Guidelines for the Implementation of Internationalized Domain Names

Final Draft for the Public Comment – 3 March 2017

1 Introduction

These Guidelines are about the implementation of Internationalized Domains Names (IDN) under Internet Domains. IDN is standardized by IETF in IDNA 2008.

The main audience of this document is Top-Level Domain (TLD) registries that offer or plan to offer registrations of IDNs under their Registry Agreements. For other registries (e.g. Country Code Top Level Domain Name registries) this document is intended as the best current practice. These Guidelines are also intended for registrars offering registration of IDNs.

The document has been prepared by members of the IDN Guidelines Working Group (IDNGWG), listed in Appendix A, constituted following the Call for Community Experts.

1.1 Normative Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119.

1.2 Document Version

This document supersedes version 3.0 of the Guidelines, following the expansion of the DNS under the 2012 New gTLD Program.

2 IDN Guidelines

2.1 Transition

1. TLD registries supporting Internationalized Domain Names (IDNs) must do so in strict compliance with the requirements of the IETF protocol for Internationalized Domain Names in Applications, as defined in the standards track RFCs 5890, 5891, 5892 and 5893.

2. Code points permitted in IDNA 2003 but disallowed in IDNA 2008 must not be accepted for registration regardless of the extent to which such code points appear in domain names registered prior to the protocol revision. The registrant of a domain name that is no longer supported by IDNA 2008 should be notified that there may be unanticipated consequences for a user attempting to reach it, and such domain names should be replaced, held, or deleted at registry initiative.

3. When a pre-existing domain name requires a registry to make transitional exception to any of these Guidelines, the terms of that action must 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 IDNA 2008 must not be permitted even by exception.

4. No label containing hyphens in the third and fourth positions must be registered unless it is a valid A-label, with reservation for transitional action. 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.

5. TLD registries with pre-existing domain names that do not conform to these guidelines should take the following actions to reduce disruption to registrants and Internet consumers:
   a. Make clear in their registration policy whether registered domain names or currently activated labels, which do not conform to the guidelines, will continue to be published in the TLD zone file.
   b. In cases where non-conforming registered domain names will continue to be published in the zone file, make clear any additional restrictions placed on usage.
      i. Include restrictions that may influence the lifecycle of the domain name, such as restrictions on renewals, transfers and change of registrant
      ii. Include restrictions on the activation or usage of variants.
      iii. Clearly state whether the continuing publication in the zone file of non-conforming labels will cease after a period of time.
         1. If publication of non-conforming labels into the zone file will cease, then clearly state the date at which the labels will be removed from the zone file.
   c. Publish relevant changes to the TLD's registration policy at a publicly accessible location on the TLD registry's website.
   d. Encourage registrars to notify registrants of non-conforming registered domain names of the change of policy and of all relevant dates and conditions which may apply to such domain names.

2.2 Format of IDN Tables

6. A registry must publish one or several repertoires of Unicode code points[1] that are permitted for registration and must not accept the registration of any domain name containing an unlisted code point. Each such list must 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 variant rules and the policies attached to it must be clearly articulated.

7. IDN tables must be placed in the IANA Repository for IDN Practices. Further, (a) Except as applicable in 7(b) below, registries must use Label Generation Ruleset (RFC 7940) format to represent an IDN table; (b) Registries with existing legacy IDN tables already

[1] Code points can be individual or could also include code point sequences, as suggested in RFC 7940.

present within the IANA Repository for IDN Practices at the time these guidelines are published are encouraged to transition to the LGR format; (c) The IDN table must include the complete repertoire of code points, any variant code points and any applicable whole-label evaluation rules which the registry uses to determine if a label is acceptable for registration.

2.3 Consistency of IDN Tables and Practices

8. TLD registries are encouraged to 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 to address common current and emerging challenges in the development and use of IDNs.

9. TLD registries seeking to implement new IDN tables or to modify existing ones may use available Reference Second Level LGRs as is or as a reference. IDN tables may deviate from Reference Second Level LGRs. Notwithstanding the foregoing, Registry Operators seeking to implement IDN tables (i.e. new or modifications of existing ones) that pose any security[2] and/or stability[3] issues must not be authorized to implement such LGRs.

10. TLD registries offering registration of IDN labels with the same language or script tag (RFC 5646) are encouraged to cooperate and contribute toward the development and update of the Reference Second Level LGRs with the goal of minimizing the difference between the reference LGRs of that language or script and the implemented IDN tables for the same language or script.

11. Any information fundamental to the understanding of a registry's IDN policies that is not published by IANA must be made directly available online by the registry. This documentation must include references to the linguistic and orthographic sources used in establishing policies and IDN tables. The registry should also encourage its registrars to call attention to these policies for all IDN registrants. If material is provided both via the IANA Repository of IDN Practices and other channels, the registry must ensure that its substance is concordant across all platforms.

---

[2] **Security** - An effect on security by the proposed Registry Service shall mean (A) the unauthorized disclosure, alteration, insertion or destruction of Registry Data, or (B) the unauthorized access to or disclosure of information or resources on the Internet by systems operating in accordance with all applicable standards.

