Internationalized Domain Names Expedited Policy Development Process

A6 & A7 Introduction

EPDP on IDNs Team Call #15 | 9 December 2021

A6 Introduction

Charter Question A6

Since RZ-LGR can be updated over time, the WG needs to consider the implications for existing TLD labels and their variant labels (if any), including any potential changing of status or disposition value.

The TSG further recommends that the Generation Panel (GP) must call out the exception where an **existing TLD is not validated** by their proposed solution during the public comment period and explain the analysis and reasons for not supporting the existing TLD in their script LGR proposal. This will allow the community and the GP to review such a case to confirm that an exception is indeed warranted.

Does the WG agree with **TSG's suggested approach**? If so, to what extent should the TLD policies and procedures be updated to allow an existing TLD and its variants (if any), which are not validated by a script LGR, to be **grandfathered**? If not, what is the recommended approach to address changes to the current version of the RZ-LGR that assign different disposition values to existing TLDs? Consider this question by taking into account the data to be collected in the "Data and Metric Requirements" section of this charter.

Scope of Question A6

- Focus on **future update of the RZ-LGR and its implication** on existing gTLDs in the future
  - Based on data collected, all current existing TLDs (including those in the pipeline for delegation) are valid

Trigger Events for RZ-LGR Update

Analysis: *Trigger events seem to result in adding more materials to the RZ-LGR, rather than subtracting*

Some trigger events include, but not limited to:

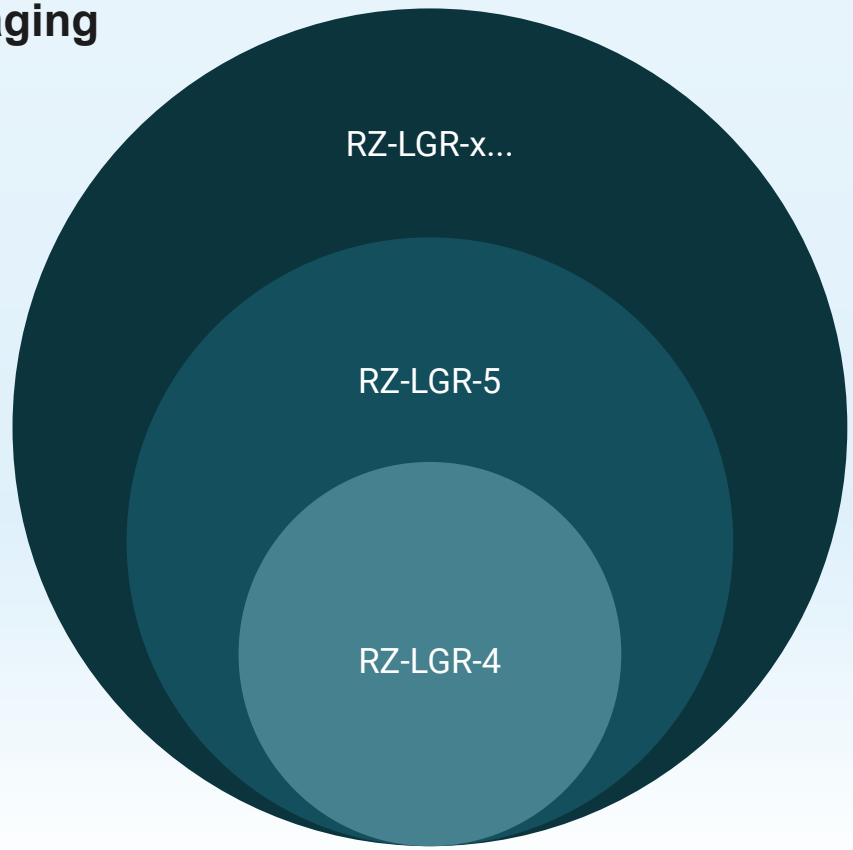
- Evidence that an additional existing code point is needed for one of the languages considered
- Additional language being considered, not considered before for the script; with reasons why the language should be considered now (e.g., change in language EGIDS value)
- A constraint on labels in a script can be relaxed without issues to accommodate a particular language
- Update in Unicode version, with additional code points available for a script

RZ-LGR Backward Compatibility

Analysis: *Invalidating an existing gTLD and its variant labels (if any) by the proposed RZ-LGR update is extremely unlikely, as it will cause instability in the root-zone*

RZ-LGR-4 Overview and Summary | 4.3 Comprehensiveness and Staging

Ideally, the Root Zone LGR would be comprehensive, that is, **include all scripts eligible for the root zone from its first version**....The goal for all future versions of the LGR must be to **retain full backward compatibility**, so that they **preserve the output of any label registration against the old LGR**, when applied to an updated LGR. Consequently, the IP anticipates that **succeeding versions of the LGR will be strict supersets of their predecessors**. It is expected that **registrations that predate the initial release of an LGR** covering the respective script **will be allowed to remain, even if in conflict**, but without becoming a binding precedent for the LGR itself. To date, **there is no known instance of such a conflict**.



TSG Recommendation 12

GPs should aim to make the RZ-LGR backward compatible with existing TLDs, to the extent possible, towards maintaining the stability of the Root Zone as recommended by SSAC's SAC060.

In the event that backward compatibility cannot be achieved, the GP must call out such an exception (that an existing TLD is not validated by their proposed solution) during the public comment period and explain the analysis and reasons for not supporting the existing TLD in their script LGR proposal. This will allow the community and the IP to review such a case to confirm that an exception is indeed warranted.

