

Latin Generation Panel:

# Analysis of public comments received for Latin script RZ-LGR proposal

---

January 27, 2022

Latin Generation Panel (GP) published the Latin Script Root Zone LGR proposal for the [public comment](#) on 23 September 2021. This document collects all nine comments and GP analysis as well as conclusion for each comment.

No. 1	Submitted by At-Large Advisory Committee (ALAC)
Comment	<p>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.</p> <p><b>Principle of Least Astonishment</b> 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.</p> <p><b>Underlining</b> 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.</p> <p><b>Capital Letters</b> 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 <code>www.example.com</code> and of <code>www.example.COM</code> will take the</p>

	<p>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.</p> <p>As a result, if the user is presented with a domain name using Cyrillic lower case for the TLD, such as www.example.com, 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.</p> <p>Accordingly, the ALAC encourages the LGP to revisit the instructions, attributed to the IDN Project, and consider treating capital letters as a special case.</p> <p><b>Repertoire</b></p> <p>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.</p> <p><b>Conclusion</b></p> <p>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.</p> <p>The ALAC Recommends the Latin Generation Panel (LGP) to:</p> <ul style="list-style-type: none"> <li>• Establish a lower bar for possible confusion, as they are the experts - not typical end users. <b>[1]</b></li> <li>• Revise their conclusion that the underlining of domain names (as happens routinely) does not obscure diacritics below the line. <b>[2]</b></li> <li>• Revisit the instructions attributed to the IDN Project and consider treating capital letters as a special case. <b>[3]</b></li> <li>• 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. <b>[4]</b></li> </ul>
Latin GP Analysis	<p><b>[1] Principles for Variants</b></p> <p>The <a href="#">LGR procedure</a>, and guidance from the Integration Panel, gave the Latin script GP the guidelines for its work. The principle referred to in the comments — Least Astonishment Principle— is one of the many principles that the panel needed to balance in the conclusions of the work on repertoire and variants.</p>

However, the fundamental principle was that the RZ-LGR must not be subject to appeal when applied to candidate TLD strings, and therefore only clear cases of variants should be admitted in the LGR algorithm. Those include true homoglyphs, clearly visibly confusable pairs or sets, and clear cases where glyphs are used interchangeably. The Latin GP proposal includes and describes all three types. Mere visual similarity cases were classified to be included in the list of confusable pairs or sets outside the formal LGR.

In the specific case of letters e and é (in public comment #3 and comment#4 to the Latin GP proposal) these are not judged to meet the criteria that the Latin GP has followed and are described in the Latin GP proposal.

Also, it is worth noting that the RZ-LGR is not envisioned as the only tool to validate and review top-level label applications. The RZ-LGR will speed certain mechanical calculations, but it is anticipated to be a complementary to a DNS Stability and Visual Similarity Review Panels (e.g., [Final Report of the GNSO New gTLD Subsequent Procedures](#), Topic 24, under consideration by the ICANN board), also in upcoming rounds, which together would make a comprehensive final determination of eligibility for top-level domains.

## **[2] Underlining**

In the case of the underlining of a hypertext the Latin GP made two conclusions:

On one hand, that underlining can create confusion and make otherwise distinct glyphs indistinguishable. This is true because a line over a glyph may obfuscate certain characteristics of it.

On the other hand, creating variants will not fully resolve the issue of confusable domain names, because the underlining feature obfuscates the text displayed to the users (“the link”) —the text that embeds a fully qualified domain name (FQDN) or, more generally, a Uniform Resource Locator (URL) — and not the domain name per se. The domain name remains unaltered for the end user to inspect.

Furthermore, the link is defined by the content creator, and it does not have to match the URL at all. An example of the link is the text “*click here*” which when clicked or tapped the application takes the user to the URL that is embedded in the link. No variant mechanism could avoid this. Therefore, the Latin GP resolved not to create variant relationships due to underlining.

A more detailed discussion of the issue can be found in Section 6.4.1. of the [Proposal](#).

## **[3] Capital Letters**

Capital letters as such are not IDNA 2008 protocol valid and can therefore never be included in the repertoire. Capital letters can, however, in many cases be used instead of the equivalent lower-case letter. Web browsers will down-

case IDN names before IDNA preparation. The Latin GP, however, has only considered capital letters in the special case of dotted and dotless letter I, both small letter and capital letter as part of the IDNA 2003 compatibility analysis. See Section 6.4.2.2 of the [Proposal](#) for details.

#### [4] Repertoire

The Latin GP agrees with the assertion that, in general, the availability and adoption of Internationalized Domain Names enable communities around the world to create and use online identifiers that best represent their needs.

#### Repertoire selection

The Latin Script is used by 1,189 languages as listed on [Omniglot page](#) (accessed on 13 January 2022). The core alphabet set is comprised by letters A through Z. The usage of these languages varies as does the composition of their alphabets. Some languages show strong usage and are well documented, while other languages lack of authoritative sources for validation.