[3] **Stability** - An effect on stability shall mean that the proposed Registry Service (A) is not compliant with applicable relevant standards that are authoritative and published by a well-established, recognized and authoritative standards body, such as relevant Standards-Track or Best Current Practice RFCs sponsored by the IETF or (B) creates a condition that adversely affects the throughput, response time, consistency or coherence of responses to Internet servers or end systems, operating in accordance with applicable relevant standards that are authoritative and published by a well-established, recognized and authoritative standards body, such as relevant Standards-Track or Best Current Practice RFCs and relying on Registry Operator's delegation information or provisioning services.

2.4 IDN Variant Labels

12. IDN Variant Labels generated by an IDN table must be a) allocated to the same registrant, or b) blocked from registration.

13. Only IDN Variant Labels with a disposition of "allocatable" may be included in the DNS. IDN Variant Labels must only be delegated into the DNS ("activated") as requested by the registrant (or corresponding registrar), except in cases where a registry-side approach is explicitly expressed in the IDN policies for a particular language/script.

    In cases of registry-side approach, the registry must carefully take into consideration the security and stability impacts: (i) as advised in the relevant documents from SSAC; (ii) different user experience perspectives as explained in the document Examining the User Experience Implications of Active Variant TLDs; (iii) the IDN Variant Issues Project: Final Integrated Issues Report; (iv) the IDN policies and LGRs adopted by the relevant respective language communities; as well as (v) the evidenced operational experiences from such communities, before implementing any IDN policy that includes registry-side activation of IDN Variant Labels.

    For example, the Chinese Domain Name Consortium, the related informational RFC on preferred variants relevant to the Han script (RFC3743) and the Report on Chinese Variants in Internationalized Top-Level Domains.

2.5 Similarity and Confusability of Labels

Commingling of cross-script code points in a single label

14. All code points in a single label must 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.

Harmonization of variant rules across same-script IDN tables

15. TLD registries must ensure that all applicable same-script IDN tables with a variant policy have uniform variant rules that properly account for symmetry and transitivity properties of all variant sets. Exceptions to this guideline vis-à-vis symmetry and transitivity properties should be clearly documented in registries' public policy. At the same time, TLD registries shall re-evaluate potential variant relationships that may require to create new variant sets due to the introduction of additional IDN tables by the registry. Registries may use relevant work for the Root Zone LGR and other sources to determine the variant sets.

Cross-script homoglyph labels

16. TLD registries may apply whole-label evaluation rules to new registrations that minimize whole-script confusables as determined by Unicode Technical Standard #39: Unicode Security Mechanisms http://unicode.org/reports/tr39/tr39-1.html#Whole_Script_Confusables. Registries may use data references such as Unicode's

intentional.txt, the cross-script variants in the Root Zone LGR or other authoritative sources. Any policy and its sources must be clearly documented in the registry's public website.

Limitations of IDN tables and policies

17. In the case of any exceptions made allowing mixing of scripts, visually confusable characters from different scripts must not be allowed to co-exist in a single set of permissible code points unless a corresponding policy and IDN table is clearly defined to mimimize confusion between domain names. TLD registries should also consider policies to minimize confusion between domain names arising from visually confusable characters within a same script.

    It is important to understand that not all visual confusing similarity issues can be addressed by IDN tables and IDN policies. Other policies such as dispute resolution policies may be necessary to mitigate against abusive registrations exploiting visually similar characters. For example, even for ASCII letters digits and hyphen (LDH) repertoire, whereas the digit "0" and letter "O", or the capital letter "I", small letter "l" and digit "1", may be considered visually confusable characters the mitigation policy for abuse is often addressed by dispute resolution policies, leveraging other bodies of knowledge (e.g. Trademark Law) to evaluate whether similarities between domain names causes confusion and abuse.

2.6 Terminology

18. The community is encouraged to adopt the relevant terminology used in these Guidelines as defined in Appendix B

2.7 Registration Data

This topic was considered by the IDN Guidelines WG. The WG does not have any recommendations on this topic. In case the community has any suggestions they should provide their feedback.

2.8 EPP

This topic was considered by the IDN Guidelines WG. The WG does not have any recommendations on this topic. In case the community has any suggestions they should provide their feedback.

Appendix A: Members of IDN Guidelines WG

	Name	Supporting Organization/ Advisory Committee
1	Satish Babu	**ALAC**
2	Wael Nasr	**ALAC**
3	Mats Dufberg	**ccNSO**
4	Pablo Rodríguez	**ccNSO**
5	Edmon Chung	**GNSO**
6	Christian Dawson	**GNSO**
7	Chris Dillon	**GNSO**
8	Kal Feher	**GNSO**
9	Dennis Tan	**GNSO**
10	Jian Zhang	**GNSO**
11	Ram Mohan	**SSAC**
12	Patrik Fältström (will only review work)	**SSAC**

