

Guidelines for the Implementation of Internationalized Domain Names 4.0

27 April 2018 – IDNGWG FINAL CALL - DRAFT VERSION

1 Introduction

These Guidelines are about the implementation of Internationalized Domain 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 registration of IDNs under their Registry Agreements. For other registries (e.g. Country Code TLD registries) this document is intended as the best current practice. These Guidelines are also intended for registrars offering registration of IDNs.

The sections on Additional Notes and Glossary of Relevant Terms are considered an integral part of these guidelines.

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 Document Version

This document supersedes [version 3.0](#) of the Guidelines, following the expansion of the DNS under the 2012 New gTLD Program.

1.2 Scope

With regards to the contents of the TLD zone file, the scope of this document is limited to only the owner-name of the DNS records which are added to the zone file by the registration system. Excluded from scope are any glue records and right-hand or target names.

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 or any RFC that replaces or updates the listed RFCs.
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.

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. Also see 18(a).
4. No label containing hyphens in both the third and the fourth positions may be registered unless it is a valid A-label, with reservation for transitional action. Labels with hyphens in both the third and the fourth 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.

2.2 Format of IDN Tables

5. A registry must publish one or several repertoires of Unicode code points¹ 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.
6. IDN Tables must be placed in the IANA Repository for IDN Practices. Further:
 - (a) Except as applicable in 6(b) below, registries must use RFC 7940: Label Generation Ruleset (LGR) Using XML format to represent an IDN Table;
 - (b) Registries with existing IDN Tables already 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 rules and any applicable contextual rules which the registry uses to determine if a label is acceptable for registration.

2.3 Consistency of IDN Tables and Practices

7. TLD registries are encouraged to collaborate on issues of shared interest. Registries may form or join an existing 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. The maturity and needs of particular IDN communities will vary greatly. Therefore, while collaboration is considered good practice, the assessment of the importance and utility of such consortia is left to the Registry Operator.
8. TLD registries seeking to implement new IDN Tables or to modify existing ones may use available Reference Second Level LGRs (<https://www.icann.org/resources/pages/second-level-lgr-2015-06-21-en>) as is or as a reference. IDN Tables may deviate from Reference Second Level LGRs. Notwithstanding the foregoing, registries seeking to implement IDN Tables (i.e. new or modifications of existing ones) that pose any security and/or stability issues must not be implemented.

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

9. 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.
10. 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. Including references to the linguistic and orthographic sources used in establishing IDN policies and tables is useful for implementers to understand the context of such policies. 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.

2.4 IDN Variant Labels

11. IDN Variant Labels generated by an IDN Table must be either (a) allocatable only to the same registrant as the primary IDN label, or (b) blocked from registration. Also see 18(b).
12. TLD Registries may activate an IDN Variant Label, provided that i) such IDN Variant Label is requested by the same registrant or corresponding registrar as the Primary IDN Label, ii) such IDN Variant Label is registered to the registrant of the Primary IDN Label, and iii) such IDN Variant Label conforms with the registry policy and IDN Tables.

In exceptional cases, i) to support a widely acceptable practice within Internet users of a language or script community, or ii) to abide by language or script established conventions, a TLD Registry may opt to activate a limited number of IDN Variant Labels at its discretion, according to its policies. In such cases, the TLD Registry must have mechanism to limit automatic activation of IDN Variant Labels to a minimum. Also see 18(c) and Additional Note V.

2.4.1 Harmonization of variant rules across same-script IDN Tables

13. TLD registries must ensure that all applicable IDN Tables with a variant policy for a particular TLD have uniform variant rules that properly account for symmetry and transitivity properties of all variant sets across these IDN Tables. 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. Also see Additional Notes I and II.

2.5 Similarity and Confusability of Labels

2.5.1 Within-script homographs

14. TLD registries are encouraged to consider policies to minimize confusion of IDN labels with other labels within the same script, specifically arising due to homoglyphic characters. Also see Additional Note III.

2.5.2 Commingling of cross-script code points in a single label

15. All code points in a single label must be taken from the same Unicode 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 Unicode scripts. Also see Additional Notes VI and VIII.
16. In the case of any exceptions made allowing mixing of Unicode 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 minimize confusion between domain names. Also see Additional Note III.

2.5.3 Whole-script confusables

17. TLD registries are encouraged to apply additional constraints on registrations that minimize Whole-Script Confusables as determined by Unicode Technical Standard #39: Unicode Security Mechanisms (<http://unicode.org/reports/tr39>), and Unicode Technical Report #36: Unicode Security Considerations (<http://unicode.org/reports/tr36>). Also see 18 (d) and Additional Note IV.

2.6 Publishing IDN Registration Policy and Rules

18. TLD Registries should publish policies or guidance related to registration of IDN labels at publicly accessible location on the TLD Registry's website. In addition to general policies or guidance on IDN registrations, these should include the following:
 - (a) A timeline related to resolution of transitional matters, if applicable
 - (b) IDN Variant Label allocation policy, if applicable
 - (c) IDN Variant Label automatic activation policy, if applicable
 - (d) Policy for minimizing Whole-Script Confusables and data sources used, if applicable. Also see Additional Note IV
 - (e) IDN Table as per Guideline 6 above.

