

Root Zone Label Generation Rules (RZ-LGR)

Motivation, Design, Usage and Status

Sarmad Hussain
Sr. Director IDN and UA Program
GNSO's IDN EPDP WG Meeting
7 October 2021



Overview

- ⦿ History of RZ-LGR
- ⦿ RZ-LGR Design and Development Process
 - Design Principles
 - Generation Panel Formation
 - Capturing Linguistic Diversity
 - Process to Develop RZ-LGR
- ⦿ Summary of Work on RZ-LGR
 - Summary of Generation Panel Work
 - Timeline for RZ-LGR Development
 - Status of RZ-LGR
- ⦿ How Does RZ-LGR Work?
 - How will RZ-LGR Be Integrated into the Process?
 - IDN Variant Label (IDL) Set States and Changes
 - Managing the Number of Variant TLDs Delegated
- ⦿ Possible Triggers and the Process for Updating the RZ-LGR
- ⦿ Ongoing Work for RZ-LGR

History of RZ-LGR

A Brief History of Root Zone Label Generation Rules (RZ-LGR)

- ICANN community has identified the need for variant top-level domains (TLDs): technically distinct strings considered the “same” by the community.
 - The definition of “same” depends on the script.
- Initial work was done to identify variant issues in different scripts by the respective script communities in 2010-11, organized by ICANN org as part of Variant Issues Project.
- The community work, incorporated into the [Integrated Issues Report](#) in 2012, identified the need for defining variant TLDs, as a prerequisite to work on managing variant TLDs.
 - “ICANN must have a way to validate potential IDN variant TLD labels when submitted, and to validate all IDN TLDs requested for variant labels and variant conflicts.”
 - “By the same token, because the root is a single, shared zone, it is necessary to adopt a single, internally consistent set of label generation rules that governs the operation of this single zone.”

A Brief History of Root Zone Label Generation Rules (RZ-LGR)

- The community developed RZ-LGR as the single mechanism for defining variant TLDs and specified the [LGR Procedure](#) to formulate RZ-LGR.
- In 2013, the LGR Procedure was [approved](#) by the ICANN Board for implementation for use with gTLDs and IDN ccTLDs.
- Security and Stability Advisory Committee (SSAC) also advised in [SAC060](#):
 - Recommendation 1: The root zone must use one and only one set of Label Generation Rules (LGR).

