Guidelines for the Implementation of Internationalized Domain Names | <mark>Interim Version for Community Feedback</mark>

This is an interim report drafted to get community feedback on the issues and corresponding recommendations being suggested by the IDN Implementation Guidelines WG.

1 Introduction

//Legacy text (to be edited): *Internationalized Domain Names (IDNs) Implementation Guidelines (IDN Guidelines or the Guidelines) address the IDN registration policies and practices, designed to minimize the risk of cybersquatting and consumer confusion, and respect the interests of communities using local languages and scripts.*

These guidelines are contractually binding for both registries and registrars offering generic Top Level Domains (gTLDs) with IDNs at the second level and recommended for IDN ccTLDs. For example, it part of the Registry Agreement (Specification 6 Clause 1.4) and 2013 Registrar Accreditation Agreement (Additional Registrar Operation Specification Clause 3) and through the Final Implementation Plan for the IDN ccTLD Fast Track Process.

This version of the Internationalized Domain Names (IDNs) Implementation Guidelines ("IDN Guidelines" or the "Guidelines") reviews version 3.0 of the Guidelines following the expansion of the DNS under the 2012 New gTLD Program. The IDN Guidelines are written for gTLD registries and registrars, however the IDN Guidelines are also intended as a support document for other registries establishing IDN policies (e.g. ccTLDs).

This document supersedes version 3.0 of the Guidelines. It was prepared by members of the IDN Guidelines Working Group (IDNGWG) constituted following the Call for Community Experts, and comprises of the following members:

	Name	Supporting Organization/ Advisory Committee
1	Satish Babu	**ALAC**
2	Wael Nasr	**ALAC**
3	Mats Dufberg	**ccNSO**
4	Pablo Rodríguez	**ccNSO**

	Name	Supporting Organization/ Advisory Committee
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**

2 IDN Guidelines

2.1 Transition

IDNA2008 has been adopted by the registries and registrars offering IDNs at the second level. WG should identify and recommend how to address any residual issues from IDNA2003.

//current recommendation 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.*

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, as defined in RFCs 5890, 5891, 5892, 5893, and 5894.

1.2. //current recommendation 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.

2.3. //current recommendation 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.

3.4. //current recommendation 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.

//Also relevant are Appendix A (Comparison of IDNA2003 with IDNA2008) and Appendix B (Additional transitional issues) of version 3.0.

//suggested revision:

In the case of code points permitted in IDNA2008 but disallowed in IDNA2003, those allowed for use in the Root Zone, typically for scripts added in Unicode versions since 2003, should be allowed in labels at other levels. Code points, however, added to IDNA2008 for other reasons should generally be disallowed in the interests of a good user experience and respecting the Longevity Principle in the Procedure.

Commented [SH1]: Chris Dillon: The word "respecting" is meant to cause discussion. We may want to consider "in the spirit of" or a weaker form of words. This touches on an interesting issue — the Procedure is binding at the Top Level, but what is its relationship to the other levels?

Recommendation:

2.2 Terminology

There has been considerable terminology introduced through the work on Label Generation Rules, relevant RFCs and additional IDN work at ICANN for definition and adoption. These include, but are not limited to, the following: RFC 5564, 5890, 5891, 5892, 5893, 5894, 5895, 5992, 6912. Additional work includes the Procedure and additional supporting documents to develop the root zone LGR, the User Experience Study for IDN variant TLDs, the Maximal Starting Repertoire (MSR) and the root zone LGR.

IDNGWG has identified the relevant terms and documented it in Appendix A.

Recommendation:

~~Option 1: Any relevant terminology used in the guidelines is explained inline, when the term is introduced, with no explicit section on terminology/glossary.~~

4.5. ~~Option 2: Any r~~Relevant terminology used in the guidelines is defined in a separate terminology/glossary section in ~~the~~ this document with the intention that these definitions will be adopted by the community and used consistently across it.

~~Option 3: WG, with input from the community, identifies a broader list of terms relevant to the IDN implementation beyond, and not limited to, the terms used in the guidelines document and define these in a separate terminology/glossary section in the document with the intention that these definitions will be adopted by the community and used consistently across it.~~

2.3 Format of IDN Tables

~~Based on work by the community, a formal~~An alternate ~~machine readable~~ specification for representing IDN tables (i.e.~~aka~~ Label Generation Rule~~sets~~s or LGR) is now available ~~and being converted to a standards track RFC by IETF. This format should be encouraged for adoption at second level, as it is being done for Root Zone LGR.~~

5.6. ~~//current recommendation 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.

