

IDN Variant TLD Implementation: Rationale for Root Zone Label Generation Rules

Contents

- 1 Background..... 2
- 2 Introduction 2
- 3 ICANN's Role in Coordinating the DNS 3
- 4 Motivation of IDN Variant TLDs for the DNS 4
- 5 Requirements for Compliance with Standards..... 6
- 6 Expected User Experience 8
- 7 The Solution through Label Generation Rules..... 10
 - 7.1 Generation Panel: Developing a Script-Specific LGR 10
 - 7.2 Integration Panel: Creating a Unified LGR 11
 - 7.3 Public Feedback on Proposed LGR 12
- 8 IDN Variant Labels in ICANN's TLD Allocation Processes 13
- 9 IDN Variant Analysis vs. String Similarity 13
- 10 Conclusion 17

1 Background

The current document is part of the set of the following six documents released for public comment:

- A. [IDN Variant TLD Implementation – Executive Summary](#)
- B. [IDN Variant TLD Implementation – Motivation, Premises and Framework](#)
- C. [IDN Variant TLD Implementation – Recommendations and Analysis](#)
- D. [IDN Variant TLD Implementation – Rationale for RZ-LGR](#)**
- E. [IDN Variant TLD Implementation – Risks and their Mitigation](#)
- F. [IDN Variant TLD Implementation – Appendices \(A: Definitions, B: Use of ROID, C: Limiting Allocated Variant TLDs\)](#)

2 Introduction

As ICANN, through the IANA function, is responsible for management of the Internet's Domain Name System (DNS) root zone, it implies that ICANN also needs to specify relevant rules for determining the labels for the root zone. Traditionally, domain labels have been formed by ASCII characters (letters a-z and A-Z, digits 0-9 and hyphen "-", known as the Letter-Digit-Hyphen or LDH scheme). Top-level domains have had additional constraints from the outset because these labels are in the Internet's root zone. [RFC 1123](#) has limited top-level domain labels to alphabetic (letters) only. The domain name label mechanism has since been extended to allow for domain names in multiple scripts based on the Unicode standard, called internationalized domain names, for which IDNA 2008 (RFCs 5890-5893) is the current applicable standard.

Like the ASCII based labels, there need to be specialized rules for the top-level domain labels, as required by IDNA2008 and other relevant standards. ICANN's multistakeholder community has developed the Root Zone Label Generation Rules (RZ-LGR) Procedure, [adopted](#) by the ICANN Board in 2013, to develop these rules, and has subsequently used the procedure to develop the Root Zone Label Generation Rules (RZ-LGR).

Because the DNS root zone is a resource shared by all Internet users worldwide, the RZ-LGR has been developed to minimize conflicts, end-user risks, and compatibility issues, regardless of language or script. Ensuring that the needs of a global audience using various scripts are supported in a secure and stable manner may require design compromises in some cases, which may not be considered optimal from the perspective of a single language community.

This report describes work that has been done on IDN variant labels for the DNS, as defined by the RZ-LGR. The scope of the discussion in this report is limited at present to considering these issues at the top level of the DNS.

3 ICANN's Role in Coordinating the DNS

As confirmed by the international Internet community in its work on the global stewardship of the IANA functions, the organizational [Bylaws](#) provide that the mission of ICANN is to ensure the stable and secure operation of the Internet's unique identifier systems. Specifically, ICANN:

(i) Coordinates the allocation and assignment of names in the root zone of the Domain Name System ("DNS") and coordinates the development and implementation of policies concerning the registration of second-level domain names in generic top-level domains ("gTLDs"). In this role, ICANN's scope is to coordinate the development and implementation of policies:

- For which uniform or coordinated resolution is reasonably necessary to facilitate the openness, interoperability, resilience, security and/or stability of the DNS including, with respect to gTLD registrars and registries, policies in the areas described in [Annex G-1](#) and [Annex G-2](#); and
- That are developed through a bottom-up consensus-based multistakeholder process and designed to ensure the stable and secure operation of the Internet's unique names systems.

...

(iv) Collaborates with other bodies as appropriate to provide registries needed for the functioning of the Internet as specified by Internet protocol standards development organizations. In service of its Mission, ICANN's scope is to provide registration services and open access for registries in the public domain requested by Internet protocol development organizations.