RZ-LGR Design and Development Process

RZ-LGR Design Principles

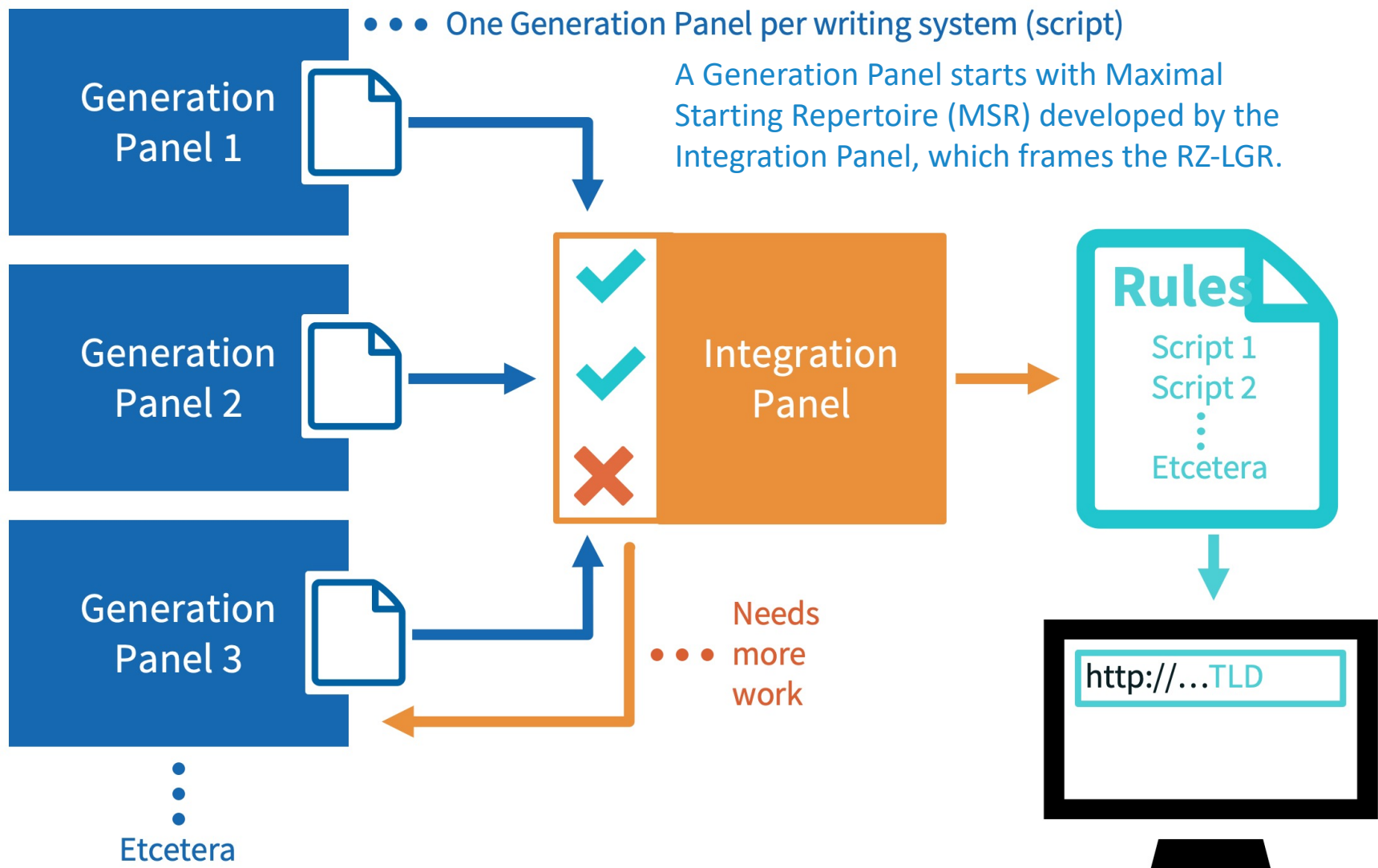
- **Longevity Principle:** A Code Point in the Zone Repertoire should have stable properties across multiple versions of Unicode.
- **Least Astonishment Principle:** A Code Point in the Zone Repertoire should not present recognition difficulties to the zone's intended user population and should not lend itself to malicious use.
- **Contextual Safety Principle:** A code point in the Zone Repertoire or any of its Variants that present unacceptable risks of being used in malicious ways should not be permitted.
- **Inclusion Principle:** The zone repertoire is built up by specific inclusion; the default status for any code point is that it is excluded.
- **Simplicity Principle:** Overly complex rules are to be avoided, in favor of rules easily understood by users with only some background.

RZ-LGR Design Principles

- **Predictability Principle:** People with reasonable knowledge of the topic should by and large reach the same conclusions about which code points should be included.
- **Stability Principle:** Once a code point is permitted, it is almost impossible to stop permitting it: the act of permitting a code point cannot be undone. This is particularly true once a label containing this code point has been registered.
- **Letter Principle:** Only Assigned Code Points normally used to write words should be permitted.
- **Conservatism Principle:** Any doubt should be resolved in favor of exclusion of a code point rather than inclusion.

The principles are generally applicable to the RZ-LGR, including all of its components: code points, variant code points and rules.

RZ-LGR Proposal Development Process



Integration Panel Composition

- As per the LGR Procedure, the Integration Panel (IP) should consist entirely of experts selected by ICANN on the basis of established expertise.
- Duly qualified against ethical conflicts, and in a contractual relation with ICANN, for impartial and unbiased evaluation.
 - At least one expert in Unicode issues.
 - At least one expert in IDNA and DNS issues (or one for each).
 - At least one expert in linguistics and writing systems (who could be the same as the first expert but need not and often will not be).
- It is worth emphasizing that the supply of actual general experts in any of the relevant areas of expertise is extremely limited. This represents a potential risk to the procedure.
- [The Integration Panel](#) has five members:
 - Two Unicode experts.
 - Two IDNA and DNS experts.
 - One linguistics and writing systems expert.

Generation Panel Composition

- The work on Generation Panels (GPs) started by an open [Call for Generation Panels to Develop Root Zone Label Generation Rules](#) in 2013, explaining the [purpose and requirements](#).
 - **Expertise:** “All generation panels should have significant expertise in the writing system(s) concerned, but need have neither overall expertise in all of Unicode, nor expertise in any other writing system.”
 - **Diversity:** “Generation panels need to have some diversity of participation in order to be useful. They must have sufficient numbers of participants [and] should be diverse in economic interest.... the work of these panels is concerned with technical issues and involves linguistic expertise, and is not “representative” in nature.”
 - **Composition:** Chair (1), Community Representatives (2+), Linguistic Expert (1-2), Registry/Registrar (1), DNS/IDN/Unicode Expert (1). Actual members vary from 7 to 60+.