The universe of letters and marks — or starting repertoire — available to the Latin GP was limited in two ways. On a protocol level, IDNA calculates which Unicode code points match the criteria (e.g. lower case letters, etc.); on a policy level, the Integration Panel excludes certain code points that have been identified as a risk to be used at the root level — in addition to the exclusion of digits and the hyphen. This initial repertoire is called the Maximal Starting Repertoire (MSR) — which all Generation Panels used for their own LGR proposals.

The Latin GP needed to find a practical method to determine the selection of its own repertoire — using the MSR as a starting point. But most importantly, determine a process to navigate through potentially hundreds of alphabets with the limited time and resources available to the panel.

The Expanded Graded Intergenerational Disruption Scale ([EGDIS](#)) provides a measure what could be considered as vitality or strength of a given language. The Latin GP worked through the list of languages that were in the levels 0 through 4 of the EGDIS scale, because these levels showed attributes of sustainability and widespread usage. Levels 5 and higher signal attributes from lower widespread usage to being in the path to extinction.

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.

	<p>3 Wider Communication The language is used in work and mass media without official status to transcend language differences across a region.</p> <p>4 Educational The language is in vigorous use, with standardization and literature being sustained through a widespread system of institutionally supported education.</p> <p>Exceptionally, the Latin GP panel decided to include languages in the EGDIS scale 5 but limited to those languages that exceeded one million native speakers. This was necessary to limit the scope of the work due to limited resources.</p> <p>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.</p> <p>The Latin GP panel was confident that this method, while not comprehensive of all languages, was a good representation of the Latin alphabet used in many languages.</p> <p>With respect to the examples referred to in the comments. The Hawaiian language had, at the time of review, an EGDIS scale of 2, while Obolo (<a href="https://www.ethnologue.com/size-and-vitality/ann">https://www.ethnologue.com/size-and-vitality/ann</a>) with 300,000+ native speakers an EGDIS scale of 5-6a. On this point, the Latin GP Panel would like to note that while a specific Language's alphabet was reviewed there is no guarantee that all words in that specific language can be represented in a top-level domain name. Quite the opposite. For example, the Okina letter from the Hawaiian alphabet resembles the apostrophe, which is prohibited by the protocol. The MSR excludes all letters whose visual representation is too close to a prohibit code point's visual representation. In contrast, the Obolo alphabet (despite not being reviewed) is well represented in the Latin Script repertoire except for one letter ('a' with caron which is used to indicate tone, but usage in writing is not conclusive).</p> <p>The proposal for a Latin Script LGR for the root zone is meant for mnemonics, not words. Even if a language is listed, it does not mean all words from that language can be represented as a domain name in the DNS, quite the contrary. Also, adding more languages to the review cycle does not guarantee finding new letters to be incorporated in the LGR. To conclude, while not all languages were reviewed individually, because of a cost efficiency decision, the Latin GP asserts that the overlap among language's alphabets provides reasonable coverage to all languages using the Latin script in the selection of top-level labels.</p> <p>The GP will include a note in the LGR Proposal to clarify this topic.</p>
Conclusion	<p>[1], [2], [3] No requirement to update the proposal.</p> <p>[4] The note on repertoire is added to the LGR proposal. See section 5.2.1 of the <a href="#">Proposal</a>.</p>

No. 2	Submitted by Registries Stakeholder Group (RySG)
Comment	<p>The Registries Stakeholder Group (RySG) welcomes the opportunity to comment on this report and thanks ICANN and the Community for the continued work to develop Reference Label Generation Rules (LGRs). The development and dissemination of additional resources is appreciated and the RySG continues to see the value of Internationalized Domain Names (IDNs) to further the goals of competition and consumer choice for internet users globally.</p> <p>The RySG again thanks ICANN for developing and disseminating these LGR resources. We look forward to further discussions on developing a process for adopting Reference LGRs within the multistakeholder policy development process.</p>
Latin GP Analysis	The GP thanks the RySG for the comment.
Conclusion	No requirement to update the proposal.

No. 3	Submitted by CORE Association
Comment	<p><b>Consideration of diacritics as variants [1]</b></p> <p>We disagree with the general exclusion of diacritics as variants of the base characters. that is, a and a acute or grave or a and a with tilde are not considered as variants. We have the impression that this exclusion is not warranted in the real world of DNS usage. The fact is that if we were to ask any user whether são paulo.tld, méxico.tld or québec.tld are the same domains as saopaulo.tld, mexico.tld and quebec.tld everybody (but some IDN experts, typographers or language teachers, perhaps) would say they are the same. The reason is not simply visual. It is the obstinate reality of over 25 years of strong conviction that words with diacritics are used in domain names without them. It is fair to say that as an industry we have failed to provide a comprehensible and easy use of IDNs so far. But the net result is what it is: users will universally see the string without the diacritic as the "main" version of the one with such diacritic. Perhaps not in German where Müller.tld "may" also be mueller.tld, but it is certainly the overwhelming view of Latin-script users as of today. We believe that the standard that should be applied when considering variants is not only, and not mainly, the perception of, say, typography experts or IDN experts but, rather, that of the average DNS user. We also believe that the result seems guided by the principle of limiting the number of variants, but in our opinion the goal should be to serve the best interests of the users and prevent excessive confusion, not a quantitative, aprioristic, question of principle.</p>