Its Bylaws also commit ICANN to:

- Preserve and enhance the administration of the DNS and the operational stability, reliability, security, global interoperability, resilience, and openness of the DNS and the Internet;
- Maintain the capacity and ability to coordinate the DNS at the overall level and work for the maintenance of a single, interoperable Internet;
- Make decisions by applying documented policies consistently, neutrally, objectively, and fairly, without singling out any particular party for discriminatory

treatment (i.e., making an unjustified prejudicial distinction between or among different parties).

ICANN undertakes its responsibility for the allocation and assignment of names in the DNS root zone primarily in a coordination role. In developing rules and procedures for internationalized domain names in the root zone, ICANN's processes are designed to support the security and stability of the DNS and to be informed by stakeholder participation from those affected. ICANN has also worked collaboratively with other bodies creating standards in this area, such as the Internet Engineering Task Force (IETF) and the Unicode Consortium.

4 Motivation of IDN Variant TLDs for the DNS

Internationalized Domain Names (IDNs) enable people around the world to use domain names in local languages and scripts. IDNs are formed using characters from different scripts, encoded by the Unicode standard, and used as permitted by relevant IDN protocols (RFCs 5890-5893).

ICANN organization has instituted the IDN Program to assist in the development and promotion of a multilingual Internet using IDNs to provide better accessibility to the global internet community. The program is primarily focused on the planning and implementation of IDN top-level domains (TLDs), including IDN country code TLDs and generic TLDs. The IDN Program also supports projects geared towards effective use of IDNs at the second-level of the DNS, as guided by the ICANN multistakeholder community.

Some script communities have identified that technically distinct domain labels may be considered indistinguishable or interchangeable with other domain labels. Such labels must minimally be identified and managed to ensure that end-users are prevented from any security threats that their use in the DNS may cause. Activation of such labels may further be required to promote accessibility of the IDNs, as different language communities using the script may use a different version of the label. Achieving this security and accessibility goal in a stable manner is one key issue being addressed in the IDN program

For example, the following may be considered variant labels by the script communities for different reasons. Examples are also illustrated in Figure 1 below.

- Chinese
 - 中國 and 中国 (traditional and simplified Chinese)

- Arabic
 - كتاب - کتاب (same character represented by different code points)
 - شبكة - شبكه (alternate word ending considered interchangeable)
 - امارات - إمارات (under-specification by not writing diacritics)
- Cross-Script (scripts contain visually indistinguishable characters)
 - Latin and Cyrillic
 - epic – epic
 - Armenian and Greek
 - ղԼԾ – ղԼԾ
 - Kannada and Telugu
 - ల౞గ౞ - ల౞గ౞