Generation Panel Work Organization

- The GPs have expertise in the script, but not required to cover all the languages supported by the script.
- The GPs have been open, and take on many more more members than stipulated, if available. ICANN org set no upper limit on the members. The GPs aim to be as diverse as possible.
- For the languages supported, additional expertise sought through outreach to the relevant community, online research, and consultations with invited experts as needed.
- Most GPs had regular online calls for members to be able to conveniently attend, supported by ICANN org.
- ICANN org supported open mailing lists and wiki pages, if requested by the GPs, to allow members to be able to follow the work.
- Where the community is less geographically distributed, face to face meetings and outreach events also arranged by GPs.
- The GPs reached out to additional experts and community through their internal outreach, presentations at local forums.
- ICANN org organized regular outreach at ICANN meetings and formal public comment for the GPs.

Capturing the Linguistic Diversity Using EGIDS Value

Level	Label	Description
0	International	The language is widely used between nations in trade, knowledge exchange, and international policy.
1	National	The language is used in education, work, mass media, and government at the national level.
2	Provincial	The language is used in education, work, mass media, and government within major administrative subdivisions of a nation.
3	Wider Communication	The language is used in work and mass media without official status to transcend language differences across a region.
4	Educational	The language is in vigorous use, with standardization and literature being sustained through a widespread system of institutionally supported education.
5	Developing	The language is in vigorous use, with literature in a standardized form being used by some though this is not yet widespread or sustainable.
6a	Vigorous	The language is used for face-to-face communication by all generations and the situation is sustainable.
6b	Threatened	The language is used for face-to-face communication within all generations, but it is losing users.
7	Shifting	The child-bearing generation can use the language among themselves, but it is not being transmitted to children.
...		

Capturing the Linguistic Diversity

- Based on the design principles, the Integration Panel (IP) suggested GPs to consider at least all the languages with Expanded Graded Intergenerational Disruption Scale ([EGIDS](#)) value of 4 or lower to ensure linguistic diversity is captured.
- Languages with EGIDS value of 5 (Developing) may be considered if the GP finds sufficient evidence of general-purpose use of the language.
- Each GP has been requested to document the relevant languages in their proposals and is reviewed by the IP to make sure linguistic diversity is maintained.

