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

7 October 2019

Table of Contents

Executive Summary

With the availability of the Root Zone Label Generation Rules (RZ-LGR), the ICANN Board asked the ICANN community to investigate any issues in technically employing the RZ-LGR for validating 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.

The SG has deliberated a range of issues from the viewpoint of the different stakeholders and has proposed a total of twelve technically motivated recommendations categorized into three groups:

- **For generic and country code supporting organizations (GNSO and ccNSO):** Recommendations 1 to 7 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 Internet Assigned Numbers Authority (IANA):** Recommendations 8 to 11 deal with publishing and enabling the use of RZ-LGR for the community.
- **For Generation Panels (GPs):** Recommendation 12 is addressed to GPs for maintaining the stability of the Root Zone.

Considerations beyond the use of the RZ-LGR such as purpose of TLD (e.g., brand gTLD, ccTLD, community gTLD, etc.), allowing single-character TLDs (as discussed in Security and Stability Advisory Committee's (SSAC's) SAC052[1]), or reducing allocatable variant labels (as suggested in SSAC's SAC060[2]) are considered policy matters, which are beyond the scope of this SG.

[1] https://www.icann.org/en/system/files/files/sac-052-en.pdf
[2] https://www.icann.org/en/system/files/files/sac-060-en.pdf

1 Background

Supporting Internationalized Domain Name (IDN) Top Level Domains (TLDs) has been an important topic for the ICANN community for several years. To address the need for IDN variant TLDs, in 2010, the ICANN Board requested the ICANN org and the community to investigate feasible approaches for evaluation, possible delegation, allocation and operation of variant TLDs.

Following the direction from the Board, in 2011 the community undertook case studies for six scripts to identify the issues associated with IDN variant TLDs. The resulting analysis, collected in the Integrated Issues Report (IIR)[3] in 2012, identified two significant issues to be addressed:

1. There is no accepted definition for what may constitute a variant relationship between top-level labels, and
2. There is no variant management mechanism.

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)[4]. 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 requested the ICANN org and the community to undertake it.

To date Generation Panels (GPs) have completed their Root Zone Label Generation Rules (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, sixteen scripts have already been integrated into the third version of the RZ-LGR (RZ-LGR-3) by the Integration Panel (IP). Many of the remaining script communities are now in the process of finalizing their LGR proposals. In addition, a detailed rationale for using RZ-LGR was published as part of the public comment on the recommendation for the variant TLD management mechanisms. After incorporating the feedback from the community, the recommendations were presented to the ICANN Board by ICANN org. The ICANN Board approved the recommendations and requested ccNSO and GNSO to take these into account in their policy development processes.

With the availability of the RZ-LGR, the ICANN Board requested the ICANN community (including SOs, ACs and IAB) to study the technical use of the RZ-LGR consistently across all IDN TLDs. Accordingly, the RZ-LGR Study Group (SG) was formed from the nominees of SOs, ACs, IAB and additional volunteers from the ICANN community following the call for formation circulated in February 2018, to address the request from the ICANN Board.

---

[3] https://www.icann.org/en/system/files/files/idn-vip-integrated-issues-final-clean-20feb12-en.pdf
[4] https://www.icann.org/en/system/files/files/draft-lgr-procedure-20mar13-en.pdf

2 Principles

The RZ-LGR SG agreed to the following principles while undertaking the technical study on technically 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 TLD labels and determine their variant TLD labels.

3 Scope of Analysis

The RZ-LGR, at its core, is a resource that has been developed by the community to:

I.   Offer a consistent and predictable mechanism to validate TLD labels, and
II. Calculate their variant TLD labels, if any.

Therefore, the SG deliberated a range of issues from the viewpoint of those producing RZ-LGR (GP 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 technical use of the RZ-LGR such as purpose of TLD (e.g., brand gTLD, ccTLD, community gTLD, etc.), allowing single-character TLDs (as discussed in Security and Stability Advisory Committee's (SSAC's) SAC052 report), or reducing allocatable variant labels (as suggested in SSAC's SAC060), are beyond the scope of this SG's work. Where relevant, some technical considerations are presented by the SG for consideration in the policy development processes by the relevant supporting organizations.

