Recommendations for the Technical Utilization of the Root Zone Label Generation Rules (RZ-LGR)

EXECUTIVE SUMMARY

With the availability of the Root Zone Label Generation Rules (RZ-LGR), the ICANN Board asked the ICANN community to recommend how to technically apply the RZ-LGR in a harmonized way for existing and future country code top-level domains (ccTLDs) and generic top-level domains (gTLDs). Following the Board request, a call for experts was announced and a Study Group (SG) on the technical use of the RZ-LGR was formed from the Supporting Organizations (SOs), the Advisory Committees (ACs) and the Internet Architecture Board (IAB). The SG has deliberated a range of issues from the viewpoint of those producing the RZ-LGR (Generation Panels (GPs) and the Integration Panel (IP)), those using the RZ-LGR (e.g. TLD registries or future TLD applicants) and those developing policy for the use of the RZ-LGR (Generic Names Supporting Organization (GNSO) and country code Names Supporting Organization (ccNSO)).

The SG makes a total of eleven recommendations categorized into three groups:

- **For GNSO and ccNSO:** Recommendations 1 to 6 seek to assist each SO to incorporate the RZ-LGR into their policies in a consistent manner, for existing and future TLDs.
- **For ICANN Organization and PTI/IANA:** Recommendations 7 to 10 deal with making the normative and non-normative information, part of proposals for RZ-LGR produced by GPs and included in the integrated RZ-LGR by IP, available for other stakeholders to use.
- **For GPs and IP:** Recommendation 11 is addressed to GPs and IP for maintaining the stability of the Root Zone.

Considerations beyond the use of the RZ-LGR such as purpose of TLD (e.g., brand, ccTLD, community, etc.), allowing single-character TLDs (as discussed in Security and Stability Advisory Committee's (SSAC's) SAC052), or reducing allocatable variant labels (as suggested in SSAC's SAC060) are considered policy matters, which are beyond the scope of this SG.

BACKGROUND

Internationalized Domain Name (IDN) Top Level Domains (TLDs) have been a priority for the ICANN Board for several years, based on the input from the community. The variant labels for IDN TLDs have also been an important component for some script communities. Therefore, in 2010, the ICANN Board asked the ICANN org and the community to investigate feasible approaches for evaluation, possible delegation, allocation and operation of variant TLDs.

In 2012, the community undertook various case studies for six scripts. The examination collected in the Integrated Issues Report [IIR] identified two items to be addressed:

1. there is no universally acceptable definition of what may constitute a variant relationship between IDN TLD labels
2. there is no variant management mechanism defined

Following this report, the ICANN org and the community developed the Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA Labels [LGR Procedure]. This procedure allows to define label generation rules for different scripts to determine valid TLD labels and their variant labels. In 2013, the ICANN Board endorsed this procedure and asked the ICANN org and community to undertake it.

To date GPs have completed and submitted their RZ-LGR proposals for Arabic, Armenian, Cyrillic, Devanagari, Ethiopic, Georgian, Gujarati, Gurmukhi, Hebrew, Kannada, Khmer, Korean, Lao, Malayalam, Oriya, Sinhala, Tamil, Telugu and Thai scripts. From these, Arabic, Ethiopic, Georgian, Khmer, Lao and Thai scripts have already been integrated into the second version of the RZ-LGR [RZ-LGR-2] by the Integration Panel (IP), and additional 10 scripts are being integrated in the third version of RZ-LGR, which has been released for public comment. Many of the remaining script communities are now in the process of finalizing their LGR proposal development work. A detailed rationale for using RZ-LGR was published as part of the public comment on the variant TLD management mechanisms. After incorporating the feedback from the community, the ICANN Board approved the documents and requested ccNSO and GNSO to take these into account in their policy development process.

With the availability of the RZ-LGR, the ICANN Board asked the ICANN community (including SOs/ACs and IAB) to study and recommend how to technically apply the RZ-LGR in a harmonized way to TLDs in order to define variant labels for IDN TLDs, the first item noted in the Integrated Issues Report [IIR], as a prerequisite to determine possible variant TLD management mechanism. The RZ-LGR Study Group (SG) has been formed from the nominees of SOs, ACs, IAB and additional volunteers from the ICANN community following the call for formation in February 2018 to address the request from the ICANN Board.