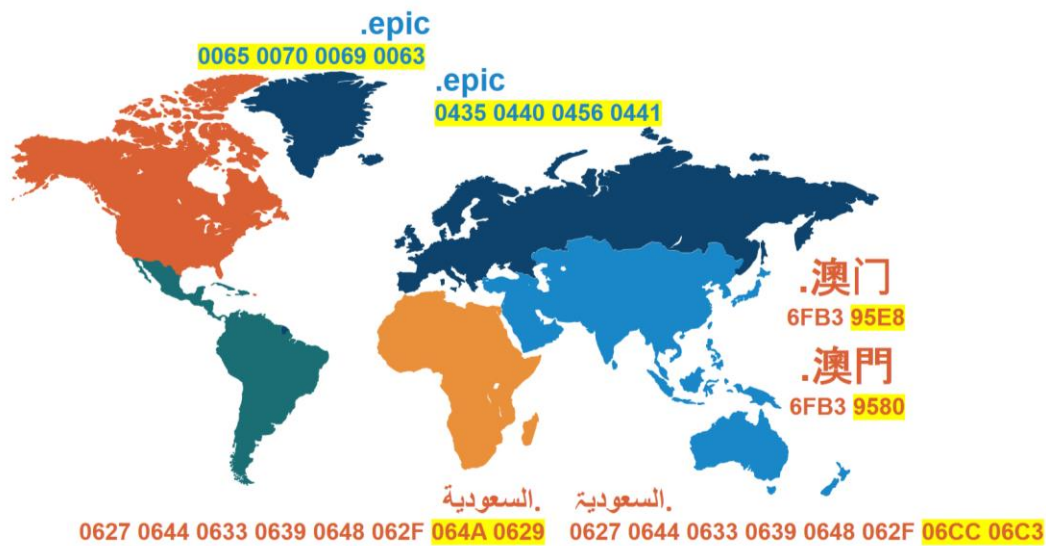


Figure 1: Examples of Cross-Script and Within-Script Variant Labels

To address these complex linguistic, technical and policy issues, ICANN org first [undertook an initiative](#) to engage six script communities that identified requirements of variants for these scripts. For a holistic understanding of the range of possible challenges, the scripts studied had represented a range of script types: Alphabetic (Latin, Cyrillic and Greek), Abjad (Arabic), Abugida (Devanagari) and Ideographic (Han Chinese). Experts from these communities undertook detailed analyses, undertook open public feedback on the issues identified for these scripts ([Arabic](#), [Chinese](#), [Cyrillic](#), [Devanagari](#), [Greek](#), [Latin](#)) and then published their [reports](#) in 2011. As anticipated, these case studies demonstrated a significant variety in the types of variant code point issues.

In 2012, ICANN org formed a coordination team, comprised of experts from the case study teams, to advise ICANN in completing a consolidated report (the [Integrated](#)

[Issues Report](#)) to summarize and synthesize the issues identified by the case study teams. The members of the issues report coordination team had expertise in the areas of DNS, IDNA, linguistics, security and scalability, policy, registry/registrar operations, and community representation. The coordination team assisted ICANN organization in producing an integrated report collating and synthesizing the issues associated with the possible inclusion of IDN variant labels in the DNS root zone, including a taxonomy of identified variant code point cases.

As each variant label is formed by a unique sequence of code points, each label is different in the context of the DNS. The mapping of different labels into variant label sets may be done through different processes; however, in any case, a variant set needs to be defined by some procedure. The Integrated Issues Report concluded that a need existed for a comprehensive set of label generation rules for the root zone as a prerequisite for implementation of IDN variant TLDs, stating that:

“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.”

This implied that the community must come together to form a single set of rules to define variants of IDN TLDs, because defining TLD labels arbitrarily or on a case to case basis may produce contradicting results, leading to security and stability issues.

5 Requirements for Compliance with Standards

Without adherence to the RZ-LGR, it may not be possible to maintain adherence to a number of pre-existing technical standards, creating added risk of technical instability and end-user issues.

The IDN standards are based on the [Unicode standard](#). [Unicode Technical Report #36](#), Unicode Security Considerations, notes that:

“because Unicode contains such a large number of characters, and incorporates the varied writing systems of the world, incorrect usage can expose programs or systems to possible security attacks. This document describes some of the security considerations that programmers, system analysts, standards developers, and users should take into account.”

[Unicode Technical Report #39](#), Unicode Security Mechanisms, also specifies mechanisms that can be used to detect possible security problems.

[RFC 1591](#), Domain Name System Structure and Delegation, provides information on the structure of domain names and the administration of delegated domains. RFC 1591 provides that: “The Internet Assigned Numbers Authority (IANA) is responsible for the overall coordination and management of the Domain Name System (DNS), and especially the delegation of portions of the name space called top-level domains.”

RFCs 5890-5984 comprise the IDNA2008 standard. Together, this collection of documents describes the protocol and usage context for a revision of Internationalized Domain Names for Applications (IDNA) that was largely completed in 2008, known within the series and elsewhere as "IDNA2008."

[RFC5890](#), Internationalized Domain Names for Applications (IDNA): Definitions and Document Framework, notes that “DNS zone administrators may impose restrictions, beyond those imposed by DNS or Internationalized Domain Names in Applications (IDNA), on the characters or strings that may be registered as labels in their zones. Because of the diversity of characters that can be used in a U-label and the confusion they might cause, such restrictions are mandatory for IDN registries and zones even though the particular restrictions are not part of these specifications.” It is further explained that “DNS zone administrators may impose restrictions ... that try to minimize characters that have similar appearance or similar interpretations.” RFC 5890 also emphasizes that: “the character of IDNA requires that it be understood and properly used by those whose responsibilities include making decisions about:

- what names are permitted in DNS zone files,
- policies related to names and naming, and
- the handling of domain name strings in files and systems, even with no immediate intention of looking them up.”

[RFC 5891](#), Internationalized Domain Names in Applications (IDNA): Protocol notes that “Registries at all levels of the DNS, ... [including] the top level, are expected to establish policies about label registrations.”