4 Users of RZ-LGR

The intended users of the RZ-LGR are likely to be:
1. TLD or variant TLD applicants (primary focus)
2. Stakeholders associated with TLD evaluation and delegation processes, including ICANN org, IANA, application evaluation as well as dispute resolution panels
3. Others, including:
    a. Those with a vested interest 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, search engines to discover allocatable variant TLDs, etc.

5 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 label is valid or invalid based on the code point repertoire, context 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.

6 Recommendations

Recommendations for GNSO and ccNSO

1. All TLD labels, IDN and ASCII labels[5], 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, these 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[6] which have been integrated into the RZ-LGR, the RZ-LGR must be the only source for processing 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

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

---

[5] The terms "IDN Label" and "ASCII Label", here and thereafter, have the same meaning as in the LGR Procedure. More precisely, per RFC 5890, the relevant "IDN label" is an IDNA-valid label in an A-Label or U-Label format, and the relevant "ASCII Label" is a Non-Reserved Letter Digit Hyphen (NR-LDH) label.
[6] A writing system refers to a system like Japanese, which uses different Unicode scripts (i.e. Han, Hiragana, and Katakana)

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

    3.1. **Delegated TLDs:** These are cases that have occurred under special circumstances in which labels generally deemed as the same (i.e. variant TLDs under RZ-LGR) 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. The self-identified "variant" labels which are also variant labels based on RZ-LGR will need to be assigned a variant disposition based on RZ-LGR calculation. Further, self-identified "variant" labels that are not variant labels based on the RZ-LGR definition should not be considered as variant TLD labels and it needs to be determined on how to address such labels previously identified by the applicants.

    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 in 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 SAC060 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 applicable 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 subsequent 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 should be dealt with (Recommendation 2 in SAC060), the following technical input should be considered by the relevant policy development process:

4.1. **Conformance with IDNA2008.** An applied-for label must be in Normalization Form C[7] and must conform to IDNA2008.

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 into 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. **Script LGR can be updated, if justified, using the LGR Procedure.** In general, GPs make design choices[8] based on current knowledge and available information. These choices determine the code point repertoire and its context rules, the whole-label evaluation rules and variant sets. If and when there is new information available, the LGR Procedure defines the process to update the RZ-LGR[9].

4.4. **Re-validation of applied-for label is possible**. The applied-for TLD label may be re-validated when a new RZ-LGR version becomes available.

5. For an applied-for TLD label whose script is not yet supported by the applicable version of the RZ-LGR, the application should not proceed until the relevant script is integrated into the RZ-LGR. It is implied that the application should remain on-hold (or other appropriate status) until the relevant script is integrated into the RZ-LGR[10]. This approach will ensure that any applied-for TLD label and its applicable variants, if any, are algorithmically validated using the RZ-LGR.

   It is important to recognize that the RZ-LGR can be updated to include additional scripts as long as it is done in compliance with the LGR Procedure. The practical limitation, however, is that the time to create an LGR script proposal varies greatly (i.e. months or years). To this end, GNSO and ccNSO should develop a policy and implementation steps to manage these cases.

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 variant labels to be a policy matter. Refer to Appendix C of IDN Variant TLD Implementation: Appendices[11] for some suggested approaches.

---

[7] Unicode Normalization Forms, Standard Annex #15: https://unicode.org/reports/tr15/#Norm_Forms
[8] 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.
[9] 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.
[10] See Appendix A for additional context and rationale.
[11] https://www.icann.org/resources/pages/idn-variant-tld-implementation-2018-07-26-en

7. It is expected that the RZ-LGR be revised throughout its lifecycle, either as a result of a new script LGR being integrated or a revision of an existing script LGR being adopted. There may be cases where a script LGR does not support an existing TLD (See Recommendation 12 for additional context). In such cases, it is possible that the existing TLD(s) may need to be grandfathered. GNSO and ccNSO should consider reviewing their policies and procedures to deal with the grandfathering process of existing TLDs as a result of changes of the RZ-LGR.

Recommendations for ICANN Organization and IANA

8. The RZ-LGR is the result of the integration of individual LGR proposals and each of them is developed based on a particular script or a writing system. These LGR proposals contain normative and non-normative information, but only the normative information (e.g., code point repertoire, context rules, 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[12] for the individual proposals by the GPs, so that this non-normative information is also easily accessible by the ICANN community.

