

# At-Large Workspace: Proposal for Latin Script Root Zone Label Generation Rules

Public Comment Close	Statement Name	Status	Assigned Working Group	Assignee(s)	Call for Comments Open	Call for Comments Close	Vote Open	Vote Close	Date of Submission	Staff Contact and Email
23 November 2021	<a href="#">Proposal for Latin Script Root Zone Label Generation Rules</a>	ADOPTED 11Y, 1A, 0N	CPWG	<a href="#">Jonathan Zuck</a>	17 November 2021	20 November 2021	19 November 2021	23 November 2021	23 November 2021	<a href="mailto:pitinan.koo@icann.org">pitinan.koo@icann.org</a>

Hide the information below, please click [here](#) >>

## See 13 Oct Latin LGR Issues



Latin LGR Issues (1).docx

## Reference 6 Oct [PPT](#) and [PDF](#)

CategoryTechnical  
RequestersICANN org  
ICANN org Contact(s)[pitinan.koo@icann.org](mailto:pitinan.koo@icann.org)

## What We Need Your Input On

Root Zone Label Generation Rules ([RZ-LGR](#)) are developed to define a conservative mechanism to determine valid Internationalized Domain Name (IDN) top-level domains (TLDs) and their variant labels. The successful development of Label Generation Rules depends on having community-based Generation Panels for each script that will be used in the root zone.

The Latin Script GP has finalized the proposal for Latin Script Root Zone Label Generation Rules (Latin Script RZ-LGR). The proposal is being published for Public Comment before being finalized and submitted to the Integration Panel for integration into the Label Generation Rules for the Root Zone (RZ-LGR).

Please review and provide your input regarding the code point repertoire, variant definitions, and the rules proposed in the proposal.

### Proposals For Your Input

[Latin Script RZ-LGR Appendices](#)

[Latin Script RZ-LGR Proposal \(XML\)](#)

[Latin Script RZ-LGR Proposal \(HTML\)](#)

[Latin Script RZ-LGR Supporting Document](#)

[Provide Your Input](#)

## Background

Root Zone Label Generation Rules ([RZ-LGR](#)) are developed to define a conservative mechanism to determine valid Internationalized Domain Name (IDN) top-level domains (TLDs) and their variant labels.

The successful development of Label Generation Rules depends on having community-based Generation Panels for each script that will be used in the root zone. Generation Panel members are representatives from each of the communities that use a particular script or writing system and their role is to identify the valid characters, variants, and Whole Label Evaluation (WLE) rules for the given script or writing system. Each Generation Panel starts by considering the code points for the relevant script that are present in the Maximal Starting Repertoire ([MSR](#)) and based on these, develops a Label Generation Rules proposal to be used to generate TLD labels for that script. In doing so, they may need to coordinate efforts with other Generation Panels whenever their respective scripts are closely related. These proposals are then reviewed by the community through a Public Comment proceeding and then by an expert Integration Panel for approval and integration into the Label Generation Rules for the Root Zone (RZ-LGR). The details are defined in the [LGR procedure](#).

Currently, there are multiple script communities that have formed Generation Panels and are actively working towards finalizing Label Generation Rules (LGRs) for the root zone including Japanese and Myanmar scripts. LGRs for the following 18 scripts have already been [completed](#) and integrated: Arabic, Bangla, Chinese, Devanagari, Ethiopic, Georgian, Gujarati, Gurmukhi, Hebrew, Kannada, Khmer, Lao, Malayalam, Oriya, Sinhala, Tamil, Telugu, and Thai. In addition, Armenian, Cyrillic, Greek, and Korean scripts have also been completed and are waiting to be integrated.

## Next Steps

Following the Public Comment proceeding and submission of the proposal by the Latin Script GP, which will reflect any changes due to community feedback, the Integration Panel will evaluate the proposal. If the proposal is successfully evaluated, it will be integrated into a subsequent version of the [RZ-LGR](#). If not, the proposal will be sent back to the Latin Script GP with comments for further consideration and revision.

