

GUIDELINES FOR THE IMPLEMENTATION OF INTERNATIONALIZED DOMAIN NAMES

Version 3.0

2 September 2011

Introduction

This document supersedes version 2.2 of these Guidelines to reflect the IDNABIS revision (“IDNA2008”) of the initial IDNA protocol (“IDNA2003”). It was prepared by members of the IDN Guidelines Revision Working Group (of gTLD and ccTLD registries with IDN experience):

gTLD Registry Constituency Representatives:

Cary Karp, MuseDoma
Jimmy Lam, Afilias
Will Shorter, VeriSign

ccNSO Representatives:

Mohammed EL Bashir, Qatar Domains Registry (ictQATAR)
Hiro Hotta, JPRS

ICANN Support Staff

Naela Sarras
Francisco Arias
Patrick Jones

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 was defined in RFCs 3454, 3490, 3491, and 3492. A revised version is defined in RFCs 5890, 5891, 5892, 5893, and 5894. Both will be in parallel use in applications for an indeterminate transitional period but registries will conform fully with IDNA2008 in the shortest practicable order.
2. No code point permitted in IDNA2003 but disallowed in IDNA2008 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 IDNA2008 should be notified that there may be unanticipated consequences for a user attempting to reach it, and such names should be replaced, held, 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 such 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. All code points in a single label will be taken from the same script as determined by the Unicode Standard Annex #24: Script Names <<http://www.unicode.org/reports/tr24>>. Exceptions to this guideline are permissible for languages with established orthographies and conventions that require the commingled use of multiple scripts. Even in the case of 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 is clearly defined.

6. Any information fundamental to the understanding of a registry's IDN policies that is not published by the IANA will be made directly available online by the registry. The registry should also encourage its registrars to call attention to these policies for all prospective IDN registrants. 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.

7. When a preexisting 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. The excepted registrations themselves are, however, not part of this documentation. At the end of the transitional period, code points that are prohibited by IDNA2008 will not be permitted even by exception.

8. 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 is only one instantiation. These guidelines are not intended to assist with any other instantiations.

9. TLD registries should 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: Comparison of IDNA2003 with IDNA2008

A1. IDNA2008 makes several changes to the initial IDNA2003 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 IDNA2003 but are treated differently under IDNA2008. The most directly relevant protocol details are described in separately numbered sections below.

A2. IDNA2003 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. IDNA2008 supports code points that appear in new versions of Unicode without need for fundamental adjustment to the protocol. If, however, a new Unicode version changes the properties of preexisting code points, the validity of those code points may also change. (This is discussed further in Appendix B4.)

A3. IDNA2003 places greater restrictions on the use of scripts written from right to left than it does on scripts written from left to right. IDNA2008 reduces that imbalance and clarifies rules about the commingled use of characters with both directional properties in a single label.

A4. IDNA2008 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. IDNA2003 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. IDNA2008 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: Additional transitional issues

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 IDNA2003 remapping bars their inclusion in registered names but does allow them to appear in queries directed to the DNS. IDNA2008 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 may need to deal with preexisting names that registrants wish to modify or complement, prior to making the newly introduced form available for autonomous registration.

B3. IDNA2008 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 IDNA2008 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.

B4. IDNA2008 was finalized when Unicode version 5.2 was in effect. The subsequently released version 6.0 changed the properties of three code points with the effect that two which were previously disallowed in IDNA2008 became valid, and one that was valid became disallowed (U+19DA NEW TAI LUE THAM DIGIT ONE). The IETF did not feel that this required changes to the underlying component of IDNA2008 (RFC 5892) and will reexamine need for such action with each successive release of Unicode. Registries should be aware of this but can expect it not to have disruptive

consequence. If the status of a code point that is deemed likely to appear in registered IDNs should reverse due to a change to its Unicode properties, IDNA2008 includes an exception mechanism that can override those changes and maintain the validity of the code point.