Appendix B: Glossary of Relevant Terms

Term	Acronym	Definition	Additional Notes	Other related Terms
Internationalized Domain Names	IDNs	Domain names containing characters not included in the traditional DNS preferred form ("LDH"). IDNs under discussion are implemented using IDNA		
Internationalized Domain Names in Applications 2003	IDNA 2003		Defined by standard track RFCs 3454, 3490, 3491, 3492	IDNA 2008
Internationalized Domain Names in Applications 2008	IDNA 2008		Defined by standard track RFCs 5890, 5891, 5892 and 5893	IDNA 2003
Code Point		A value, or position, for a character, in any coded character set	As defined by Unicode at http://unicode.org/glossary/#code_point	Code Point Sequence
Code Point Sequence		A sequence of two or more Code Points (e.g. as specified in an LGR)	As explained in RFC 7940, Section 5.1	Code Point
Blocking of a label		An action taken on a given label with respect to a zone, according to which the label is	As defined in Integrated Issues	Blocked

Term	Acronym	Definition	Additional Notes	Other related Terms
		unavailable for allocation to anyone	Report of Variant Issues Project	
Allocation of a label		A label with respect to a zone, whereby the label is associated administratively to some entity that has requested the label	As defined in Integrated Issues Report of Variant Issues Project	Allocatable, Allocated
Delegation of a label		A label with respect to a zone, indicating that in that zone there are NS resource records at the label and that there is no SOA resource record at the label (i.e., that this is the parent zone: there are also NS records with the same owner name in the child zone, but in that child zone there must be an SOA record as well).	As defined in Integrated Issues Report of Variant Issues Project	Delegated
Variant		The term "variant" is used generally to identify different types of linguistic situations where different words are considered to be the same (i.e. a variant) of another word. Because of the wide-ranging understanding of the term, to avoid confusion more specific terms such as "Variant Code Point"		IDN Variant Code Point, IDN Variant Label

Term	Acronym	Definition	Additional Notes	Other related Terms
		or "IDN Variant Label" should be used.		
Label Generation Ruleset, or Label Generation Rules	LGR	LGRs are algorithms used to determine whether, and under what conditions, a given identifier label is permitted, based on the code points it contains and their context. These algorithms comprise a list of permissible code points, variant code point mappings, and a set of rules that act on the code points and mappings. LGRs form part of an administrator's policies. In deploying Internationalized Domain Names (IDNs), they have also been known as IDN tables	As introduced in RFC 7940. Format specified in RFC 7940. Additonal formats include those specified in RFC 4290 and RFC 3743	IDN Table
Code Point Repertoire for the Zone		Also known informally as a zone repertoire. A set of code points permitted in U-labels in a zone	As defined in Integrated Issues Report of Variant Issues Project. Used synonymously for Code Point Repertoire or just Repertoire	Repertoire, Code Point Repertoire

Term	Acronym	Definition	Additional Notes	Other related Terms
Homoglyph		An abstract character or a conceptual character that is represented with the same glyph as another abstract character or conceptual character.	As defined in Integrated Issues Report of Variant Issues Project	
Glyph		A synonym for *glyph image*. In displaying Unicode character data, one or more glyphs may be selected to depict a particular character. These glyphs are selected by a rendering engine during composition and layout processing	As defined by Unicode at http://unicode.org/glossary/#glyph	
Whole Label Evaluation Rules	WLE Rules	Context-based and whole label rules. The also contain the character classes that they depend on, and any actions that assign dispositions to labels based on rules or variant mappings	As explained in RFC 7940, Seciton 6	
Internationalized Domain Name Table	IDN Table	Specification of permitted code points and combition of those in domains name labels . Also see LGR.	Formats specified in RFC 7940, RFC 4290 and RFC 3743	LGR
Allocatable		An IDN label which can be Allocated		Allocated, Allocation of a Label

Term	Acronym	Definition	Additional Notes	Other related Terms
Allocated		State of an IDN label after Allocation The resulting string should be reserved for use by the same operator of the origin string but not automatically allocated for use.	As defined in RFC 7940, Section 7.3	Allocatable, Allocation of a Label
Activated		State of an IDN label after Activation; The resulting string should be activated for use. (This is the same as a Preferred Variant [RFC3743].)	As defined in RFC 7940, Section 7.3	
Withheld				
Blocked		State of an IDN label after blocking. The resulting string is a valid label but should be blocked from registration. This would typically apply for a derived variant that is undesirable due to having no practical use or being confusingly similar to some other label	As defined in RFC 7940, Section 7.3	Blocking of a Label

Term	Acronym	Definition	Additional Notes	Other related Terms
Variant Code Point(s)		Code point(s) that may be used as alternative for code point(s) in the zone repertoire based on a given IDN Table		
IDN Variant Label		A label generated as a variant of a Primary IDN Label based on a given LGR (or IDN Table and IDN registration rules)		Label, IDN Label, Primary IDN Label
Primary IDN Label		An IDN Label applied-for or submitted by a registrant		Label, IDN Label, IDN Variant Label
Label		Part of a domain name separated by dots		
Internationalized Domain Name Label	IDN label	A label valid as per IDNA 2008		Label