## Supporting Information

This additional information from ICANN org provides more context for this Public Comment Proceeding and may help you review the proposals for input and publish a submission.

Supporting Information
<a href="#">Repository of Guidelines and other relevant documents</a>
<a href="#">LGR proposals already received</a>
<a href="#">Guidance on Designing LGRs Supporting Variant Labels</a>
<a href="#">LGR Toolset</a>
<a href="#">Machine readable specification for Label Generation Rules</a>
<a href="#">Root Zone Label Generation Rules</a>
<a href="#">Maximal Starting Repertoire</a>
<a href="#">Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA Labels (LGR Procedure)</a>

## FINAL VERSION SUBMITTED (IF RATIFIED)

*The final version to be submitted, if the draft is ratified, will be placed here by upon completion of the vote.*



AL-ALAC-ST-1121-01-00-EN.pdf

---

**FINAL DRAFT VERSION TO BE VOTED UPON BY THE ALAC**

*The final draft version to be voted upon by the ALAC will be placed here before the vote is to begin.*



AL-ALAC-ST-1121-01-00-EN.docx

See [Google Doc \(comment-only\)](#)

### ALAC Statement on Proposal for Latin Script Root Zone Label Generation Rules

The At-Large Advisory Committee (ALAC) welcomes the opportunity to comment on the proposed Latin Script Root Zone Label Generation Rules. The ALAC appreciates the efforts of the Latin Generation Panel (LGP) but questions some of the instructions received by and assumptions made by the panel. These “quirks” have led to a proposal that might not be the best for multilingual expression or the mitigation of DNS abuse.

#### Principle of Least Astonishment

The principle of least astonishment proposes that a component of a system should behave in a way that most users will expect it to behave. The now expert nature of the Latin GP suggests that it is no longer, if ever, representative of “most users.” Therefore, establishing a high bar for possible confusion is counter intuitive. If even a single Panel member considers that two code points should be a variant, that conclusion should be applied to the Panel findings. Otherwise, the Panel is implying that non-experts will be more successful in spotting differences than some of its experts. This is an entirely unreasonable expectation of a *typical* end user and making this assumption further paves the way for misdirection and consequent DNS abuse.

#### Underlining

The Latin GP has concluded that underlining of domain names (as happens routinely) does not obscure diacritics below the line. While *some* software products put a blank space around such diacritics, it is not at all clear that end users are cognizant of this practice. Furthermore, many software products, including ICANN's own Public Comment system, do not use these spaces. Accordingly, the ALAC recommends revising that conclusion.

#### Capital Letters

The ALAC notes, in the Latin Generation Panel's report, that the Internationalized Domain Name (IDN) Project has directed the Generation Panels (GPs) to ignore upper case letters when identifying variants. The apparent rationale for this direction is that domain names are restricted to lower case. Unfortunately, a half century of experience has taught end users that, in a domain name, upper and lower case are entirely interchangeable. Thus, a domain name of [www.example.com](#) and of [www.example.COM](#) will take the user to the same place. The users neither know nor care that this is an artifact of their browser, rather than a feature of the DNS.

As a result, if the user is presented with a domain name using Cyrillic lower case for the TLD, such as [www.example.](#), they will naturally interpret it to be the usual .com, just with the TLD capitalized. What they will NOT do, despite in expectation of the IDN Project, is look at the TLD and notice the third letter doesn't look like a Latin lower case M and conclude this is not a .com domain name. The potential for DNS abuse is obvious.

Accordingly, the ALAC encourages the LGP to revisit the instructions, attributed to the IDN Project, and consider treating capital letters as a special case.

#### Repertoire

The objective of the IDN project is to make domain names available in the languages of all non-native speakers of English. To restrict the repertoire for the Latin script to less than half of the living languages which use that script is contrary to that goal. The ALAC believes that the LGP should go back and include all of the languages which use the Latin script. (At a minimum, languages which have more native speakers than the smallest of the “official languages” which are already included, should be added.) For example, Hawaiian has perhaps 25,000 native speakers, but is included because it is an official language of the State of Hawaii, despite the stated threshold of 1 million speakers.