PRINCIPLES

The RZ-LGR SG agreed to the following principles while undertaking the technical study on utilizing the RZ-LGR:

1. **Recommendations must not go against the security and stability of the root zone.**
2. **Recommendations must follow the LGR Procedure**, because the Procedure has been developed by the community and adopted by the ICANN Board, and is the basis of the community work by the different script-based GPs to develop RZ-LGR.
3. **Recommendations must not challenge the contents of RZ-LGR,** because the SG focuses on technical use of RZ-LGR, whereas the responsibility for its contents lies with the GPs and IP as per the LGR Procedure.
4. **Given multiple options, the recommendations should take the conservative approach,** as this would be the first time RZ-LGR is being used to validate labels and determine their variant labels.

SCOPE OF ANALYSIS

The RZ-LGR, at its core, is a resource that has been developed by the community to offer consistent and predictable validation of strings intended for TLDs, and calculation of their variant labels. Therefore, the SG deliberated a range of issues from the viewpoint of those producing RZ-LGR (GPs and IP), using RZ-LGR (e.g. TLD registries and future TLD applicants) and developing policy to use RZ-LGR (GNSO and ccNSO). Considerations beyond the use of the RZ-LGR such as purpose of TLD (e.g., brand, ccTLD, community, etc.), allowing single-character TLDs (as discussed in SSAC's SAC052), reducing allocatable variant labels (as suggested in SSAC's SAC060), or determining the process to resolve issues derived from RZ-LGR (as suggested in SSAC's SAC060), are beyond the scope of this SG. In some of these cases, where relevant, some technical considerations are presented by the SG for consideration in the policy development process by the appropriate supporting organizations.

Users of RZ-LGR

The intended users of the RZ-LGR are likely to be:

1. TLD or TLD Variant Label applicants (primary focus).
2. Others associated with TLD evaluation and delegation, including ICANN org, PTI/IANA, application evaluation panels
3. Others, including:
    a. Those involved in the application process, e.g. trademark holders, to engage in the objection process during application evaluation
    b. Technical solution providers, e.g. for universal acceptance of domain names, such as browser developers to calculate variant labels

What does “Technical Utilization” of RZ-LGR mean?

Technical utilization of RZ-LGR constitutes the following functions:

1. **Label syntax validation** - To determine whether an applied-for string is valid or invalid based on the code point repertoire and Whole-Label Evaluation (WLE) rules of the RZ-LGR.

2. **Variant label calculation** - To calculate the variant labels of an input string, and corresponding label disposition values (allocatable or blocked), based on the variant code points and their types defined in the RZ-LGR.

RECOMMENDATIONS

Recommendations for GNSO and ccNSO

1. All top-level domain (TLD) labels, IDN and ASCII labels[1], must be processed using the RZ-LGR**.** Lowercase alphabetic ASCII labels are, as a practical matter, a subset of the Latin script labels defined by RZ-LGR; therefore, ASCII Labels must be subject to RZ-LGR processing to determine their cross-script variant labels, e.g. with Armenian, Cyrillic, Greek, and other applicable scripts. Consequently, GNSO and ccNSO should incorporate the use of RZ-LGR into their TLD application processes accordingly and in a consistent manner.

2. For the scripts and writing systems[2] which have been integrated into the RZ-LGR, the RZ-LGR must be the only source for the following cases:

    2.1. Validate an applied-for TLD label and determine its variant labels with corresponding dispositions;

    2.2. Calculate variant labels, and corresponding disposition values, for each one of the already allocated or delegated TLD labels; and

    2.3. Calculate variant labels, and corresponding disposition values, for each one of the reserved TLD labels.

3. GNSO and ccNSO should work collaboratively and consider the policy, procedure and/or contract changes to address any existing deviations from the calculation of the RZ-LGR in two specific areas:

[1] The terms “IDN Label” and “ASCII Label”, here and thereafter, have the same meaning as in the LGR Procedure. More precisely, per RFC 5890, an “IDN Label” refers to an A-Label or U-Label, and “ASCII Label” refers to Non-Reserved Letter Digit Hyphen (NR-LDH) labels.
[2] A writing system refers to a system like Japanese, which uses different Unicode scripts (i.e. Han, Hiragana, Katakana and a subset of Latin letters)

3.1. **Delegated TLDs:** These are cases that have occurred under special circumstances in which labels generally deemed as variant TLDs now as per the RZ-LGR calculations, were previously delegated as independent TLDs, albeit with special considerations (e.g. synchronized TLDs). Any such variations should be considered for alignment with RZ-LGR.