Process to Create and Integrate a Script Proposal in RZ-LGR

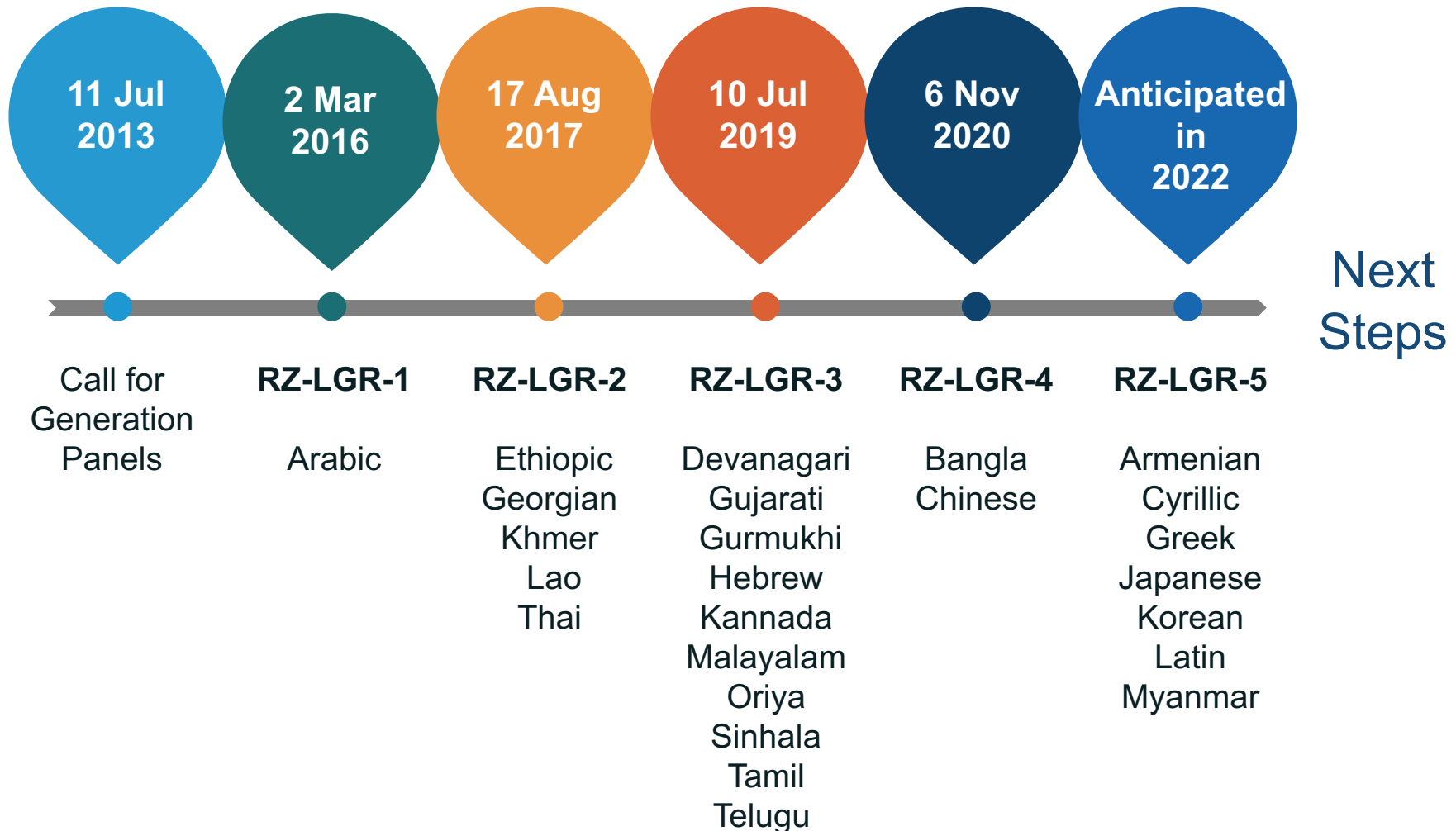
- A script community forms a GP.
- The GP explores the languages that should be supported and their needs in terms of code points, variants or contextual constraints.
- The GP develops a proposal for the RZ-LGR, including the formal XML definition and the explanatory supporting document(s).
- Generally, the GP has multiple consultations with IP while developing the proposal.
- The GP publishes the script proposal for public comment.
- The GP finalizes the proposal and submits to the IP.
- The IP reviews the proposal based on the RZ-LGR principles.
- If the IP accepts the proposal, it is integrated into the next version of the RZ-LGR, else the proposal is returned to the GP.
- The IP publishes the version or RZ-LGR for public comment.
- The IP finalizes and publishes the next version of RZ-LGR.

Summary of Work on RZ-LGR

Summary of Generation Panel Work

Script	Start	End	Days	2014	2015	2016	2017	2018	2019	2020	2021
Arabic	14-Feb-14	18-Nov-15	642								
Armenian	3-Feb-15	5-Nov-15	275								
Bangla	26-May-15	20-May-20	1821								
Chinese	24-Sep-14	26-May-20	2071								
Cyrillic	10-Dec-15	3-Apr-18	845								
Devanagari	26-May-15	22-Apr-19	1427								
Ethiopic	22-Dec-15	17-May-17	512								
Georgian	17-Jun-16	24-Nov-16	160								
Greek	31-Oct-16	15-Jul-21	1718								
Gujarati	26-May-15	6-Mar-19	1380								
Gurmukhi	26-May-15	22-Apr-19	1427								
Hebrew	15-Oct-18	24-Apr-19	191								
Japanese	17-Mar-15	30-Sep-21	2389								
Kannada	26-May-15	6-Mar-19	1380								
Khmer	17-Jun-15	15-Aug-16	425								
Korean	1-Feb-16	1-May-21	1916								
Lao	15-Sep-15	31-Jan-17	504								
Latin	15-May-17	23-Sep-21	1592								
Malayalam	26-May-15	26-Jun-20	1858								
Myanmar	28-Jun-18	ongoing	-								
Oriya	26-May-15	6-Mar-19	1380								
Sinhala	3-Jan-18	22-Apr-19	474								
Tamil	26-May-15	6-Mar-19	1380								
Telugu	26-May-15	7-Jun-19	1473								
Thaana	TBD										
Thai	6-Oct-15	25-May-17	597								
Tibetan	TBD										

