

Internationalized Domain Names Expedited Policy Development Process

A4 Analysis & A5 Introduction



EPDP on IDNs Team Call #14 | 2 December 2021

A4 Analysis

Charter Question A4 (emphasis added)

For future gTLD applications, the SubPro PDP proposes an implementation guidance that if a script is not yet integrated into the RZ-LGR, applicants should be able to apply for a string in that script, and it should be processed up to but not including contracting.

Applicants under such circumstances should be warned of the possibility that the applied-for string may never be delegated and they will be responsible for any additional evaluation costs. The burden in this case is on the applicant, who may have to wait for an indeterminate amount of time but is not aware of any other serious concerns. The SubPro PDP developed this implementation guidance by taking into consideration the TSG recommendation that the application should remain on-hold (or other appropriate status) until the relevant script is integrated into the RZ-LGR.

The WG and the SubPro IRT to coordinate and consider the following questions in order to develop a consistent solution: **should the SubPro recommendation be extended to existing TLDs that apply for a variant TLD label whose script is not yet supported by the applicable version of the RZ-LGR?** Consider this question in tandem with **b4)** and by taking into account the data to be collected in the “Data and Metric Requirements” section of this charter. If not, what should be the process for an existing TLD registry who wishes to apply for a variant TLD label whose script is not yet supported by the applicable version of the RZ-LGR?

Scripts of Existing gTLDs & RZ-LGR Status

Q1: Which scripts are used in existing gTLDs?

Arabic, Chinese, Cyrillic, Devanagari, Hebrew, Japanese, Korean, Latin, Thai

Q2: Which scripts of the existing gTLDs are supported by the latest RZ-LGR?

RZ-LGR-4 has integrated Arabic, Chinese, Devanagari, Hebrew, and Thai

Q3: What's the status of scripts of existing gTLDs not yet supported by the RZ-LGR?

- Cyrillic & Korean: script proposals completed; waiting for integration into RZ-LGR-5
- Japanese & Latin: public comment for script proposals completed; community feedback for Japanese proposal is all positive, but has some divergence for Latin proposal;
Next Steps: GP to finalize script proposals and IP to consider integration into RZ-LGR-5

Q4: What's the estimated timeline to launch RZ-LGR-5?

RZ-LGR-5 is estimated to launch in Mid-2022, covering all scripts of existing gTLDs

Further Analysis of Existing Latin Script gTLDs

- 1,170 existing gTLDs in Latin script, including .web and .webs in the pipeline
- All Latin script gTLDs, except for .vermögensberater and .vermögensberatung, are ASCII labels
- Delegated ASCII gTLDs do not have any allocatable variants, only blocked ones
- Only allocatable variants within the Latin script are ß (sharp S in German) and ı (dotless i in Turkish), but no existing gTLDs contain these letters

Conclusion: *Even if the RZ-LGR-5 launch is delayed, it is unlikely that existing Latin script gTLD operators could apply for allocatable variants that do not exist*

A5 Introduction

Charter Question A5

SAC060 notes that variant code points in LGR may introduce a “permutation issue”, possibly creating a large number of variant domain names, which “presents challenges for the management of variant domains at the registry, the registrar and registrant levels.” SAC060 advises that “ICANN should ensure that the number of strings that are activated is as small as possible.” The TSG agreed with this SSAC advice. Appendix C of the Staff Paper reviewed the factors causing numerous variant labels and suggested measures to address this issue.

Should there be a ceiling value or other mechanism to ensure that the number of delegated top-level variant labels remains small, understanding that variant labels in the second level may compound the situation? Should additional security and stability guidelines be developed to make variant domains manageable at the registry, registrar, and registrant levels?

SAC060 Recommendation 14: ICANN should ensure that the number of strings that are activated is conservative

- Variants introduce a permutation issue both at the top-level and with combinations of top-level and second-level
- **Large number of variant strings presents challenges for the management of variant domains at the registry, the registrar and registrant levels**
- Some registries have imposed additional rules for variants, e.g., “no mixing”
- A variant TLD application must be accepted only if the TLD applicant clearly demonstrates the necessity for activating the string
- Variants that are not necessary, but are desired, must not be allocated and activated
- Early input from SSAC reiterates this point

TSG Recommendation 14: SSAC advises in SAC060 that too many variant labels should not be delegated. The SG considers that the matter on limiting the number of allocatable variant labels to be a policy matter. Refer to Appendix C of IDN Variant TLD Implementation: Appendices for some suggested approaches.

