

GUIDELINES FOR THE IMPLEMENTATION OF INTERNATIONALIZED DOMAIN NAMES

Discussion draft

15 November 2010

Introduction

This discussion draft version of the IDN Guidelines is released to facilitate an interactive IDN session on the subject of IDNA Protocol implementation and transition during the ICANN meeting in Cartagena, Colombia, December 2010.

This discussion draft is based on the [latest version of the IDN Guidelines, version 2.2.](#)

This discussion draft is established on the [revised version of the IDNA protocol](#) and was prepared by members of the IDN Guidelines Revision WG (a working group of gTLD and ccTLD registries with IDN experience):

- gTLD Registry Constituency Representatives:
 - Cary Karp, MuseDoma
 - Pat Kane, VeriSign
 - Ram Mohan, Afiliis
- ccNSO Representatives:
 - Hiro Hotta, JPRS
 - Mohammed EL Bashir, .sd Registry
- ICANN Staff
 - Tina Dam

Discussion Draft IDN Guidelines

1. Top-level domain ("TLD") registries supporting Internationalized Domain Names ("IDNs") will do so in strict compliance with the requirements of the IETF protocol for Internationalized Domain Names in Applications. The initial version of this protocol ("IDNA" aka IDNA2003) is defined in RFCs 3454, 3490, 3491, and 3492. A revised version ("IDNABIS" aka IDNA2008) is defined in RFCs 5890, 5891, 5892, 5893, and 5894. Differences between them are discussed separately in Appendix A.

This is extended in Appendix B with suggested practices for accommodating transitional issues.

2. No code point permitted in IDNA but disallowed in IDNABIS will be accepted for registration regardless of the extent to which such code points appear in names registered prior to the protocol revision. The registrant of a domain that is no longer supported by IDNABIS should be notified that there may be unanticipated consequences for a user attempting to reach it, and such names should be replaced or deleted at registry initiative.

3. A registry will publish one or several lists of Unicode code points that are permitted for registration and will not accept the registration of any name containing an unlisted code point. Each such list will indicate the

script or language(s) it is intended to support. If registry policy treats any code point in a list as a variant of any other code point, the nature of that variance and the policies attached to it will be clearly articulated.

4. All code point listings will be placed in the IANA Repository for IDN TLD Practices in tabular format together with any rules applied to the registration of names containing those code points, before any such registration may be accepted.

5. Any material fundamental to the understanding of a registry's IDN policies that is not published by the IANA will be made readily available online directly by the registry, which should also ensure that its registrars call the attention of prospective registrants of IDN names to it. This documentation will include references to the linguistic and orthographic sources used in establishing policies and code point repertoires. If material is provided both via the IANA and other channels the registry must ensure that its substance is concordant across all platforms.

6. When a pre-existing name requires a registry to make transitional exception to any of these Guidelines, the terms of that action will also be made readily available online, including the timeline for the resolution of such transitional matters. At the end of this period, code points that are prohibited by the IDNABIS standard will not be permitted even by exception.

7. No label containing hyphens in the third and fourth positions will be registered unless it is a valid A-label, with reservation for transitional action in accordance with the preceding Guideline. Hyphens in these positions are explicitly reserved to indicate encoding schemes, of which IDNA/IDNABIS is only one instantiation.

8. TLD registries will collaborate on issues of shared interest, for example, by forming a consortium to coordinate contact with external communities, elicit the assistance of support groups, and establish global fora.

Appendix A

A1. IDNABIS makes several changes to the initial IDNA specification that are of material consequence for TLD registries supporting IDN. The operator of any such registry should therefore be aware of key aspects of the protocol revision and make special provision for the registration of names that are valid under IDNA but are treated differently under IDNABIS. The most directly relevant protocol details are described in separately numbered sections below.

A2. IDNA is locked to Unicode version 3.2. There have, however, been several subsequent additions to the Unicode repertoire (now at version 6.0) that would immediately extend the benefit of IDNs if they were permitted by the protocol. IDNABIS supports code points that appear in new versions of Unicode without need for separate adjustment to the protocol.

A3. IDNA places greater restrictions on the use of scripts written from right to left than it does on scripts written from left to right. IDNABIS reduces that imbalance and clarifies rules about the commingled use of characters with both directional properties in a single label.

A4. IDNABIS prohibits graphic symbols and similar devices that have code points but are not used as basic elements of any writing system. Previous Guidelines explicitly prohibiting these symbols are now redundant and have been removed.

A5. IDNA remaps a number of code points to other code points while preparing the ASCII-encoded sequence that is actually entered into the DNS. It is therefore possible for a single A-label to be generated from a number of different U-labels. The A-label will, however, only decode to one of those U-labels. IDNABIS removes all such remapping from the protocol, ensures a unique equivalence between any A-label and a corresponding U-label, and eliminates any confusion about the label that has actually been registered.

Appendix B

[This is a seed list of recommended practices and will be modified and extended in the ongoing revision of these Guidelines.]

B1. Whenever an IDN registry adds support for a new code point there is need for dealing with the registrants of names that would likely have included that code point if it had been possible at the time of initial registration. These registrants need special accommodation before the modified form is made available for registration by anyone else and it is assumed that the registry either has preexisting policies for dealing with such situations or recognizes situations where they are needed. The concepts normally applied to such policies include sunrise, bundling, and blocking, but no general recommendations are currently being put forth in these Guidelines. The following two points do, however, describe situations that lack counterpart in previous practice and therefore require special consideration.

B2. Two specific consequences of the elimination of remapping require particular attention. The U+03C2 GREEK SMALL LETTER FINAL SIGMA (ς), and the U+00DF LATIN SMALL LETTER SHARP S (ß) are accepted elements of Greek and German orthographies, respectively. The IDNA remapping bars their inclusion in registered names but does allow them to appear in queries directed to the DNS. IDNABIS makes them available for actual registration and this change may initially result in unexpected behavior on the query side. As discussed in the preceding point, a registry supporting the two new characters will need to deal with preexisting names that registrants may wish to modify or complement, prior to making the newly introduced form available for autonomous registration.

B3. IDNABIS makes certain code points available under the explicit condition that a registry supporting them imposes clearly-stated contextual rules on their use. This is of particular importance to the use of non-spacing Unicode control characters ("join controls"), which IDNABIS permits to extend support for the correct display of characters in complex scripts that take various forms depending on their position in a label, and on the characters to which they are adjacent.