Timeline for RZ-LGR Development



Status of Scripts

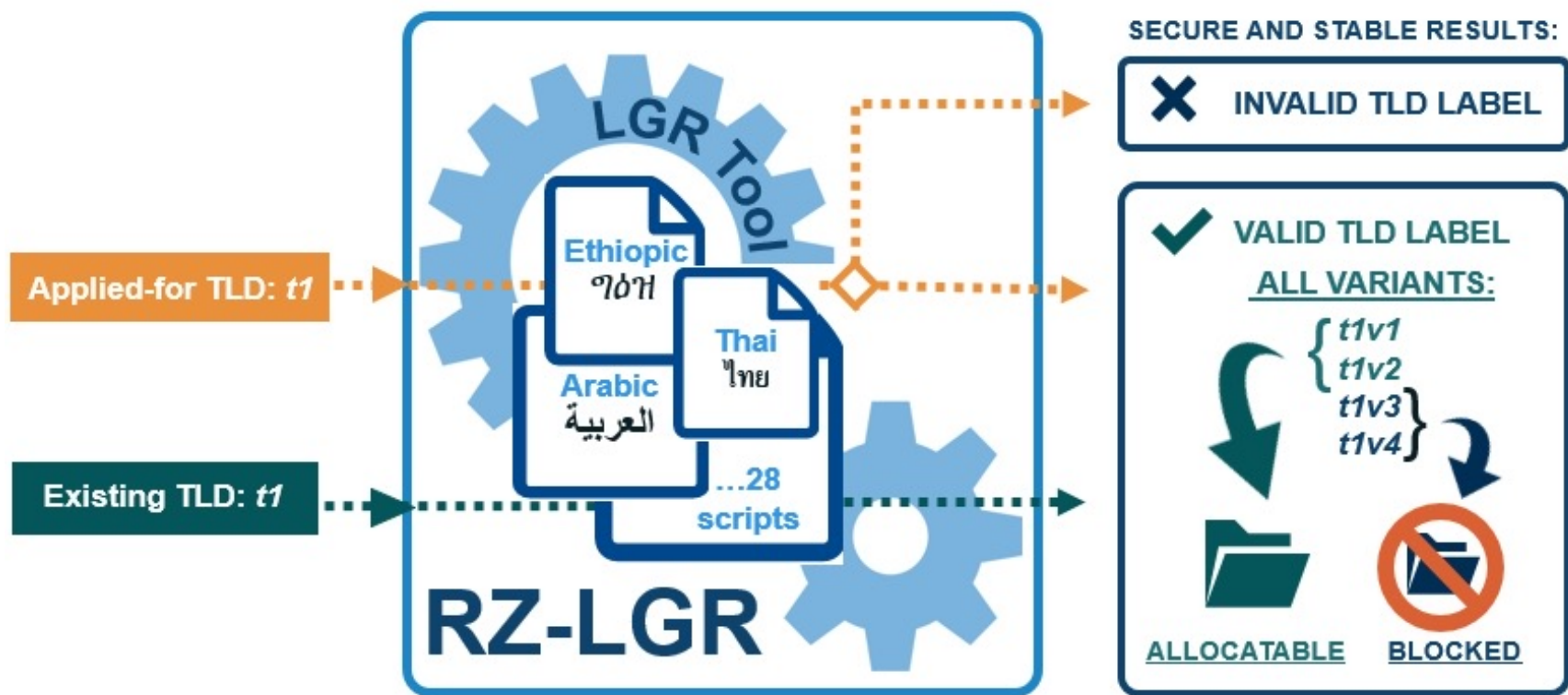
- Integrated:
 - Arabic, Bangla, Chinese, Devanagari, Ethiopic, Georgian, Gujarati, Gurmukhi, Hebrew, Kannada, Khmer, Lao, Malayalam, Oriya, Sinhala, Tamil, Telugu, and Thai scripts.
- Completed, awaiting integration:
 - Armenian, Cyrillic, Greek (waiting for Latin) and Korean scripts.
- In public comment:
 - Japanese and Latin scripts.
- Finalizing for public comment:
 - Myanmar script.
- Not started:
 - Thaana and Tibetan scripts.