Appendix C of Staff Paper: Limiting the IDN Variant Domain Names with the Delegation of IDN Variant TLDs

Reviewed the factors causing numerous variant labels and suggested measures to address this issue

Appendix C: Cause of Over-Production of Allocatable Variant TLDs

RZ-LGR is optimized on script level but may not be on language level, but **over-production is not always the case**

- **Difference in analyzing variant labels for RZ-LGR compared to their use by end users**

GP designs RZ-LGR by looking at the wider use across all languages and across all communities using the relevant script, whereas TLD applicant focuses on a particular language when identifying a TLD label or its variant label

- **Use of Same Script across Different Writing Systems**

Variant code points may be over-generated in cases where the same script is used across different writing systems

- **Usage Conventions**

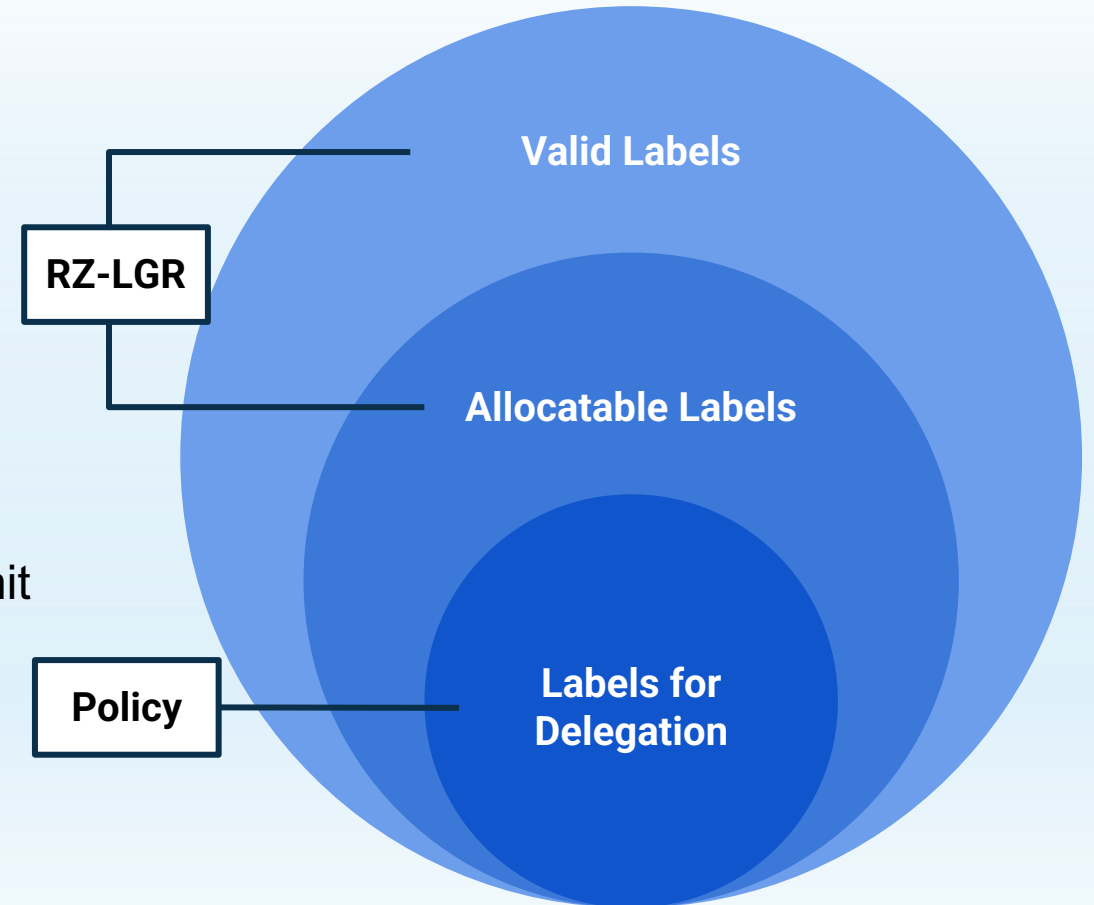
RZ-LGR may allow some allocatable variant labels which are not well-formed according to usage conventions

- **Meaningfulness of Variant TLD Labels**

Lack of association or meaningfulness for certain categories of labels (e.g., brands, geographic names, community names, country codes, etc.) may be considered as their overproduction

