

Internationalized Domain Name (IDN) Implementation Guidelines

Publication Date: 11 September 2017

Prepared By: IDN Guidelines Working Group (IDNGWG)

Section III: Summary of Comments

General Disclaimer: This section intends to summarize broadly and comprehensively the comments submitted to this public comment proceeding but does not address every specific position stated by each contributor. The preparer recommends that readers interested in specific aspects of any of the summarized comments, or the full context of others, refer directly to the specific contributions at the link referenced above (View Comments Submitted).

The text in blue below will be moved to the next section once finalized. It has been interspersed in this section for ease of review.

CL1: A clarification is suggested regarding the use of IDN labels inside a TLD zone for records that are not-authoritative, e.g. the NS and glue records. So a TLD can restrict a Unicode point for registration purposes, but it could exist inside the TLD zone. So “labels inside a zone” should be clarified to mean “labels of authoritative names inside a zone”.

CL1 Analysis: This is a useful observation and ~~we-the WG~~ agrees that the IDN label the guidelines are currently referring to are those which are authoritative in the zone. We will clarify this is meant by default and call out any specific reference to non-authoritative data in the guidelines explicitly to ensure there is no ambiguity.

CL2: Add a new guideline: "a TLD can't restrict the codepoints of names inside its zone for which it's not authoritative (such as delegations to sibling zones or glue records names), but should check such labels are syntactically valid U-labels (in RFC7940 sense)".

CL2 Analysis: The WG is currently considering if the guidelines should cover data which is not authoritative for the relevant zone. If the WG considers it relevant after discussions, appropriate text will be added, either as part of the discussion on scope or else as a guideline specific for such data.

O-X refers to a recent post online which directed users to disable display of IDN URLs in browsers to prevent phishing using whole-script confusable domain names. O-X opines that this mindset would widely reject IDNs throughout the internet. Therefore, it is important to

prevent dangers associated with whole-script confusable domain names.

O-X1: Such dangers can be avoided by establishing a basic principle that two domain names that look confusable to an average internet user must be considered variants and must never be registered to different registrants. This be established as a cornerstone for IDN Guidelines.

O-X2: While confusability is a subjective feature, technical standards like UTR#39 of Unicode provide implementable definition and algorithm. These standards should be implemented.

O-X1-O-X2, Gol1-Gol4 Analysis: There are two aspects to this comment. First, that confusable labels be considered as variant labels, and second that variant labels be allocated to same registrant.

The second restriction is captured by the WG in a separate recommendation.

Regarding the first part of the comment, the WG ~~has considered~~ is considering visual confusability of labels in detail. The WG considers visual confusability is not binary in nature, but a continuum from homographs on one end to labels which are visually distinct on the other end, ~~and having with~~ many different ~~kind of~~ cases in between which may not be true homographs or completely distinct, but what may be termed as similar but still confusing. It is difficult to delimit this continuum of cases into two categories, without having authoritative data as reference. It is out of the scope of the WG to create such data. The WG has reviewed data published by the Unicode consortium, through IETF RFCs and by ICANN and still has not found a comprehensive source. Specifically, in the context of UTR 36 and UTR 39, the WG has found that the data file *intentional.txt* is not complete at this time and the file *confusables.txt* has many cases which are not strictly homoglyphic in nature.

The WG notes that putting a strict requirement, without having an authoritative source of data to implement such a requirement consistently across the TLDs, may not address the problem effectively. Thus, ~~still~~ noting the significance of the issue, the WG has put in a series of guidelines for the TLD registries in Section 2.5, but without the force suggested. The WG hopes that as the clear and implementable authoritative data sources for scripts becomes available, e.g. through the Root Zone Label Generation Rules being developed by the community, the guidelines can be made more precise. Until that time such analysis has been left to the registry to undertake.

O-X3: Point 16 of the draft guidelines should be reworded to the following: "TLD registries must apply to new registrations whole label evaluation rules that minimize whole-script confusables as determined by Unicode Technical Standard #39: Unicode Security Mechanisms; new domain names that according to those rules are whole-script confusables in respect to an existing domain name must be a) allocated to the same registrant of the existing domain name, or b) blocked from registration."