RFC 5891 specifically refers to the rationale in [RFC 5894](#): Internationalized Domain Names for Applications (IDNA): Background, Explanation, and Rationale that any domain name registry, including that of the root zone,

“should develop and apply additional restrictions as needed to reduce confusion and other problems ... For many scripts, the use of variant techniques ... may be helpful in reducing problems that might be perceived by users. ... In general,

users will benefit if registries only permit characters from scripts that are well-understood by the registry or its advisers,” suggesting some cases, e.g. “reduce opportunities for confusion by constructing policies that disallow characters used in historic writing systems or characters whose use is restricted to specialized, highly technical contexts.”

Additional guiding principles are defined in [RFC 6912](#), Principles for Unicode Code Point Inclusion in Labels in the DNS, which provides that some principles apply to every DNS zone. Some additional principles apply to all public zones, including the root zone, and some principles apply only to the root zone. This means that zones higher in the DNS tree tend to have more restrictive rules, and zones lower in the DNS tree tend to have less restrictive rules, since they [the latter] are used within a more narrow context. In general, the relevant context for a principle is that of the zone, not that of a given subset of the user community; for the root zone, for example, the context is “the entire Internet population.”

The work of the RZ-LGR on restricting code points and defining their variants for a zone is in line with and expected by IDNA2008 and other relevant standards. As ICANN, through the IANA function, is responsible for management of the DNS Root Zone, it follows that ICANN also needs to specify relevant rules for determining the valid labels and their variant labels for the Root Zone. These guidelines are part of the RZ-LGR Procedure developed by the ICANN community and are realized through the RZ-LGR.

6 Expected User Experience

A [2013 report](#) by ICANN org and consulting experts examined the user experience implications of active variant TLDs, taking into account the security and stability of the DNS as well as exploring issues of usability to the communities the DNS is designed to serve. The report provided a set of guiding principles for activating the variants of IDN TLDs, including:

- Security (variants must minimize the risks introduced by IDNs)
- Predictability (variants should behave and function as users expect in their language and script environments)
- Equivalency (variants must be managed by the same entity and direct users to related content)
- Consistency (variants should behave similarly within and across TLDs and supporting technology)

The paper also recommended that ICANN define technical requirements and engage with standards organizations, such as the IETF, to determine how IDN variants should be consistently implemented.

Use of the RZ-LGR supports these principles by facilitating the involvement of script communities in developing transparent and predictable rules and promoting compliance with relevant standards and protocols. In an environment without such an established and recognized mechanism, variant determinations may be unreliable and the principles of security and consistency for users are more difficult to fulfill.

ICANN's Security and Stability Advisory Committee (SSAC) commented on the user experience report, stating that:

“The root zone is necessarily shared by everyone on the Internet, and needs a set of LGR that ensures minimal conflict, minimal risk to all users (independent of the language or script they are using and independent of gTLD or ccTLD), and minimal potential for incompatible change over time.”

The SSAC's recommendations in its [SAC060 advice](#) included:

- Recommendation 1: The root zone must use one and only one set of Label Generation Rules (LGR).
- Recommendation 2: 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.

As noted by SSAC in its comments, the root zone lacks other contexts that can be used by a registry to restrict LGR for that particular TLD. Due to difference in context for second-level domain names, different rules for the same script may exist across TLDs. However, SSAC cautions that, “Applying such a model to the root zone would cause stability issues.” Therefore, SSAC emphasizes that **“The root zone must use one and only one set of rules for the Root LGR procedure.”**

The LGR Procedure contains multiple opportunities for feedback and comment on proposed rulesets, and allows the LGR to be updated by the community based on new information, but requires continued prioritization of security and stability. Accordingly, the LGR Procedure anticipates evolution and revision. If the members of the community do not agree with the calculation of a variant set according to the RZ-LGR, they may reconvene the relevant Generation Panel, present their case to the relevant Generation Panel, proposing the changes along with the rationale. If the proposed update proceeds, the Integration Panel must agree that the reasons are valid and cause

no harm from a security and stability point of view, and then update the RZ-LGR accordingly. Thus, both community and technical safeguards are built into the process and questions are resolved as the variant set is formulated, rather than afterward.

7 The Solution through Label Generation Rules

The label generation rules govern the way a zone is operated. The RZ-LGR provides a mechanism for creating and maintaining the rules with respect to IDN labels for the root. This mechanism can be used to determine which Unicode code points are permitted for use in U-labels in the root zone, what variant labels (if any) are possible to allocate in the root zone, and what variant labels (if any) are automatically blocked.