Appendix C: Limitations of Using Label Disposition in RZ-LGR

- RZ-LGR mechanically or algorithmically create a minimal set of valid and allocatable variants
- In some cases, more allocatable variant labels could be generated by RZ-LGR for an applied-for label
- “Allocatable” label does not mean it necessarily will or should be delegated
- Other steps in the registration process are expected to limit the set of labels for delegation
- Labels for delegation could only be further limited through a policy → a separate subsequent evaluation process to choose suitable labels from among the allocatable labels



Appendix C: Manage the Permutational Challenge

Staff Paper Suggestion: Community should deliberate on developing appropriate **policy and procedures** to manage the permutational challenge using a three-tier approach

Minimize delegation of IDN variant TLDs

Minimize registration of second-level IDN variant labels

Minimize IDN domain registration (by combining variant at multiple levels)

Appendix C: Minimize Delegation of IDN Variant TLDs

1. The IDN variant TLD label must be **specified for a specific language community**, which should have been supported by the **relevant script Generation Panel** in their proposal and is present in the **Common Locale Data Repository (CLDR)**
2. The IDN variant TLD for the target language should be **valid label using that language based Reference Second Level LGR** released by ICANN, **if available**
 - In case the **reference LGR is not available** or the applicant requires additional characters to support the language, the IDN variant TLD label should be valid using the **relevant language based second-level IDN table proposed for it**
3. The variant label **must be usable**. People of the target community **should be able to compose** the variant TLD using **generally available input method editors (IME)**
4. The variant label should **follow the orthographic conventions** of the script and the relevant language community
5. The variant label must **demonstrate association or meaningfulness** in relevant cases

Appendix C: Minimize Delegation of IDN Variant TLDs (Cont.)

6. **Ceiling Value:** additional policy may be developed to propose a ceiling value to ensure that the number of delegated variant TLD labels remains small

- E.g., Chinese community allows for three variant labels: 1) applied-for label, 2) simplified Chinese version, 3) Traditional Chinese version
- Similar limit may be imposed in the beginning to **ensure conservatism**, and which **may be relaxed over time**
- Take further input from the script-based Generation Panels who have developed the RZ-LGR proposals

Appendix C: Minimize Registration of Second-Level IDN Variants

1. **Language-based IDN tables** (instead of script-based IDN tables) should be used where possible
2. IDN Tables should be designed to include the following:
 - Code points, based on the principles in RFC 6912
 - Variant code points, using variant code point types which resolve to either “blocked” and “allocatable” dispositions for all valid labels as per RFC 7940, **maximizing blocked variant labels**
 - Label-level rules to **further reduce the valid or allocatable labels** generated
3. Variant labels produced using code points **not contained in a single language-based IDN table** should be given a **blocked disposition**
4. Script communities considering **free or automatic activation of variant labels** should develop guidelines on **how to limit such labels to a small number**
5. Additional policy may be developed to propose a **ceiling value** to ensure an arbitrary upper limit on the variant labels registered at the second-level

Appendix C: Minimize IDN Domain Registrations

1. **Registration policies** need to be considered and extended with **additional constraints** to manage consequences of combining variant labels at top-level and second level.
2. **Registration policy being applied for second-level should be consistent with the top-level**
 - e.g., usability argument could be extended to reduce domain names
3. Additional restrictions, such as **putting a ceiling value**, should be considered in case the algorithmic solution and policy remain insufficient to contain the domain name permutations
 - e.g., an overall limit of three variant domain names could be proposed, as practiced by the Chinese community
 - Community needs to deliberate on what could be a good starting value and then relax it over time based on experience
4. Similar practices may also be promoted for the **third and other levels**, as applicable

Distilling Charter Question A5

Question A5: Should there be a ceiling value or other mechanism to ensure that the number of delegated top-level variant labels remains small, understanding that variant labels in the second level may compound the situation? **Should additional security and stability guidelines be developed** to make variant domains manageable at the registry, registrar, and registrant levels?

1. Should the number of allocatable variant TLD labels for delegation be minimized?
2. What is the criteria for minimizing delegation of allocatable variant TLDs?
3. What is the process to determine whether an allocatable variant TLD meets the criteria for delegation?

This question also ties to Charter Question B4:

What should an application process look like in terms of timing and sequence for an existing and future Registry Operator with respect to applying or activating their allocatable variant TLD labels?