2.7 Terminology

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

2.8 Additional Notes

- I. For Guideline 13: The use of "uniform" here means that (i) two variant code points or variant code point sequences in one IDN Table cannot be non-variant code points or non-variant code point sequences in another IDN Table implemented under the same TLD, and (ii) all code points in all the IDN Tables under the same TLD must be collectively considered for analysis of variants of code points for each of these IDN Tables. These

two measures are suggested to prevent cases of IDN Variant Labels being generated by different IDN Tables under the same TLD to be allocated to different registrants.

- II. For Guideline 13: Registries may use relevant work for the Root Zone LGR and other sources to determine the variant sets.
- III. For Guidelines 14 and 16: 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.
- IV. For Guideline 17: Registries may use data references such as Unicode's intentional.txt, the cross-script variants in the Root Zone LGR or other authoritative sources.
- V. For example, automatic activation may be considered acceptable practice for Chinese language.
- VI. For example, Japanese language normally mixes Hiragana, Katakana and Han scripts. Also, for Chinese, Japanese and Korean languages, the IDN tables commonly mix "a-z" Latin letters.
- VII. This guideline does not aim to preclude the use of relevant subset of code points with "common" or "inherited" script property in the Unicode standard with the particular language and script, e.g., digits and hyphen.

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 (until 7 April 2017)	GNSO
11	Patrik Fältström (will only review work)	SSAC

Appendix B: Glossary of Relevant Terms

Term	Acronym	Definition	Additional Notes	Other related Terms
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	
Allocatable		An IDN label which can be Allocated		Allocated, Allocation of a Label
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
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
Blocked		State of an IDN label after blocking		Blocking of a Label

Term	Acronym	Definition	Additional Notes	Other related Terms
		The resulting string is a valid label [generated based on a given LGR (or IDN Table and IDN registration rules)] 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		An action taken on a given label with respect to a zone, according to which the label is unavailable for allocation to anyone	As defined in Integrated Issues Report of Variant Issues Project	Blocked
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 Used in the context of Unicode standard in this document	Code Point Sequence
Code Point Repertoire for the Zone		Also known informally as a zone repertoire. A set of code points	As defined in Integrated Issues Report of Variant Issues Project.	Repertoire, Code Point Repertoire

Term	Acronym	Definition	Additional Notes	Other related Terms
		permitted in U-labels in a zone	Used synonymously for Code Point Repertoire or just Repertoire	
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
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
Glyph		A synonym for <i>glyph image</i> . 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	

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	
IDN 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
Internationalized Domain Name Label	IDN label	A label valid as per IDNA 2008		Label
Internationalized Domain Name Table	IDN Table	Specification of permitted code points and combination of those in domains name labels. Also see LGR	Formats specified in RFC 7940, RFC 4290 and RFC 3743	LGR
Internationalized Domain Names	IDNs	Domain names containing characters not included in the traditional DNS preferred form (“LDH”). IDNs under discussion are		

Term	Acronym	Definition	Additional Notes	Other related Terms
		implemented using IDNA		
Internationalized Domain Names in Applications 2003	IDNA 2003		Defined by standard track RFCs 3454, 3490, 3491, 3492 IDNA2003 has been superseded by IDAN2008	IDNA 2008
Internationalized Domain Names in Applications 2008	IDNA 2008		Defined by standard track RFCs 5890, 5891, 5892 and 5893	IDNA 2003
Label		Part of a domain name separated by dots		
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.	As introduced in RFC 7940. Format specified in RFC 7940. Additional formats include those specified in RFC 4290 and RFC 3743	IDN Table

Term	Acronym	Definition	Additional Notes	Other related Terms
		<p>LGRs form part of an administrator’s policies. In deploying Internationalized Domain Names (IDNs), they have also been known as IDN Tables</p>		
Primary IDN Label		An IDN Label applied-for or submitted by a registrant		Label, IDN Label, IDN Variant Label
Variant		<p>The term "variant" is used generally to identify different types of linguistic situations where different code points or labels are considered to be the same (i.e. a variant) of another. Because of the wide-ranging understanding of the term, to avoid confusion more specific terms such as "IDN Variant Code Point" or "IDN Variant Label" should be used</p>		IDN Variant Code Point, IDN Variant Label
Whole Label Evaluation Rules	WLE Rules	Context-based and whole label rules. The “rule” element also contain the character classes that they depend on, and any	As explained in RFC 7940 , Section 6	

Term	Acronym	Definition	Additional Notes	Other related Terms
		actions that assign dispositions to labels based on rules or variant mappings		
Whole-Script Confusables		It may be possible to compose an entire label in a script that will be essentially always identical in form to a label in another script, such as "scope" in Cyrillic looking just like "scope" in Latin. Such strings are called whole-script confusables	Definition derived from http://unicode.org/reports/tr36/#Mixed Script Spoofing	