The procedure for developing IDN TLD variant label generation rules consists of two passes.

7.1 Generation Panel: Developing a Script-Specific LGR

The first pass creates a set of label generation rules specific to a given script or writing system; this task is carried out by Generation Panels composed of community members with deep experience or interest in the script or writing system used by some community of Internet users. The Generation Panels are community-based panels that have the task of proposing the LGR for the particular script used in each community for the Root Zone. Generation Panels are formed based on expertise and diversity requirements. Each Generation Panel is generally comprised of a chair and a number of representatives from the relevant script community, with technical expertise in DNS, IDNA, Unicode, and linguistics. In some cases, a community already has a working group on IDN and variant issues which forms the basis of a Generation Panel. Generation Panel members are responsible for researching and providing input on issues in developing the LGR submission, participating in drafting work, and engaging regularly with the panel. The Generation Panel membership has been open; panels have generally accepted any volunteer interest from the community and have not restricted membership. For membership details and activities, included archives of emails and call recordings (where online calls were done), see the panel specific wiki pages in the left menu of the [community webpages for the Generation Panels](#).

Generation Panels have also worked in coordination with each other to ensure that cross-script issues are addressed in developing the LGR for a particular script. For example, the Khmer, Lao and Thai Generation Panels met multiple times at ICANN

meetings to coordinate on evaluating cross-script variants. Similarly, the Chinese, Japanese, and Korean Generation Panels worked together on finding common solutions to the overlapping code point repertoire between these writing systems, as well as other topics of mutual interest. Similarly, Latin, Cyrillic and Greek panel members have also been interacting and discussing how to address cross-script variant code points, which are in many cases homoglyphs. And Neo-Brahmi and Sinhala panels who have been working on scripts in South Asia have also been interacting to address cross-script variant cases for Bangla, Devanagari, Gujarati, Gurmukhi, Kannada, Malayalam, Oriya, Sinhala, Telugu and Tamil scripts.

7.2 Integration Panel: Creating a Unified LGR

Each Generation Panel submits its LGR proposal to an independent Integration Panel, coordinated by ICANN, that has responsibility for the second pass. The Integration Panel consists of experts in DNS, Unicode and scripts. This second pass involves integrating the proposals into a single unified LGR for the root zone, taking into account the need for a secure, stable and reliable DNS root zone. The decisions by the Integration Panel are required to be unanimous, representing the caution and conservatism required to keep the root zone stable and secure.

The RZ-LGR procedure recognizes that a shared resource, like the root zone, requires cross-script expertise, but that each script and writing system will bring its own issues. Because of the two-panel structure, the procedure relies upon a high degree of expertise, openness and transparency by not confining itself only to experts. Initial development is undertaken by generation panels with expertise in their respective scripts. The Integration Panel, on the other hand, has expertise in areas of DNS, IDNA, Unicode and Linguistics with a responsibility for total review and is ultimately responsible for the label generation rules as deployed in the root zone.

These two kinds of panels may be assisted by advisors, who observe the activities of any or all active panels; who provide comment and advice on the topics of IDNA, Unicode, DNS, linguistics, ICANN policy and process, or other matters; but who do not otherwise have a formal role in making a decision. On the request of the panels, advisors are appointed based on an open call for expertise and qualified by ICANN organization.

Because there are many scripts and writing systems, there are multiple generation panels, but because there is only one root zone, there is only a single Integration Panel. The proposed LGRs produced by the generation panels are submitted to the integration panel for review, and, if approved, for integration into the LGR for the root zone.

7.3 Public Feedback on Proposed LGR

Each proposal from a generation panel undergoes multiple public comment phases:

1. Upon submission of the proposal by the respective Generation Panel, it is released for public comment to allow those who have not participated in the panel to make their views known to the members of the panel and the community at-large. Based on the feedback, the panel finalizes the proposal for submission to the Integration Panel for potential integration into the RZ-LGR.
2. Once the proposal is approved by the Integration Panel, the integrated version of the RZ-LGR is released for public comment again for the community to comment on the updated version of RZ-LGR (e.g. see [public comment](#) announcement for the second version of the RZ-LGR).