O-X3 Analysis: As discussed earlier, due to lack of comprehensive authoritative data which can be consistently implemented, the WG considers that requiring such analysis will not address the confusability (as the analysis will remain inconsistent across TLDs, which will have to determine their own data). Therefore, the WG is not requiring but only encouraging this analysis at this time.

IAB finds the document to be a good step in support of the deployment of IDNs in TLDs, promoting use of IDNs, supporting specification and conservatively implementing registry and registrar policy.

IAB1: IDN Guidelines is not a protocol document. Thus, it will be helpful to clarify the RFC 2119 terminology defines what it means to comply with the guidelines but cannot specify protocol compliance.

IAB1 and RySG1 Analysis: The WG agrees with the comment and, therefore, reference to RFC 2119 ~~has been~~will be removed to avoid any confusion.

IAB2: Section 2.1 is clear statement of compliance with IDNA 2008 as a target with transition guidelines from other specification is appropriately conservative.

IAB2 Analysis: ~~The has been the intention of the WG has been to make a clear statement for compliance with IDNA 2008.~~ The WG thanks IAB for supporting it.

IAB3: In Sections 2.2 or 2.3, it may be useful to explicitly state that sometimes various combinations of languages and a script (and variants) are not fully compatible and registry has to make a decision.

IAB3 and IAB5 Analysis: WG considers that IAB is asking to make ~~a general statements on motivation behind, in support of the proposed guidelines.~~ General statements may not be actionable so the WG ~~has tried~~is attempting to follow up with IAB for further clarification ~~and still awaiting a response.~~ In case, any additional actionable details are received by IAB beyond what has been already proposed, the WG would consider including them in the relevant guidelines.

IAB4: In Section 2.4, no. 13, the final paragraph is not clear. Is it an example of guideline in previous paragraphs for registry-side processing of IDN variants?

IAB4 Analysis: The WG ~~has split~~is splitting the recommendation into three parts, and ~~deleting~~are being referred to, ~~to clarify~~for clarifying the guideline. The paragraphs ~~being referred to have been~~are being deleted as the WG ~~has considered~~is that these were explanatory texts and thus not needed as part of the guideline.

IAB5: In Sections 2.2 to 2.5, the document is trying to operationalize provisions of RFC 5894 that registries have a policy and that they allow only those characters they fully understand. It

will be helpful to state that as a goal for these guidelines.

IAB5 Analysis: ~~TBD – WG to discuss the statements in RFC 5890-94 and 6912 and finalize how to address this. See IAB3 analysis above.~~

GoI emphasizes the universal acceptance of internationalized domain names and email addresses.

GoI1: The recent security advice to disable display of IDN URLs in browsers to prevent phishing due to whole-script confusable domain names is pointed out. It is noted that this would widely reject IDNs throughout the Internet. Thus, ICANN should prevent such issues.

GoI2: To avoid such confusion, it is suggested to establish a basic principle that any two domain names that look confusable to an average Internet user must be considered variants of the same domain name and must never be registered to different registrants.

GoI3: Though confusability definition is subjective, technical standards like Unicode TR#39 provide implementable definition and algorithm to detect confusable domain names, and should be implemented through the guidelines.

GoI4: Confusing registrations can not only hamper IDN adoption, but also has financial costs, even before phishing occurs. Thus it is efficient to detect and prevent these at the registry level. Thus, for whole-script confusables, “may” should be changed to “must”. Further, Point 16 should be rephrased to: “TLD registries must apply to new registrations whole label evaluation rules that minimize whole-script confusables as determined by Unicode Technical Standard #39: Unicode Security Mechanisms; new domain names that according to those rules are whole-script confusables in respect to an existing domain name must be a) allocated to the same registrant of the existing domain name, or b) blocked from registration.”

~~**For Analysis of GoI1- GoI4,:** In addition to the points already discussed by the WG in response to **see O-X1, O-X2 Analysis** (as these comments are similar), the WG would also like to add that it considered the additional point made by GoI regarding ASCII labels. However, such cases regarding ASCII labels are not being analyzed by the WG as these are considered out of scope of the current review focused on IDNs.~~

GoI5: There should be a separate guideline to deal with emojis.