Status on Use of RZ-LGR

- In 2019, ICANN org published a set of [Recommendations for Managing the IDN Variant TLDs](#), based on RZ-LGR, including a report explaining why RZ-LGR is needed:
 - [IDN Variant TLD Implementation – Rationale for RZ-LGR](#)
- Following the publication, in 2019 ICANN Board [resolved](#) that GNSO and ccNSO take into account the Recommendations for Managing the IDN Variant TLDs, which integrated the use of RZ-LGR, in their policy development processes.
- In 2020, the ICANN Board [resolved](#) that GNSO and ccNSO take into account the [Recommendations for the Technical Utilization of the Root Zone Label Generation](#) in their policy development processes.
- In 2021, GNSO published its [Report on New gTLD Subsequent Procedures](#) which incorporates the use of RZ-LGR for the next round of new gTLDs.

How Does RZ-LGR Work?

How Does the RZ-LGR Work?

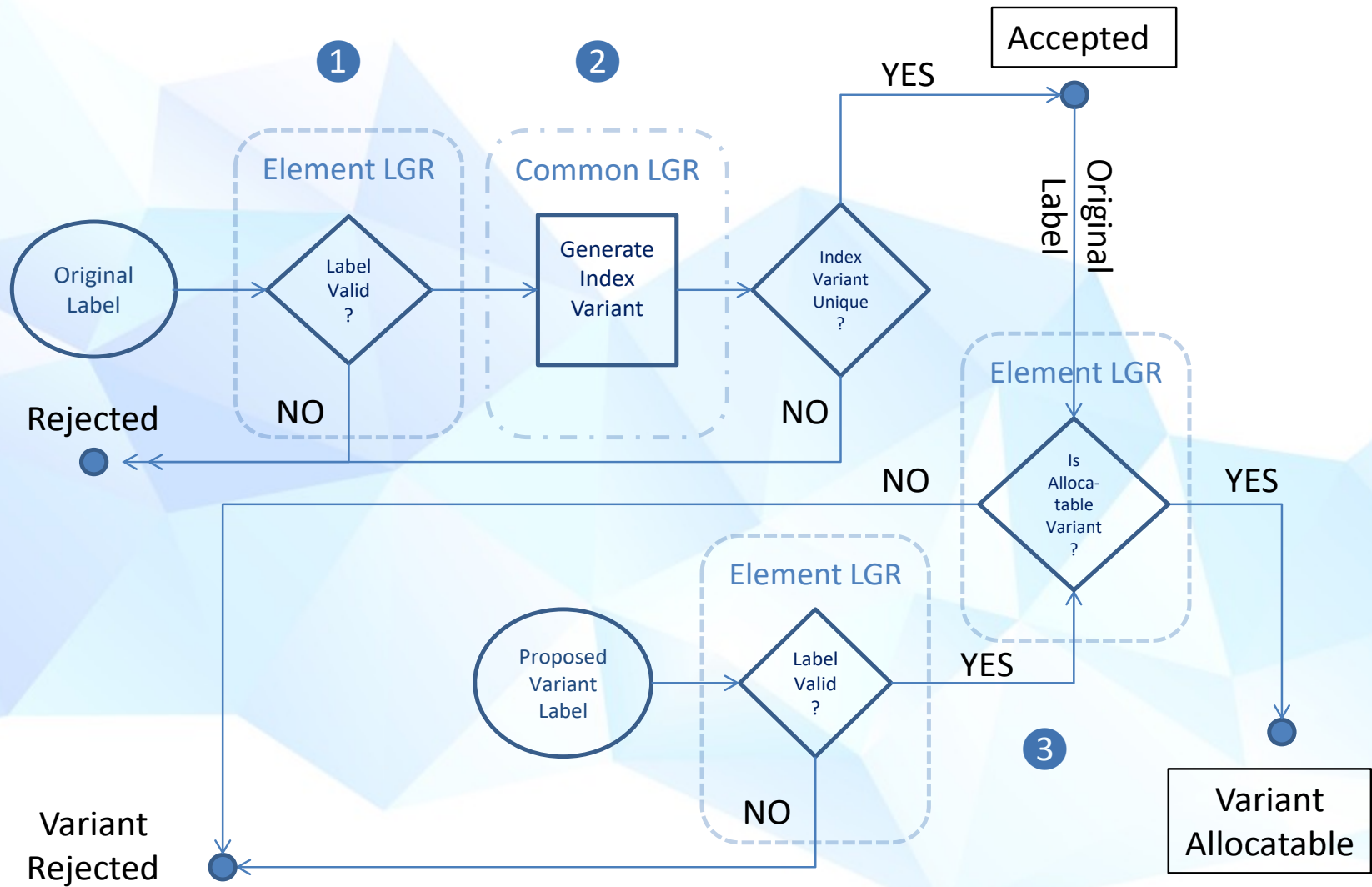


How Are (In)Valid and Variant Labels Calculated?

