offsets to blocks does not have enough capacity to reference the higher-numbered planes and blocks. This condition is unlikely even in the long term and certain to not arise in the next few years.”

- g. Internet Draft, Proposed Issues and Changes for IDNA - An Overview (IDNAbis Issues), February 23, 2007¹²

(Note: This is version 01, an update to the previously listed Internet Draft of the same name, version 00.)

¹² IDNAbis Issues: <http://www.ietf.org/internet-drafts/draft-klensin-idnabis-issues-01.txt> (J. Klensin)

Section 8.1, Design Criteria, says the following regarding tagged names:

“3. Anyone entering a label into a DNS zone must properly validate that label -- i.e., be sure that the criteria for an A-label are met -- in order for Unicode version-independence to be possible. In particular:

- Any label that contains hyphens as its third and fourth characters MUST be IDNA-valid. This implies in particular that, (i) if the third and fourth characters are hyphens, the first and second ones MUST be "xn" until and unless this specification is updated to permit other prefixes and (ii) labels starting in "xn--" MUST be valid A-labels, as discussed in Section 3 above.”

Section 8.3, The Question of Prefix Changes, says:

“The conditions that would require a change in the IDNA "prefix" ("xn--" for the version of IDNA specified in [RFC3490]) have been a great concern to the community. A prefix change would clearly be necessary if the algorithms were modified in a manner that would create serious ambiguities during subsequent transition in registrations. This section summarizes our conclusions about the conditions under which changes in prefix would be necessary.

“8.3.1. Conditions requiring a prefix change

“An IDN prefix change is needed if a given string would resolve or otherwise be interpreted differently depending on the version of the protocol or tables being used. Consequently, work to update IDNs would require a prefix change if, and only if, one of the following four conditions were met:

1. The conversion of a Punycode string to Unicode yields one string under IDNA2003 (RFC3490) and a different string under IDNA200x.
2. An input string that is valid under IDNA2003 and also valid under IDNA200x yields two different Punycode strings with the different versions of IDNA. This condition is believed to be essentially equivalent to the one above.

Note, however, that if the input string is valid under one version and not valid under the other, this condition does not apply. See the first item in Section 8.3.2, below.

3. A fundamental change is made to the semantics of the string that is inserted in the DNS, e.g., if a decision were made to try to include

language or specific script information in that string, rather than having it be just a string of characters.

4. A sufficiently large number of characters is added to Unicode so that the Punycode mechanism for block offsets no longer has enough capacity to reference the higher-numbered planes and blocks. This condition is unlikely even in the long term and certain not to arise in the next few years.”

“Section 8.3.2, Conditions not requiring a prefix change, says:

“In particular, as a result of the principles described above, none of the following changes require a new prefix:

1. Prohibition of some characters as input to IDNA. This may make names that are now registered inaccessible, but does not require a prefix change.
2. Adjustments in Stringprep tables or IDNA actions, including normalization definitions, that do not affect characters that have already been invalid under IDNA2003.
3. Changes in the style of definitions of Stringprep or Nameprep that do not alter the actions performed by them.”