~~**Analysis of GoI5:** After deliberating on this comment, As Emojis are not PVALID, the WG discussed and has agreed concludes that the existing requirement to adhere to the IDNA2008, already in the guidelines, addresses the guidance suggestion on for not using allowing Emojis, because Emoji are DISALLOWED by IDNA2008. , and explicitly calling them out is not needed in the context of these Guidelines.~~

RySG1: RFC 2119 limits the use of this imperative language and defines that “they MUST only

be used where it is actually required for interoperability or to limit behaviour which has potential for causing harm'. In addition, RFC 2119 asks document authors to 'elaborate the security implications of not following recommendations or requirements'. Draft Guidelines introduce requirements not strictly necessary for interoperability or to limit potential harm. So it is advised to use the imperative language 'with care and sparingly' and 'elaborate the security implications of not following recommendations'.

See IAB1 Analysis.

RySG2: Revise Guideline 1 to read “ (...) as defined in standards track RFCs 5890, 5891, 5892 and 5893 and their successors” by adding “and their successors” at the end.

RySG2 Analysis: The WG agrees and has added “or any RFC that replaces or updates the listed RFCs” to the guideline.

RySG3: Revise Guideline 4 to add “both” for emphasis and clarity to read “label containing hyphens in both the third and fourth positions (...).”

RySG3 Analysis: The WG agrees and has added “both” to the guideline.

RySG4: Guideline 5 is overly prescriptive, because registry operators are in a better position to design a communication plan to address any policy update that affects the live cycle of domain names under its TLDs.

RySG5: Guideline 5 should clarify that pre-existing domain names are not mandated to comply with these guidelines. The relevant part of the guideline should be changed to the following: “TLD registries with IDNs that were registered prior to the implementation of these guidelines and which do not conform to these guidelines are not required to comply with the guidelines, but should take the following actions for these pre-existing domain names to reduce disruption to registrants and Internet consumers.”

RySG4 and RySG5 Analysis: The WG ~~has simplified~~ is simplifying the guideline ~~and made to~~ make it less verbose. Though the registry decides on how pre-existing domain names which do not conform to changes will be addressed, and the guideline does not require the existing non-compliant domain names to be removed, the WG considers that the intentions of the registry for such domain names should be clearly communicated to the registrants for them to be aware of the latest registration policy and its implications on their registrations. The revised guideline will still capture ~~s~~ these relevant details which the WG considers are important.

RySG6.1: As per Guideline 7, the RySG acknowledges the benefits of use of the LGR RFC format as it allows a better adoption and easier comparison and supports this as a long-term strategy. It has no effect in minimizing the risk of cybersquatting or consumer confusion, two of the stated goals of these IDN guidelines, and not required for interoperability. Therefore,

registry operators should not be required to use a new format (i.e. RFC 7940) to publish IDN Tables.

RySG6.2: it is underscored that a transition to a general use of the LGR format would require a long enough transition period for Registries to create new code tables, distribute them, and put them into effect, and that ICANN or IIS should provide validation tools to review these new tables before the policy becomes mandatory.

RySG6.3: It is also noted that the new gTLD contract requires IDN Tables to be submitted to IANA for publication in the IANA Repository, but that not all submitted tables have been published.

RySG6.4: For the reasons mentioned the RySG recommends amending draft guideline 7 as follows: "IDN tables must be submitted to IANA for publication in the IANA Repository for IDN Practices. Further, (a) Except as applicable in 7(b) below, registries are encouraged to use Label Generation Ruleset (RFC 7940) format to represent an IDN table; (b) Registries with existing legacy IDN tables already submitted for inclusion in the IANA Repository for IDN Practices at the time these guidelines are published are encouraged to transition to the LGR format (...)."

RySG6 Analysis: The WG considers that the LGR format in RFC 7940 contributes significantly to the interoperability of the IDN tables. Earlier formats cannot formally capture all the information including repertoire, variant code points, variant types, label dispositions and label-level evaluation rules. The WG thinks that interoperability should be looked at in a broad sense beyond registries and registrars, e.g. also including application developers, who may find it easier to implement the LGR, due to its well defined format. As another example, it is useful for moving a TLD between RSPs Registry Service Providers. Thus, this format is ~~needed~~ useful for interoperability and promotes usability, even if does not directly address confusability or cybersquatting.