> **Commented [TTD2]:** I left recommendation #3 unchanged because this recommendation is not about format but about registry obligation to publish such list(s) (regardless of format), albeit the mandatory language or script declaration.

6.7. ~~//current recommendation 4:~~ All such code point listings will be placed in the IANA Repository for IDN TLD Practices in tabular or Label Generation Ruleset (RFC 7940~~(i.e. draft-ietf-lager-specification-13~~) format together with any rules applied to the registration of names containing those code points, before any such registration may be accepted.

> **Commented [TTD3]:** Question to the wg: what would we do if the draft doesn't make it as standard by the time this wg is ready to publish the recommendations up for public comment?. Are we comfortable using a draft (i.e. document with expiration date) as a recommended artifact to use?

> **Commented [TTD4]:** Moved to Section 2.3 for better fit (i.e. format of IDN tables)

~~Recommendation:~~

2.4 Consistency of IDN Tables

The content should be made more consistent across registries and across levels for predictable user experience. This could be done by sharing the LGRs across registries, considering reference IDN tables and other relevant work.

//current recommendation 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.*

//current recommendation 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 by including a provision in its Registry-Registrar Agreement. 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.*

//current recommendation 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 to address common current and emerging challenges in the development and use of IDNs.*

//New proposed recommendation for Ref. second level LGRs: *TLD registries should collaborate to review and contribute to the development of reference second level LGRs so that these meet the expectations of the community. The TLD registries are encouraged to adopt these LGRs, because having common LGRs provides a predictable and consistent end-user experience. In cases where the TLD registries decide to deviate from these reference second level LGRs to suit the particular needs of their specific communities, they should publish and highlight the deviation from the reference LGRs for the general end-users to better understand any differences in LGRs across the TLDs.*

Revised new recommendation re:LGRs: TLD registries seeking to implement new IDN Tables may use available Reference Second Level LGRs as is or as a reference. IDN Tables may deviate from Reference Second Level LGRs. In such case and when deviations pose security and/or stability issues they will require a justification from the registry operator.

TLD registries offering registration of IDNs with the same language tag are encourage 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 and the implemented tables for the same language.

Recommendation:

2.5 IDN Variants

Nomenclature, states of variants and their management process should be made consistent. Relevant policies, e.g. ownership, automatic activation, ceiling value, choice between variants, etc. should be considered and appropriate recommendation should be forwarded.

Topic	Question(s)	Consideration (is this something the IDN Guidelines should cover or not, and if so, what should the recommendation be)
Allocation	• Atomicity (all IDN Variants to be allocated to the same registrant. Is the domain registration (of a primary IDN along with its IDN Variants) considered ONE (1) domain registration or multiple domain registrations? Such that they also renew as ONE (1) domain?) • "Blocked" IDN Variants (are IDN Variants with "blocked" as an LGR disposition considered to be allocated to the registrant?) • Can a dispute be brought to a blocked IDN Variant? • If a dispute brought to a particular IDN Variant sustains, should only that particular IDN Variant be affected or the whole set (including primary IDN and all its IDN Variants) • If a dispute sustains, could a particular IDN Variant be disassociated with the primary IDN and be exceptionally allocated to the winning disputant?	*The group agreed to suggest that variants "should" be allocated to the same registrant, in principle – and not a "must" – but the WG agreed to review this further.* *Thus it was suggested to highlight the possible scenarios and that registries include language in their dispute resolution policy that a label or variant based on its canonical form can be applicable in a dispute. So registrant is aware of such cases and is held to the same level as if the registrant registered the canonical label.*
Delegation	• Only IDN Variants with LGR disposition "allocatable" be delegatable into the DNS • For some languages/scripts, e.g. Chinese, is it acceptable for "preferred variants" as identified by the registry policy to be automatically delegated and activated into the DNS? • For some situations, should a TLD registry limit the number of activated IDN Variants by a reasonable ceiling number (even if such limitation is an arbitrary number) • Should registrants (through registrars) be able to choose particular allocatable IDN Variants to be activated into the DNS? If that is allowed, should it affect the atomicity principle above?	
Childhosts and Nameservers	• Childhosts (when a childhost is created should multiple hosts be created for all activated IDN	