Question for the EPDP Team:

- Does the EPDP Team agree with **TSG's suggested approach**?
- If so, to what extent should the TLD policies and procedures be updated to **grandfather** an existing TLD and its variants (if any), which are not validated by a script LGR?

A7 Introduction

Charter Question A7

What mechanism or criteria should be used to identify the scripts/languages appropriate for single-character TLDs? Once those scripts/languages are identified, **what mechanism or criteria should be used to identify a specific list of allowable characters** which can be used as a single-character TLD within such scripts/languages? Should any specific implementation guidance be provided? Furthermore, **should the relevant GP tag these code points in the RZ-LGR** for a consistent analysis and to ease their identification and algorithmic calculation?

- *Single character: a character in a U-label*

A7 Origin

- SubPro Recommendation 25.4 (Feb 2021)
- TSG Report Appendix B (Oct 2019)
- SAC052 Recommendation 1 (Jan 2012)
- JIG Final Report on Single Character IDN TLDs (Mar 2011)

SubPro Recommendation 25.4

Single character gTLDs may be allowed for limited script/language combinations where a character is an ideograph (or ideogram) and do not introduce confusion risks that rise above commonplace similarities, consistent with SSAC and Joint ccNSO-GNSO IDN Workgroup (JIG) reports.

- *SSAC Report: SAC052*
- *JIG Report: JIG Final Report on Single Character IDN TLDs*

Rationale:

- It is appropriate to limit single-character gTLDs to only certain scripts and languages
- SubPro does not have the relevant expertise to make this determination
- Welcome the identification of the limited set of scripts and languages and potentially a specific list of allowable single-character gTLDs (e.g., during implementation)
- Substantially increase the predictability of what will likely still remain a case-by-case, manual process

TSG Report Appendix B

Historically, single character TLDs have not been allowed due to their confusability potential. The SG advises GNSO and ccNSO to **review SSAC's SAC052** on the delegation of single character IDN TLDs.

In the event that certain range of code points or entire scripts are permitted to be used for single character TLD applications based on certain criteria, **it may be useful that those code points are appropriately tagged by the relevant Generation Panel in the RZ-LGR for a consistent analysis** and to ease their identification and algorithmic calculation.

SAC052 Finding

Single-character TLDs are more likely to cause user confusion than TLDs with more than one character

- The more characters a domain label has, the easier it is for a user to "infer the context"
- Code points in closely related script blocks may represent characters that are confusingly similar despite belonging to different scripts
  - Latin/Greek/Cyrillic
  - Gurmukhi/Bangla/Devanagari
  - Thai/Lao
- Not aware of a comprehensive inventory of confusable scripts
- Ideographic scripts such as Han, not only can a single character represent a complete “word” or idea, but in some cases different single characters can represent the same “word” or idea

SAC052 Recommendation 1

Given the potential for user confusion and the currently unfinished work on string similarity and IDN variants, the SSAC recommends a very conservative approach to the delegation of single-character IDN top-level domains.

Until ICANN completes its work on user confusion/string similarity and IDN variants, the SSAC recommends:

1. Delegation of all single-character IDN TLDs in all scripts should be disallowed by default.
2. Exceptions may be made for some scripts, but only after careful consideration of potential confusability both within and across scripts. Such consideration should invite comments from the technical and linguistic community, and from ICANN's advisory committees.
3. Single-character TLD applications in an exceptionally allowed script should be accepted only **when there is clear evidence that there is no risk of user confusion**. **Each applied-for single-character TLD label must be explicitly examined across scripts** to ensure that there is absolutely no possibility of user confusion within or across scripts.
4. **ICANN should consult with the technical and linguistic community** to determine which scripts, if any, should be restricted with respect to the delegation of single-character TLDs, and how any such restrictions should be defined, and how such restrictions may be relaxed if appropriate.
5. ICANN should take into consideration the outcome of the **IETF work on the creation of a concise specification of the TLD label syntax** based on existing syntax documentation, extended minimally to accommodate IDNs.

SAC052 Recommendation 1 (Cont.)

6. ICANN should consider adopting the following guidelines regarding its consideration of which scripts and code points could be accepted as exceptions:

a) The code point must be PVALID according to IDNA2008.

b) The code point is from one of the following Unicode categories: lower case letter (Ll), upper case letter (Lu), and other letter (Lo) as defined by the Unicode Standard.

c) Some single-character IDN TLDs are composed of multiple Unicode code points, which may include non Lx-class code points. These should be subjected to a more stringent technical and confusability analysis, whose criteria should be well defined and made public.

d) The script in which an exception is made and a single character IDN is allowed **should not have characters that are intrinsically confusable with characters of another script (for example, Latin/Greek/Cyrillic, Lao/Thai, etc.).**

e) The existing and extended rules of confusability must be met. Single-character code points must explicitly be examined across scripts. Denial of a single-character TLD application does not imply blocking of the script. Similarly, acceptance of a single-character TLD application does not imply acceptance of the script.

f) If a script is allowed, **a distinct and explicit specification of which subset of the script is available for single-character TLDs** should be required prior to the acceptance of a single-character TLD application. **By default all characters are disallowed**, even when a script is allowed, and an explicit single-character TLD-allowed list must be generated for each case.

JIG Final Report on Single Character IDN TLDs

Implementation Recommendation on Single Character IDN TLDs

D. Requested Single Character IDN TLD strings should be analyzed on a case-by-case basis in the new gTLD process depending on the script and language. Single Character IDN TLDs should be acceptable, but must not be confusingly similar to single or two character ASCII TLDs. For alphabetic script Single Character IDN TLDs, other technical aspects of confusability may be taken into consideration, such as the likelihood of user slip with relevance to keyboard layouts.