The resulting label generation rules provide a consistent and predictable set of permissible code points for IDN TLDs and provide a way to determine whether there are variant labels (and if so, what they are). The RZ-LGR thus enables consistent and predictable definition of variant sets, which are TLD labels that are considered to be variants of a given label based on inclusion of characters that are considered variant characters to one another.

The RZ-LGR is the mechanism for establishment of variant sets but does not make all determinations about the treatment of labels in a given variant set. Rather, the output of the RZ-LGR is an input into ICANN procedures (e.g., determining whether a particular applicant may be allocated its requested label and whether that requested label may be delegated into the root zone) and other processes. For the goals of security and stability to be met, downstream processes using the RZ-LGR output must maintain adherence to the variant set composition as defined by the RZ-LGR. For example, two labels that are variant labels according to the RZ-LGR could not be allocated to different entities as a result of an application evaluation process.

As described in the sections above, there can only be a single RZ-LGR. In keeping with its importance to Internet users around the globe, the RZ-LGR has been developed in a transparent and open process requiring participation by policy, IDN, and Unicode experts in a breadth of roles, as well as governments and TLD operators, taking into account broad linguistic diversity and providing multiple avenues for public comment.

Processes or decisions that attempt to incorporate other sources or substitute other definitions for variant sets in an effort to allocate names differently or enact different treatment for labels in a variant set break the principles of security and stability that are critical to the DNS.

The RZ-LGR can be found [here](#).

8 IDN Variant Labels in ICANN's TLD Allocation Processes

In discharging its responsibilities for management of the DNS root zone, ICANN's existing processes for allocation of TLDs include provisions on handling of variant labels where relevant. However, these processes do not delegate the responsibility of determining variant sets.

For example, the 2012 gTLD Applicant Guidebook provided that each application submitted would contain one applied-for gTLD string (label). An applicant also had the option to declare any variant strings for the TLD in its application. However, the Guidebook stated that "no variant gTLD strings will be delegated through the New gTLD Program until variant management solutions are developed and implemented. Declaring variant strings is informative only and will not imply any right or claim to the declared variant strings."

The IDN ccTLD Fast Track, established in 2009, predates the development of the RZ-LGR. In this process, any variant TLD labels desired by the requester for delegation must be indicated by the requester, and, if successfully evaluated, those desired variant TLDs will be allocated to the requester. However, the documentation states that: "This does not mean that the variant TLD will be delegated in the DNS root zone. It will be allocated to the requester in order to be reserved to the entitled manager for potential future delegation in the DNS root zone." Multistakeholder community is still to work on an IDN variant ccTLD management policy, which will replace the Fast-Track process. This is noted in the proposed IDN ccTLD policy by the ccNSO: "To date (March 2013) identifying the issues pertaining to the management of variant TLD's [sic] are still under discussion by the community, in particular the delineation of technical, policy and operational aspects. For this reason policy recommendations pertaining to the management of variant IDN ccTLDs, if any, are not included, but will be added at a later stage."

9 IDN Variant Analysis vs. String Similarity

Independent of the RZ-LGR, the TLD allocation processes described above include additional steps to guard against user confusion by, for example, assessing possible types of similarity (e.g., visual similarity) that would tend to cause user confusion. In this instance, labels that are proposed for inclusion in the DNS root zone are reviewed case-by-case by expert panels, and determinations as to the degree and basis of similarity may vary depending on the circumstances and elements of the decision process. These comparisons occur on labels that have already met protocol requirements for the DNS, but it would be problematic to rely on this type of process for generating the labels themselves, which requires adherence to a carefully developed and accepted set of rules. The DNS is a shared resource for all Internet users and a case by case approach to generating labels and defining variant sets would run counter to achieving goals such as security, stability, predictability, and consistency.

It is important to note that, while there may be overlapping cases where some variant labels are also seen as having visual or another type of similarity, this is not the same test as is applied by the RZ-LGR. Variant analysis is based on “same” or interchangeable code points as determined by the community process, which may or may not involve visually similar characters.

The RZ-LGR Procedure, while defining “IDN variants” says that:

- “An IDN variant, as understood here, is an alternate code point (or sequence of code points) that could be substituted for a code point (or sequence of code points) in a candidate label to create a variant label that is considered the “same” in some measure by a given community of Internet users.”

However, the Procedure also acknowledges immediately following the definition that:

- “There is not general agreement of what that sameness requires, and many of the things people seem to want from that sameness are not technically achievable.”