3.2. **Self-identified variant TLDs:** Historically IDN TLD applications, for gTLDs and ccTLDs, have asked the applicant to identify and list any variant labels (based on their own calculations) corresponding to the applied-for string. These self-identified variant labels may or may not conform to the RZ-LGR once implemented. Those conforming self-identified variant labels will need to be assigned a variant disposition based on RZ-LGR calculation. Further, it would need to be determined how to address the self-identified "variant" labels that do not conform with the RZ-LGR definition.

GNSO and ccNSO must consider a resolution of such outstanding cases that conforms to the LGR Procedure and RZ-LGR calculations.

Recommendation 4 below describes the cases by which an applied-for label, whose script is supported by the RZ-LGR, is determined to be "invalid". The SG defers to the GNSO and ccNSO to determine the process to deal with these cases (e.g. suspend or reject the applied-for TLD) as this is considered a matter of policy or procedure. While there may be merits for either choice, the SG provides items 4.1 to 4.4 as technical input for community's consideration, to help address SSAC's SAC 060 recommendation: "ICANN must maintain a secure, stable, and objective process to resolve cases in which some members of the community (e.g., an applicant for a TLD) do not agree with the result of the LGR calculations."

4. For an applied-for TLD label whose script(s) are supported by the most current version of the RZ-LGR, the RZ-LGR will calculate either of two values: "valid" or "invalid". Consequently, an applied-for TLD that is determined "valid" may proceed with the relevant evaluation process, whereas an applied-for TLD that is determined "invalid" must not proceed, because it did not pass the validation by RZ-LGR. While policy needs to determine how an "invalid" label is dealt with (Recommendation 2, SAC060), the following technical input should be considered by the relevant policy:

4.1. **Conformance with IDNA2008.** An applied-for label containing any DISALLOWED or UNASSIGNED code point(s) per IDNA 2008, or its successors, must not proceed.

4.2. **Conformance with LGR Procedure.** Policy or procedure must not override the results of the RZ-LGR. That is, policy or procedure alone cannot turn an "invalid" label to a "valid" label, or vice-versa. Doing so would invalidate the entire RZ-

LGR. Any change to the RZ-LGR (e.g. repertoire, variant rules or WLEs) must be undertaken using the process stipulated in the LGR Procedure.

4.3. **Excluded code point from MSR.** An applied-for label containing one or more code points that are allowed per IDNA2008 (e.g. PVALID or Protocol Valid) but are not included in the latest version of the Maximal Starting Repertoire (MSR) may proceed with the evaluation process only if: i) the code points are included in an updated version of the MSR in conformance with the LGR Procedure, and ii) the applied-for label is validated by the updated RZ-LGR.[3]

4.4. **Excluded code point from RZ-LGR.** If the applied-for label contains one or more code points that are not in the RZ-LGR but included in the applicable MSR for the relevant script, the applied-for label may proceed only if: i) the code points are included in the revised script proposal in conformance with the LGR Procedure, and ii) the applied-for label is validated by the updated RZ-LGR. [4]

5. For an applied-for TLD label whose script is not yet supported by the most current version of the RZ-LGR, the label should not proceed through evaluation until the relevant script is integrated in the RZ-LGR. This approach will ensure that any applied-for TLD label and its applicable variants, if any, are algorithmically validated using the RZ-LGR and are unique from existing or other applied-for TLD labels, or their applicable variants.

   At this point, where the script of an applied-for label is not supported by the RZ-LGR, the SG contemplated three scenarios and its implications that should be considered by the relevant policy to resolve the situation for such TLD applicants. The SG is seeking feedback from the community on the options provided to finalize the recommendation.

   Option A - Hold the application until relevant script is integrated into the RZ-LGR (recommended)

   This path would allow the applicant to resolve the situation of its applied-for TLD through a process that conforms with the LGR Procedure (i.e. Principle 2 above), such that the applied-for TLD may be allowed by a new version of the RZ-LGR.

[3] The LGR Procedure dictates that updates or changes to the RZ-LGR, with respect to a script, are done in a two-step process; first, a Generation Panel develops or updates a proposal; second, the Integration Panel reviews and integrates the proposal into a new version of the RZ-LGR. The review process by the Integration Panel may result in accepting or rejecting the Generation Panel's proposal.

[4] Certain code points that were considered in the MSR but explicitly excluded from the corresponding script's proposal may never be included in the RZ-LGR, e.g. for security reasons. Therefore, applicants should carefully review the reasons provided by the Generation Panel in their proposal before considering to apply for a TLD that contains a code point that was explicitly excluded from RZ-LGR.