~~In the case of complex rules there is a higher possibility of different registries implementing them differently even using the same IDN table, which can cause consumer confusion. Thus, having a clearer format helps address the problem alleviates the risk of cybersquatting and consumer confusion.~~

The WG has also considered RySG comments 6.2-6.4 and notes that some of these are implementation level comments by RySG – e.g. transition time and tool requirements, so may not be addressed by the guidelines. The WG notes that ICANN has already made a LGR Tool to process IDN tables in LGR format available online and also released its code for re-use. There are also continued discussions on making the process of publication of IDN tables at IANA platform more effective. The WG realizes the implementation challenges registries may face and, therefore, the WG will consult with RySG to determine a reasonable time to implement the LGR format, instead of requiring it immediately to come in effect immediately after the publication of the guidelines and will note it in the guidelines.

RySG7: It is suggested that for Guideline 9 the definition of stability is too broad and too open for interpretation for IDN Guidelines. The relevant standards should be only standards-track or Best Current Practice RFCs by IETF.

RySG7 Analysis: The WG does not intend to define the terms “security” and “stability”. It ~~has~~ ~~is taken away~~ deleting the reference to the definitions ~~provided, which were originally~~ reused from the Registry Services Evaluation Policy of the gTLDs, and will be using these terms without an explicit definition, with the intention that relevant applicable definitions will be inherited from existing arrangements, e.g. from the respective contracts for the gTLDs.

RySG8: The term “same registrant” should be defined in Guideline 12.

RySG8 Analysis: The wording ~~has been~~ will be updated to “same registrant as the primary IDN label” to make it clearer.

RySG9: Guideline 12 should be further clarified by stating that registry operator must publish the variant activation policy in its public website.

RySG9 Analysis: A separate recommendation is being added asking the registry operator to publish the variant allocation and activation policy in its public website.

RySG10: The term “Registry-side approach” should be defined in Guideline 13, and the recommendation should be reworded for clarification. Second and third paragraphs should be removed or moved to a separate section and change “must” to “should” or “may” to make it an advisory.

RySG10 Analysis: This guideline ~~has been~~ is being reworded based on the suggestion. ~~with~~ The second and third paragraphs are being removed. The guideline ~~is~~ will be divided in two parts, ~~where~~. The first part will relay that it is generally expected that registration is done on the request of the registrant. ~~Only~~ The second part will suggest that only in exceptional cases automatic activation by the registry should be done for a script, and such cases should remain minimal.

RySG11: The first sentence of draft guideline 15 is confusing as it is not clear if the guideline refers to cross-TLD tables of the same Registry or to a single same TLD. The recommendation may not be workable as one Registry may have two or more TLDs in the same script but directed at different languages where variants must be handled differently. Likewise, two TLDs, one script-based and the other language-based, in the same script may have different variant tables. It seems that the only viable interpretation is that the draft guideline refers to one single TLD, in which case this should be clarified in the guideline.

RySG11 Analysis: The WG intended to do it for IDN tables across the same script for a specific TLD. The recommendation will be updated to clarify that this be done “for a particular TLD”.

RySG12: The RySG supports the view that the IDN Guidelines are not the right place for recommendations on registration data and EPP.

RySG12 Analysis: The WG will remove sections pertaining to registration data and EPP from the final version of the Guidelines.

RySG13: It is suggested to replace 'word' by 'label' in the definition of "variant": 'The term "variant" is used generally to identify different types of linguistic situations where different labels are considered to be the same (i.e. variant) of another label. Because of the wide-ranging understanding of the term, to avoid confusion more specific terms such as "Variant Code Point" or "IDN Variant Label" should be used.'

RySG13 Analysis: The WG ~~has~~will updated the definition accordingly.

Section IV: Analysis of Comments

General Disclaimer: This section intends to provide an analysis and evaluation of the comments submitted along with explanations regarding the basis for any recommendations provided within the analysis.

The IDN Guidelines WG is thankful to the CL, O-X, IAB, Gol and RySG for their thoughtful comments.