While noting the benefits of defining IDN variants, the procedure also acknowledges the limitations.

- “The primary benefit of the LGR process is as a mechanism that delivers hands-off evaluation for these aspects.
- “By doing so, the process may not be able to replace case-by-case analysis altogether: there will still be a role for additional types of review, such as for String Similarity, and which are not included in the LGR process.”

So, not all matters can be settled in the LGR. A line has to be drawn between “same” and “similar” cases.

The LGR Procedure does note what is in the scope to LGR:

- “the LGR process is designed to clear the table of all the straightforward, non-subjective cases, mainly by returning a “blocked” disposition.
- “Even for variants based on visual similarity, there exists a subset of evaluation rules that could be applied in an automated manner, obviating the need for further case-by case or even contextual review.”

But the Procedure notes that this should not go too far into the string similarity discussion:

- “While the process described here could be expanded to address cases of visual similarity, that is not the primary intention”
- “Finally, in investigating the possible variant relations, Generation Panels should ignore cases where the relation is based exclusively on aspects of visual similarity.”

One could infer from these statements in the RZ-LGR Procedure that if two code points are considered “same” by the user community, these should be included as IDN variants (this is not limited to visual similarity, but could also include semantic equivalence, like in Chinese, orthographic conventions or spelling simplification, like in Arabic, homophonic relations, like in Ethiopic, etc., as determined the respective script community). The “straightforward, non-subjective cases” of visual similarity which are indistinguishable by the relevant script community or “same” could be included as IDN variant characters. Beyond these, the analysis goes into the realm of string similarity review, which is beyond the intended scope of the LGR. This is illustrated in Figure 2 below.