The practical limitation, however, is that an LGR script proposal required to evaluate the applied-for TLD may take months or even years[5] to be ready for use. Thus, the question is how long can an application be held.

Option B - Use a manual inspection that conforms with the principles of the LGR Procedure, to review the applied-for TLD:

The SG defers, to a policy development process, the issue of determining whether it is advisable to delegate an applied-for label in a script that is not supported by the RZ-LGR using a manual inspection[6]. If Option B is considered by policy as the way to proceed, then, the policy working group should consider the following additional technical considerations for a secure, stable and consistent approach:

- A manual string evaluation procedure must be consistent with the principles used for developing RZ-LGR, using the LGR Procedure and RFC 6912 (Principles for Unicode Code Point Inclusion in Labels in the DNS).
- The evaluation panel must include or co-opt expert(s) of relevant script or writing system (e.g. linguists, Unicode experts, etc.) as well as member(s) of the Integration Panel to help evaluate the string.
- The manual string evaluation procedure must be just that and nothing more; the procedure must only be used for string evaluation and must not seek to define script behavior for the RZ-LGR. The latter is the work for the GP for the script, once it is formed.
- The string evaluation must be conservative. This means, the string evaluation should use a narrow repertoire which must be derived from the latest version of the MSR. The applied-for label's intended language status on Expanded Graded Intergenerational Disruption Scale (EGIDS)[7] should be 0 through 4 only, and each one of the code points used in the applied-for label must be unequivocally a core letter used for writing the language identified[8].
- Code point variants should also be tentatively identified using liberal criteria that encompass possible visual, phonetic and semantic aspects relevant for the script (including a reasonably liberal visual similarity criteria). Following their identification, a conservative variant analysis approach should be applied (for a

[5] Generation Panels have taken as few as three months to complete their RZ-LGR proposals and as many as more than five years.
[6] Currently this is one of the functions of a DNS Stability Review Panel.
[7] See https://www.ethnologue.com/about/language-status.
[8] These are some of the criteria used by GPs to develop their respective script-based proposals.

script not supported by the RZ-LGR) which must tentatively block all potential variant labels created based on such liberal analysis until the LGR proposal for the script by the relevant GP is integrated into the RZ-LGR which can confirm that the identified labels are indeed variant labels. At that time, only the variant labels identified by RZ-LGR can proceed to further steps based on their dispositions identified by RZ-LGR. Any additional labels initially identified as possible variants can then be released new applications.

- The applied-for TLD label and possible variant labels identified must follow conventional constraints of the script as would be identified by the WLE rules by the relevant GP. The script experts in the panel should review the applied-for string in the context of such conventions.
- Use of combining marks in the applied-for TLD label must follow conventional rules and must not be used arbitrarily. Combining marks optionally required in a script, or only used for special purpose, must not be used in a TLD label. It must be the responsibility of the applicant to provide sufficient evidence for the panel to confirm that the use of combining marks is appropriate per the criteria discussed.
- The applied-for label must not mix scripts. Each one of the code points of the applied-for label must have the same Unicode script value.

Option C - Reject the application with no appeal (least recommended)

The SG believes this is the most extreme alternative and given that there is a process to modify the RZ-LGR, the process should be utilized (e.g. Option A), and rejecting an application without appeal is contrary to SSAC's recommendation in SAC 060.

6. 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 labels to be a policy matter. Refer to Appendix C of IDN Variant TLD Implementation: Appendices [Variant Management] for some suggested approaches.

Recommendations for ICANN Organization and PTI/IANA

7. The RZ-LGR is the result of the integration process of individual proposals each based on a script or a writing system. These proposals contain normative and non-normative information, but only the normative information (e.g., code point repertoire, variant sets and whole label evaluation rules) are integrated into the RZ-LGR. The non-normative sections of the individual proposal may be used to inform other procedures, such as

string similarity review. To this end, ICANN organization should develop and maintain a centralized repository[9] for the individual proposals by the GPs, so that this non-normative information is also easily accessible by the ICANN community.

8. The authoritative XML file of the RZ-LGR must be published by PTI/IANA.

9. PTI/IANA should make available an informative and non-normative list of variant labels corresponding to every delegated TLD in the root zone; variant labels must be published with their corresponding disposition value.

