Guidelines for the Implementation of Internationalized Domain Names

Interim Draft Version

1 Introduction

These Guidelines areabout the implementation of Internationalized Domains Names (IDN) under Internet domains. IDN is standardized by IETF in IDNA2008.

The main target of this document isTop-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 the best current practice. These Guidelinesarealso valuable 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 documentsupersedes 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") will do so in strict compliance with the requirements of the IETF protocol for Internationalized Domain Names in Applications, as defined in RFCs 5890, 5891, 5892, 5893, and 5894.

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. 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.

4. 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.

2.2 Terminology

5. Relevant terminology used in the Guidelines is defined in AppendixB ofthis document with the intention that these definitions will be adopted by the community and used consistently across it.

2.3 Format of IDN Tables

6. A registry will publish one or several lists of Unicode code points[1] 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 variant rulesand the policies attached to it will be clearly articulated.

7. Label Generation Rules ("LGR") 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 a LGR; (b) Registries with existing legacy 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 LGR must include the complete repertoire of code points, any variants and any applicable whole-label evaluation rules which the registry uses to determine if a label is acceptable for registration.

2.4 Consistency of IDN Tables

8. TLD registries are encouraged tocollaborate 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.

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

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 LGRs (i.e. new or modifications of existing ones) that pose any security[2]and/or stability[3] issues will not be authorized to implement such LGRs.

10. TLD registries offering registration of IDNs with the same languageor script tag (RFC 5646) are encouraged to cooperate on the contribution to the development and update of the second level reference IDN tables with the goal of minimizing the difference between the reference table of that language or script and the implemented tables for the same language or script.

2.5 User Acceptance

11. 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. This documentation will include references to the linguistic and orthographic sources used in establishing policies and code point repertoires. 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.

Comment [SH1]: LGRs?

2.6 IDN Variant Labels(Partially Discussed)

12. IDN Variant Labels generated by an IDN Table or a LGR must be allocated to the same registrant or blocked.

---

[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.

//New recommendation proposed by EC: Only IDN Variant Labels with a disposition of "allocatable" may be included in the DNS. IDN Variant Labels may be automatically delegated by the TLD registry in accordance with RFC 3743 (i.e. Preferred Variants), otherwise IDN Variant Labels may be activated when requested by the Registrant (or through a sponsoring Registrar) of the Primary IDN.

2.7 Similarity and Confusability of Labels - TBD

The different kinds of confusability of labels at the second level, arising from homoglyphs, cross-script homoglyphs, relevance of upper case, script mixing and other (e.g. semantic) mechanisms should be managed.

Commingling of cross-script code points in a single IDN table

13. 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[unicode.org]. 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

14. 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 create new variant sets due to the introduction of additional IDN Tables to registry's repertoire. Registries may use relevant work for the Root Zone LGR and other sources.

Cross-script homoglyph labels

15. TLD registries may apply whole-label evaluation rules to new registrations that minimize whole-script homoglyph labels 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 Root Zone LGR homoglyphs references or other authoritative sources. Any policy and its sources will be clearly documented in the registry's public website.

Limitations of IDN tables and policies

In the case of any exceptions made allowing mixing of scripts, 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. TLD registries should also consider policies for visually confusable characters within a same script. Nevertheless, it is important to understand that not all visual confusing similarity issues can be addressed by IDN tables, LGRs and policies. Other policies such as dispute resolution policies may be necessary to mitigate against abusive registrations exploiting

visually similar characters. For example, even for Latin 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 names causes confusion and abuse.

2.8 Registration Data - TBD

WG to look into how to represent and manage registration data for IDNs and for variants of IDNs.

2.9 EPP - TBD

WG to look into any recommendations for EPP, as raised by the community in ICANN 55.

2.10 Non-Conforming Registered Domain Names

Comment [u2]: Is this an appropriate title and place for this recommendation?

16. TLD Registries with existing registered domains 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 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 domains 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, 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 continueing 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 domains of the change of policy and of all relevant dates and conditions which may apply to such domains.

Appendix A: Members of IDN Guideliens 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

Proposed definitions to be included:

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 "IDN Variant", "IDN Variant Character" or "IDN Variant Label" should be used.

IDN Variant (IDN Variant Character and IDN Variant Label)

Variant is defined by an LGR. The term "IDN Variant" maybe used to reasonably describe an IDN Variant Character (code point or code point sequence) or an IDN Variant Label depending on its context. An IDN Variant character is defined in relation to a base character within an IDN Table, such as expressed by an LGR. An IDN Variant Label is a string generated from a Primary IDN based on a given LGR (or IDN Table and IDN registration rules).

Primary IDN

Primary IDN is the string representing the domain name applied for submitted by a registrant.

Term	Acronym	Definition	Additional Notes	Other related Terms
Internationalized Domain Names	IDNs			
	IDNA 2003			
	IDNA 2008			
Code Point				
A-Label				

Term	Acronym	Definition	Additional Notes	Other related Terms
Variant				IDN Variant
Label Generation Ruleset	LGR		Used synonymously for Label Generation Rules	IDN Table
Code Point Repertoire			Used synonymously for Repertoire	
Whole Label Evaluation Rules	WLE Rules			
IDN Table				LGR
Allocatable				
Allocated				
Activated				
Withheld				
Blocked				
IDN Variant				Variant, IDN Variant Code Point, IDN Variant Label
IDN Variant Code Point				IDN Variant

Term	Acronym	Definition	Additional Notes	Other related Terms
IDN Variant Label				IDN Variant