	Variants? If the childhost label is an IDN, should the IDN variants be considered?) • Nameserver Records (when a nameserver is assigned to an IDN, should the same be assigned to its IDN Variants? Should the entire NS RR set be identical?) • If choice activation of IDN Variants are allowed, would the same NS RR set be assigned with an activated IDN Variant automatically?	
WHOIS / Registration Data	• WHOIS search (should all IDN Variants be searchable? Should all allocatable IDN Variants be searchable? Should all activated IDN Variants be searchable?) • WHOIS result (should all IDN Variants return the same result with the primary IDN as the "domain registered"? should all IDN Variants be identified and included in the results? Should all allocatable IDN variants be identified and included? Should all activated idn variants be identified and included? Should the primary IDN always be identified and included?)	
DNSSEC	• Are there any considerations relevant for DNSSEC for IDN Variants (e.g., should the KSK/ZSK be required to be the same for the primary IDN and its IDN Variants?)	
…	•	

Recommendation:

2.6   Similarity and Confusability of Labels

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.

//recommendation 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.

Recommendation:

2.7   Registration Data

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

Recommendation:

2.8   EPP

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

Recommendation:

2.9   User Acceptance

Recommendation:

//current recommendation 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. 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.*

Commented [SH11]: KF: I think this belongs in a usability or user acceptance section. It has nothing to do with consistency.

Commented [SH12]: MD: I do not think that we should require that a registry should publish documentation including references to linguistic and orthographic sources etc. It could be useful for the community of the registries publish the purpose of an IDN table, e.g. for what language, languages or country it is meant. Many ccTLDs base their tables not on a single language but rather on the languages spoken in its country.

It is, however, much more important that the users and stakeholders can compare the IDN tables for different TLDs. That we can achieve if we require that all TLDs use the LGR format of the IDN table as the authoritative source of that table.

Appendix A: Glossary of Relevant Terms

Term	Acronym	Definition	Additional Notes	Other Related Terms
Writing System				
Whole Label Evaluation Rule	WLE Rule			
Variant Label Disposition				
Variant Label				
Variant Code Point Type				
Variant Code Point				
Variant		"Variant" is an ambiguous term, as it can refer to**Variant Code Point** or **Variant Label**, and therefore it should be further qualified whenever it is used.		
U-Label				
Script				
Punycode				

Term	Acronym	Definition	Additional Notes	Other Related Terms
Maximal Starting Repertoire	MSR			
LGR Specification				
Language				
Label: Reserved				
Label: Delegated				
Label: Blocked				
Label: Allocated				
Label: Allocatable				
Label: Activated				
Label Generation Ruleset / Label Generation Rules	LGR			
Label				
Internationalized Domain Names in Applications Protocol 2008	IDNA 2008			
Internationalized Domain Names in Applications Protocol 2003	IDNA 2003			

Term	Acronym	Definition	Additional Notes	Other Related Terms
Internationalized Domain Name Label	IDN Label			
Internationalized Domain Name	IDN	An "internationalized domain name" (IDN) is a domain name that contains at least one A-label or U-label, but that otherwise may contain any mixture of NR-LDH labels, A-labels, or U-labels.	As defined in RFC 5890	A-Label, U-Label
Homoglyph				
Glyph				
Cross-Script Variant Code Points		Variant code points across related scripts, e.g. U+0441 CYRILLIC SMALL LETTER ES 'c' and U+0063 LATIN SMALL LETTER C 'c'		
Code Point Repertoire				
Code Point				
A-Label		An "A-label" is the ASCII-Compatible Encoding form of an IDNA-valid string. It must be a complete label: IDNA is defined for labels, not for parts of them and not for complete domain names. This means, by definition, that every A-label will begin with the	As defined in RFC 5890	U-Label

Term	Acronym	Definition	Additional Notes	Other Related Terms
		IDNA ACE prefix, "xn--", followed by a string that is a valid output of the Punycode algorithm [RFC3492] and hence a maximum of 59 ASCII characters in length. The prefix and string together must conform to all requirements for a label that can be stored in the DNS including conformance to the rules for LDH labels. If and only if a string meeting the above requirements can be decoded into a U-label is it an A-label.		