- The RZ-LGR contains a list of code points, lists of variant code points and a list of rules which can apply on code points or code point sequences.
- For a TLD label, all the code points forming the label must be in the list of code points.
- None of the code points or code point sequences should violate the applicable rules on them.
- If the label meets the above conditions, it is valid, else invalid.

- For any label, the variant rules are applied to determine if it is part of the variant set.
- Any variant label is also evaluated to determine if it is valid.
- For a valid variant label, its disposition is also determined:
 - A variant label may be allocatable, i.e., is a candidate for possible allocation and delegation?
 - A variant label may be blocked, i.e., is not a candidate for possible allocation and delegation?

Schematic for Processing a Label



A Real Example Worked Out with RZ-LGR for [Arabic](#) ([HTML](#))

#	Type	U-label	A-label	Disposition	Code point sequence
1	original	شبكة	xn--ngbc5azd	valid	U+0634 U+0628 U+0643 U+0629
2	varlabel	شبكة	xn--ngbx0cq	allocatable	U+0634 U+0628 U+0643 U+0647
3	varlabel	شبكة	xn--ngbx0c15a	blocked	U+0634 U+0628 U+0643 U+06BE
4	varlabel	شبكة	xn--ngbx0c95a	blocked	U+0634 U+0628 U+0643 U+06C0
5	varlabel	شبكة	xn--ngbx0cy6a	blocked	U+0634 U+0628 U+0643 U+06C1
6	varlabel	شبكة	xn--ngbx0c26a	blocked	U+0634 U+0628 U+0643 U+06C2
7	varlabel	شبكة	xn--ngbx0c66a	allocatable	U+0634 U+0628 U+0643 U+06C3
8	varlabel	شبكة	xn--ngbx0c31b	blocked	U+0634 U+0628 U+0643 U+06D5
9	varlabel	شبكة	xn--ngbc5az1b	allocatable	U+0634 U+0628 U+06A9 U+0629
10	varlabel	شبكة	xn--ngbx2d5u	allocatable	U+0634 U+0628 U+06A9 U+0647
11	varlabel	شبكة	xn--ngbx66ayc	blocked	U+0634 U+0628 U+06A9 U+06BE
12	varlabel	شبكة	xn--ngbx66a6c	blocked	U+0634 U+0628 U+06A9 U+06C0
13	varlabel	شبكة	xn--ngbx66agd	blocked	U+0634 U+0628 U+06A9 U+06C1
14	varlabel	شبكة	xn--ngbx66akd	blocked	U+0634 U+0628 U+06A9 U+06C2
15	varlabel	شبكة	xn--ngbx66aod	allocatable	U+0634 U+0628 U+06A9 U+06C3
16	varlabel	شبكة	xn--ngbx66a0f	blocked	U+0634 U+0628 U+06A9 U+06D5
17	varlabel	شبكة	xn--ngbc5a31b	allocatable	U+0634 U+0628 U+06AA U+0629
18	varlabel	شبكة	xn--ngbx2d9u	allocatable	U+0634 U+0628 U+06AA U+0647
19	varlabel	شبكة	xn--ngbx96asc	blocked	U+0634 U+0628 U+06AA U+06BE
20	varlabel	شبكة	xn--ngbx96a0c	blocked	U+0634 U+0628 U+06AA U+06C0
21	varlabel	شبكة	xn--ngbx96a4c	blocked	U+0634 U+0628 U+06AA U+06C1
22	varlabel	شبكة	xn--ngbx96a8c	blocked	U+0634 U+0628 U+06AA U+06C2
23	varlabel	شبكة	xn--ngbx96ahd	allocatable	U+0634 U+0628 U+06AA U+06C3
24	varlabel	شبكة	xn--ngbx96arf	blocked	U+0634 U+0628 U+06AA U+06D5

Possible Process for Evaluating a TLD Label

- A gTLD label is submitted to the LGR tool as a gTLD or a variant gTLD. All gTLD applications (ASCII and IDN) follow the same process.
- A tool evaluates the label using RZ-LGR to determine if the label is valid.
- The tool determines if the label is available:
 - The label is not already delegated.
 - The label is not reserved.
 - The label is not a variant TLD of a label already delegated.
 - The label is not a variant label of a label already reserved.
- If the label is available, it informs the applicant.
- TBD – generate a list of allocatable variant gTLD labels, which could be numerous. Blocked labels would not be published as these are much more numerous than allocatable variant labels.