#### Conclusion

It is the considered position of the ALAC that, while great strides have been made by the Latin Generation Panel, more must be done to enable more users, with sufficient security, to use the DNS. Again, we appreciate the opportunity to comment on the draft proposal and look forward to further discussions on this topic.

## DRAFT SUBMITTED FOR DISCUSSION

*The first draft submitted will be placed here before the call for comments begins. The Draft should be preceded by the name of the person submitting the draft and the date/time. If, during the discussion, the draft is revised, the older version(S) should be left in place and the new version along with a header line identifying the drafter and date/time should be placed above the older version(s), separated by a Horizontal Rule (available + Insert More Content control).*

See [Google Doc \(comment-only\)](#)

### ALAC Statement on Proposal for Latin Script Root Zone Label Generation Rules

The At-Large Advisory Committee (ALAC) welcomes the opportunity to comment on the proposed Latin Script Root Zone Label Generation Rules. The ALAC appreciates the efforts of the Latin Generation Panel (LGP) but questions some of the instructions received by and assumptions made by the panel. These “quirks” have led to a proposal that might not be the best for multilingual expression or the mitigation of DNS abuse.

#### Principle of Least Astonishment

The principle of least astonishment proposes that a component of a system should behave in a way that most users will expect it to behave. The now expert nature of the Latin GP suggests that it is no longer, if ever, representative of “most users.” Therefore, establishing a high bar for possible confusion is counter intuitive. If even a single Panel member considers that two code points should be a variant, that conclusion should be applied to the Panel findings. Otherwise, the Panel is implying that non-experts will be more successful in spotting differences than some of its experts. This is an entirely unreasonable expectation of a *typical* end user and making this assumption further paves the way for misdirection and consequent DNS abuse.

#### Underlining

The Latin GP has concluded that underlining of domain names (as happens routinely) does not obscure diacritics below the line. While *some* software products put a blank space around such diacritics, it is not at all clear that end users are cognizant of this practice. Furthermore, many software products, including ICANN's own Public Comment system, do not use these spaces. Accordingly, the ALAC recommends revising that conclusion.

#### Capital Letters

The ALAC notes, in the Latin Generation Panel's report, that the Internationalized Domain Name (IDN) Project has directed the Generation Panels (GPs) to ignore upper case letters when identifying variants. The apparent rationale for this direction is that domain names are restricted to lower case. Unfortunately, a half century of experience has taught end users that, in a domain name, upper and lower case are entirely interchangeable. Thus, a domain name of [www.example.com](#) and of [www.example.COM](#) will take the user to the same place. The users neither know nor care that this is an artifact of their browser, rather than a feature of the DNS.

As a result, if the user is presented with a domain name using Cyrillic lower case for the TLD, such as [www.example.](#), they will naturally interpret it to be the usual .com, just with the TLD capitalized. What they will NOT do, despite in expectation of the IDN Project, is look at the TLD and notice the third letter doesn't look like a Latin lower case M and conclude this is not a .com domain name. The potential for DNS abuse is obvious.

Accordingly, the ALAC encourages the LGP to revisit the instructions, attributed to the IDN Project, and consider treating capital letters as a special case.

#### Repertoire

The objective of the IDN project is to make domain names available in the languages of all non-native speakers of English. To restrict the repertoire for the Latin script to less than half of the living languages which use that script is contrary to that goal. The ALAC believes that the LGP should go back and include all of the languages which use the Latin script. (At a minimum, languages which have more native speakers than the smallest of the “official languages” which are already included, should be added.) For example, Hawaiian has perhaps 25,000 native speakers, but is included because it is an official language of the State of Hawaii, despite the stated threshold of 1 million speakers.

#### Conclusion

It is the considered position of the ALAC that, while great strides have been made by the Latin Generation Panel, more must be done to enable more users, with sufficient security, to use the DNS. Again, we appreciate the opportunity to comment on the draft proposal and look forward to further discussions on this topic.