10. ICANN org should make available an implementation of the RZ-LGR. This implementation of the RZ-LGR should not be regarded as authoritative, but as a tool for community service. Only the XML file of the RZ-LGR published by PTI/IANA is to be considered authoritative or normative.[10]

Recommendation for GPs and the IP

11. The purpose of the RZ-LGR is to offer a predictable and algorithmic method to determine valid TLDs and their variant labels. GPs and the IP should pay close attention to avoid any impact to existing TLDs, that is, GPs and IP should make the RZ-LGR backward compatible with existing TLDs towards maintaining the stability of the Root Zone[11] (also see Recommendation Section 3.1 above discussing “Synchronized TLDs”). In the event that backward compatibility cannot be achieved, the GP must clearly identify and discuss the rationale in its RZ-LGR proposal.

[9] See https://www.icann.org/resources/pages/lgr-proposals-2015-12-01-en for this repository which points to original proposal by GP, public comment feedback for each proposal and the final proposals submitted for evaluation by IP.
[10] See https://lgrtool.icann.org/ as an example of such a tool.
[11] See Recommendation 6 in SAC060 located at https://www.icann.org/en/system/files/files/sac-060-en.pdf

CONTRIBUTORS

1	Edmon Chung	.asia	GNSO
2	Gaurav Vedi	Dominion Registries	GNSO
3	Dusan Stojicevic	Gransy and Cyrillic GP	GNSO
4	Dennis Tan Tanaka (Chair)	Verisign and Latin GP	GNSO
5	Mirjana Tasic	.rs, .срб and Latin GP	ccNSO
6	Wei Wang	KNET and Chinese GP	GNSO
7	Ajay Data	XGENPLUS and Neo-Brahmi GP	GNSO and ccNSO
8	Alireza Saleh	IRNIC and Arabic GP	IAB
9	Dessalegn Yehuala	Addis Ababa Univ. and Ethiopic GP	
10	Harsha Wijayawardhana	Theekshana and Sinhala GP	

ICANN org supporting team:

- Pitinan Kooarmornpatana
- Sarmad Hussain

REFERENCES

[Variant Management]	IDN Variant TLD Management Framework Recommendations, 2019 https://www.icann.org/resources/pages/idn-variant-tld-implementation-2018-07-26-en
[IIR]	Integrated Issues Report, 2012 https://www.icann.org/en/system/files/files/idn-vip-integrated-issues-final-clean-20feb12-en.pdf
[LGR Procedure]	Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA Labels. Version 2013-03-20b https://www.icann.org/en/system/files/files/lgr-procedure-20mar13-en.pdf
[RZ-LGR-V2]	Root Zone Label Generation Ruleset. Version 2 https://www.icann.org/resources/pages/root-zone-lgr-2015-06-21-en
[SAC060]	SSAC Comment on Examining the User Experience Implications of Active Variant TLDs Report https://www.icann.org/en/system/files/files/sac-060-en.pdf

Appendix A: Note on Single Character TLDs

Historically, single character TLDs have not been allowed due to their confusability potential. The SG advises GNSO and ccNSO to review SSAC's SAC 052[12] 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.

[12] See SAC052 located at https://www.icann.org/en/system/files/files/sac-052-en.pdf

Appendix B: Definitions

Allocatable Label means a Valid Label that has been processed by the RZ-LGR with a disposition value “allocatable”; the label may be eligible for allocation or delegation as a top level domain.

A-Label is an IDN Label, per IDNA2008, in its ASCII Compatible Encoding form (e.g. xn--fo-gka)

ASCII Label means an LDH (Letter Digit Hyphen) Label; this definition is consistent with its use in the LGR Procedure. ASCII Label is not to be confused with A-Label.

Blocked Label means a Valid Label that has been processed by the RZ-LGR with a disposition value “blocked”. Consequently, a Blocked Label must not be available for allocation or delegation as a top level domain.

IDN stands for Internationalized Domain Name.

IDN Label means an A-Label or U-Label per IDNA2008; this definition is consistent with its use in the LGR Procedure.

Invalid Label means an input label, or a calculated Variant Label, that does not conform to the RZ-LGR repertoire or the applicable WLE rules.

LGR stands for Label Generation Rules or Ruleset.

U-Label is an IDNA2008 valid string of Unicode characters, in Normalization Form C (NFC) and including at least one non-ASCII character, expressed in a standard Unicode Encoding Form (such as UTF-8; e.g. foö).

Valid Label means an input label, or a calculated Variant Label, which conforms to the RZ-LGR repertoire and the applicable WLE rules.

Variant Label means a label calculated by the RZ-LGR based on the variant code points defined in the repertoire. As per the RZ-LGR Procedure, Variant Labels are deemed the same by the GP.