Possible Process for Evaluating a TLD Label

- If a label is applied as a variant gTLD of a delegated gTLD by the applicant, the tool uses modified definition of availability:
 - The label is an allocatable variant TLD of the delegated gTLD.
 - The applicant is the same entity as that for the delegated gTLD.
- If a label is clear through RZ-LGR validation, the application proceeds for the DNS Stability review.
- String similarity review takes into account the similarity to variant TLDs.

IDN Variant Label (IDL) Set States and Changes

Initial state	State may change to	Remarks
<i>Withheld-same-entity</i>	<i>Allocated</i>	Allocation only to the same entity as another label in the IDL set. This change happens if a variant was not initially requested for allocation and later is.
<i>Blocked</i>	<i>Withheld-same-entity</i>	A later RZ-LGR version may broaden the available labels in the IDL set. Such possible labels automatically become Withheld-same-entity.
<i>Allocated</i>	<i>Delegated</i>	Happens when name servers are added. (Not new.)
<i>Delegated</i>	<i>Allocated</i>	If a domain is removed from the DNS, the allocation can remain in place anyway. Rare in the root zone, but not new.
<i>Rejected</i>	<i>Withheld-same-entity</i>	Every Rejected label is automatically Withheld-same-entity as well. If the Rejected status comes off, the label can be handled as any other Withheld-same-entity label.

Managing the Number of Variant TLDs Delegated

- RZ-LGR asks to:
 - Minimize allocatable variant labels – to balance usability and manageability.
 - Maximize blocked variant labels – to address security.
- Also discussed by SSAC in SAC060:
 - Notes that “large number of variant strings presents challenges for the management of variant domains at the registry, the registrar and registrant levels.”
 - Recommendation 14: ICANN should ensure that the number of strings that are activated is as small as possible.”
 - Agrees that: 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.
- Though the IP and the GPs have worked to reduce allocatable variant labels, RZ-LGR could still create numerous allocatable variant labels in some cases, and limiting delegation needs to be managed further by policy.

Possible Triggers and the Process for Updating the RZ-LGR

What Can Trigger an Update to the Existing Script in RZ-LGR?

- 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.
-
- There is no automatic periodic revision of RZ-LGR, but can be triggered at any time by an incoming proposal by a GP to the IP.
 - Generally, any update in the RZ-LGR would have to be backwards compatible for stability reasons. Also Recommendation 6 in SAC060.

Process to Update an Existing Script Proposal in RZ-LGR – 1/2

- A GP, applicant or anyone in the community can request a change at any time.
- The change request, with description and evidence, can go directly to the GP or ICANN org.
- The GP evaluates the change to decide if it would consider it.
- If the GP agrees, the GP conducts an analysis and develops an updated proposal for the RZ-LGR.

- The GP develops a proposal for the RZ-LGR, including the formal XML definition and the explanatory supporting document(s).
- Generally, the GP has multiple consultations with IP while developing the proposal.
- The GP publishes the script proposal for public comment.
- The GP finalizes the proposal and submits to the IP.

- The IP reviews the updated proposal based on the RZ-LGR principles.
 - If the IP accepts the proposal, it is integrated into the next version of the RZ-LGR, else the proposal is returned to the GP.
 - The IP publishes the RZ-LGR version for public comment
 - The IP finalizes and publishes the next version of RZ-LGR.
-
- [RZ-LGR Technical Study Group](#)'s Recommendation 4 states that that Policy or procedure must not override the results of the RZ-LGR and that any changes in RZ-LGR by a process outside the LGR Procedure would invalidate the RZ-LGR and thus the definition of the variant TLD.

Ongoing Work for RZ-LGR

Ongoing Work for RZ-LGR

- Integrate pending scripts and scripts finishing public comment into RZ-LGR-5 in 2022.
- Support community in maintaining and updating the RZ-LGR in the future for the scripts already integrated.
- Keep ongoing outreach efforts to set up Generation Panels for Thaana and Tibetan scripts.
- Support any additional scripts that might qualify for integrating into the RZ-LGR.

Engage with ICANN



Visit us at icann.org

 [@icann](https://twitter.com/icann), [@UASGTech](https://twitter.com/UASGTech)

 facebook.com/icannorg

 youtube.com/icannnews

 flickr.com/icann

 linkedin/company/icann

 soundcloud/icann

 instagram.com/icannorg