9. The XML files of the individual script LGRs and the merged LGR, collectively forming the RZ-LGR, must be published by IANA, and this publication must be considered as the only authoritative source for the RZ-LGR.

10. IANA should make available an informative and non-normative list of variant labels corresponding to every delegated TLD in the root zone based on the applicable version of the RZ-LGR; each variant label must be published along with its corresponding disposition value calculated by the RZ-LGR.

11. ICANN org should make available an implementation of the RZ-LGR[13]. This implementation of the RZ-LGR should not be regarded as authoritative, but as a tool for community service.

Recommendation for GPs and the IP

The purpose of the RZ-LGR is to offer a transparent, predictable and algorithmic method to determine valid TLDs and their variant labels. It is reasonably expected that impacts to existing TLDs should be kept to a minimum. However, there may be a case where a script LGR proposal by a GP may not support an existing delegated TLD label. The LGR Procedure states that "While existing labels will almost certainly have to be grandfathered if they are in conflict with the label generation rules established by this procedure, that precedent and conflict is not a

---

[12] See https://www.icann.org/resources/pages/lgr-proposals-2015-12-01-en for this repository which points to original proposals by GP, public comment feedback for each proposal and the final proposals submitted for evaluation by IP.
[13] See https://lgrtool.icann.org/ as an example of such a tool.

reason to invalidate any aspect of the new rules or this procedure." To this end the following recommendation will be applicable.

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.

7 Contributors

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

ICANN org supporting team:

- Pitinan Kooarmornpatana
- Sarmad Hussain

Appendix A: Options to Treat Unsupported Scripts

Given that the process to incorporate script LGRs into the RZ-LGR is staggered there may be times when a given applied-for label cannot be processed by the RZ-LGR simply because, in practical terms, the script has not been integrated yet. In this event, what are the options?

The SG group analyzed three alternatives, presented as Options A, B and C, respectively below. Each one presents pros and cons. However, Option A is what the SG recommends because is the option that most conforms with the principles of this work.

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 later version of the 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[14] to be ready for use.

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 evaluate an applied-for label in a script that is not supported by the RZ-LGR using a manual inspection[15]. 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

---

[14] Generation Panels have taken as few as three months to complete their RZ-LGR proposals and as many as more than five years.
[15] Currently this is one of the functions of a DNS Stability Review Panel.

Intergenerational Disruption Scale (EGIDS)[16] 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[17].

- Code point variants should also be tentatively identified using a liberal criteria that emcompasses 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 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 for 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.

It should be noted that even if Option B is recommended through a Policy Development Process, while the evaluation of the applied-for TLD string may be processed, the RZ-LGR must be updated, in compliance with the LGR Procedure, prior to the delegation of the applied-for TLD (and corresponding variant labels), and preferably prior to the allocation of the TLD.

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 SAC060.

---

[16] See https://www.ethnologue.com/about/language-status.

[17] These are some of the criteria used by GPs to develop their respective script-based proposals.

Appendix B: 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 SAC052[18] 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.

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

Appendix C: Definitions

Allocatable Variant Label (for the top-level) 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. See RFC 8228 for more details.

A-Label is the ASCII-Compatible Encoding form of an IDNA-Valid Label (as per RFC5890).

ASCII Label means a label formed with US ASCII encoding. Relevant subsets for domain labels include the A-labels and Non-Reserved Letter-Digit-Hyphen (NR-LDH) labels (see Figure 1 and details in RFC 5890).

Blocked Variant Label (for the top-level) means a Valid Label that has been processed by the RZ-LGR with a disposition value "blocked"; the label must not be available for allocation or delegation as a top-level domain. See RFC 8228 for more details.

IDNA-Valid Label is an IDN label which meets all of the requirements of IDNA2008 (or its successors) (as noted in RFC 5890).

Invalid TLD Label means an input label, or a calculated variant label, that does not conform to the RZ-LGR.

U-Label is an IDNA-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).

Valid TLD Label is an IDNA-Valid Label which conforms to the RZ-LGR.

Variant TLD Label is a Valid TLD Label which is a variant of a label, as calculated by using the RZ-LGR.