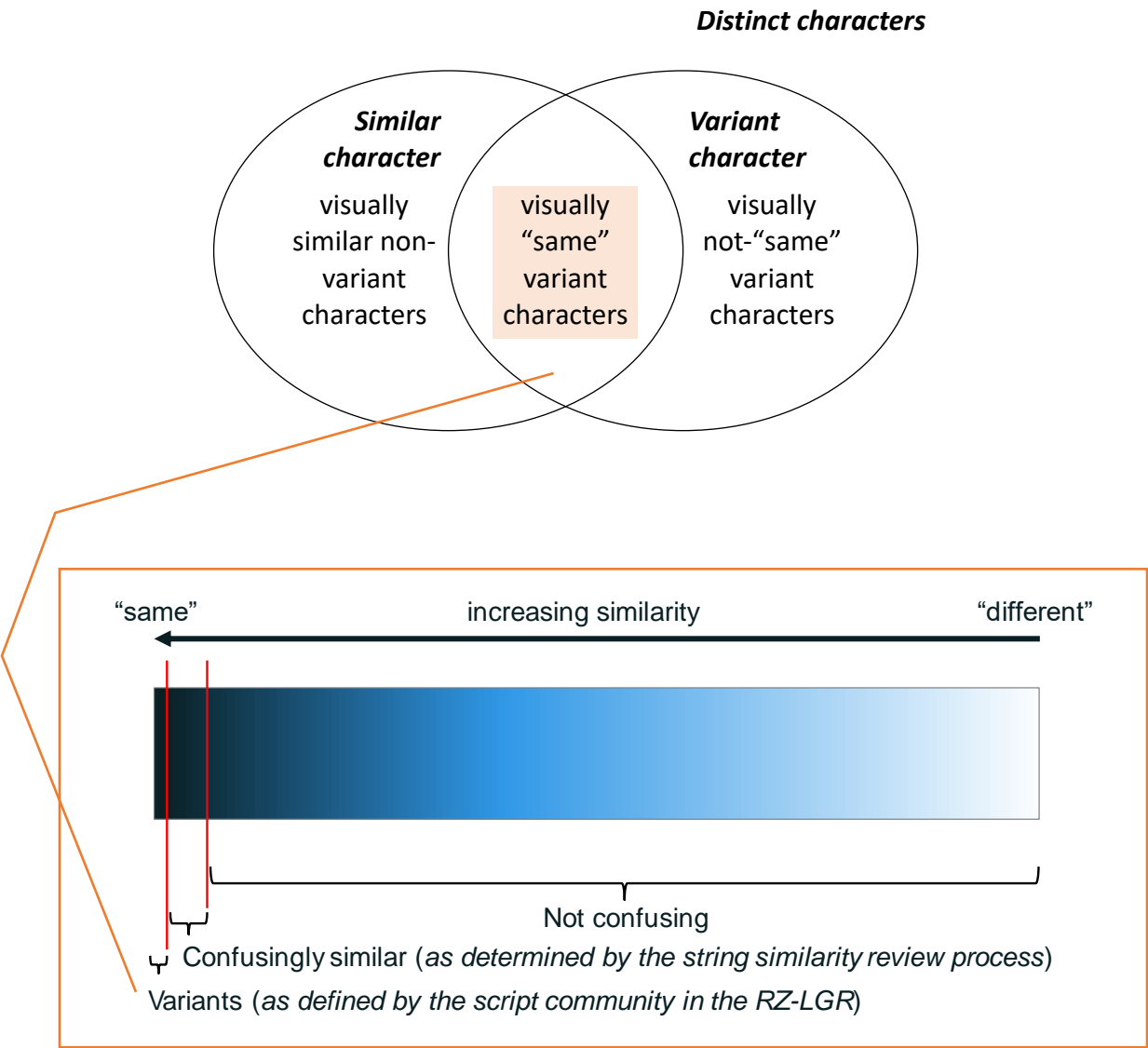


Figure 2: Variant and Similar Characters

Similarity analysis assesses confusability of whole labels, which are not produced through variant characters or code points. These tests should not be mixed: the code point variant analysis is determinative in these situations. Desired variant sets based on visual similarity arguments must yield to the principles of the RZ-LGR process, as illustrated in Figure 3 below. That is, a variant set established by the RZ-LGR cannot be broken because of an argument that certain labels appear similar or dissimilar in some respect.

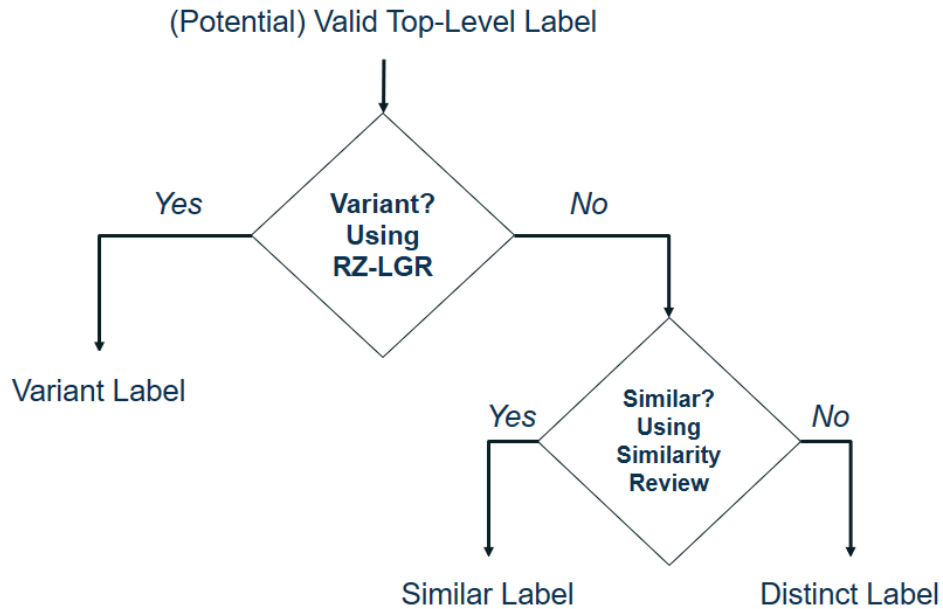


Figure 3: Tiered Process to Evaluate Variant Labels and String Similarity

As a carefully developed body of rules that creates objective and repeatable results, the RZ-LGR takes precedence over alternative formulations of variant sets to best support the objectives of a secure and stable DNS for all users globally. Management of the root zone as a resource for all users requires adherence to a single set of rules to govern variant sets as calculated from using the RZ-LGR.

10 Conclusion

The issues described here are complex and have been discussed by users, technical experts, and language communities for many years. The principles and procedures that are in place today have been developed collectively in an open and transparent process to fulfill the mission of maintaining security and stability for all Internet users. Throughout this work, it has emerged as a common finding that adherence to a single set of label generation rules for the root zone is fundamental, not only for individual cases but for the benefit of the whole system.