This point about diacritics has a clear and direct implication in existing TLDs, for instance .quebec managed by a customer of CORE. It is evident to any one in Québec, and also any DNS user aware of what a diacritic is, that .quebec and .québec can simply not be 2 different things, two different TLDs. They must be a single TLD, managed by the same Registry where second-level domains in both the version with accent and without it must belong to the same Registrants managed by the same Registrars, with the same Nameservers.... In practical terms, .quebec and .québec must be variants of the same TLDs anything else would be closer to a fraud than a confusion. The point is not that they are similar, is that they cannot be perceived as anything else than variants of the same thing. We request, therefore that the question of diacritics be revisited in view of accepting them as variants when warranted.

#### **Inclusion of additional code points such as middle dot [5]**

We respectfully disagree with the exclusion of some codepoints which are, for example, perfectly acceptable according to IDNA 2008 and used at the second level by many Registries without any known issue. One such case is the “Ela geminada” for Catalan language, which uses Middel dot (U+00B7) between two Latin letter "l" (U+006C)

The reason seems to be clear but purely formal: the "Middle dot" is in a given list and not another and needs some “context”. But the former argument is basically based in history more than function; and the second is weak in the sense that there is no real “context” in the sense that the codepoint behaves differently in different locations, before or after given codepoints etc. It is a simple, fixed and unmutable rule: can only be used between that character. Nothing else. We understand the preference for a conservative approach, and if the discussion had lasted for 2 hours, we would understand the outcome. But after so many years, we expect that the aprioristic questions of principle should not prevail without any real check over “how it really works”. We therefore request that the discussion on this concrete codepoint be reopened.

#### **Other Comments (on Language list ordering) [6]**

We would like to express our respect and gratitude to all individuals involved in this effort. We also thank them for taking into account a large number of languages, going beyond ISO 639-1 list and well into the ISO 639-2 one. Nevertheless, we find unnecessary and unfortunate to use the EGIDS Scale to “rank” languages in the provided documentation. EGIDS Scale, as developed by SIL/[Ethnologue.com](http://Ethnologue.com) may be, besides quite controversial, very useful in sociolinguistics analysis of a given language's "health". But using it here seems like establishing a layered status for languages for IDN or DNS usage. In fact, what counts is whether this work has taken into account the characters/codepoints used in a given language, but “ranking” them in your Appendices seems as odd as IANA listing ccTLDs in the root not in alphabetical order but by GDP per capita or by COVID-19 vaccination ratios: both confusing and unwarranted. Please list them by alphabetical order of their ISO 639-2 code.

Latin GP Analysis	<p><b>[1]</b> See Comment 1.</p> <p><b>[5] Inclusion of Middle Dot (U+00B7)</b></p> <p>As has been stated in the Section 5.4.1 of the <a href="#">Proposal</a>, Latin GP proposed for the inclusion of Middle Dot (U+00B7). Only code points in MSR can be part of the Latin repertoire, and the proposal was to include Middle Dot in MSR, but that was not even possible by the decision in the “RZ-LGR Procedure”, section B.3.4.2 [<a href="#">LGR Procedure</a>]. The rationale behind that decision is that code points classified as CONTEXTO in IDNA 2008 [<a href="#">RFC 5892</a>], which Middle Dot is, must not be included in MSR.</p> <p>As long as Middle Dot is outside MSR it is not even possible for Latin GP to suggest its inclusion in the repertoire.</p> <p><b>[6] List of Languages in Appendix B</b></p> <p>The list order by EGDIS scale is not a ranking system of languages, but the designation by the scale definition (e.g., 0-International, 1-National, 2-Provincial, etc.). Appendix B simply reflects, for ease of reference, the data points used by the panel in reviewing the alphabets. Languages in each group are listed in alphabetic order as well.</p>
Conclusion	No requirement to update the proposal.

No. 4	Submitted by PointQuébec
Comment	<p><b>Consideration of diacritics as variants [1]</b></p> <p>PointQuébec must respectfully disagree with the conclusion that e with acute (é) and e are not variants. In general with the decision to consider diacritics as irrelevant for variant purposes. While the characters are different from a grammatical and typographic point of view, nobody in Québec, or, as we see it, elsewhere, would consider domain.quebec and domain.québec as different, separate domains, nor, in fact, québec.domain and quebec.domain. Internet users simply got used to write without diacritics as those special characters were not available, first, and not widely used, later. But they are perfectly equivalent in their mind, and their usage. PointQuébec applied in the 2012 round for .quebec and has been running the Registry since 2014. It has always offered IDNs at the second level as bundles with optional activation: i.e. québec.quebec and quebec.quebec are treated as variants of the same domain. In the 2012 round TLDs in ASCII were not allowed to indicate any future variant, for the purely formal reason that the LGR system now being developed did not exist yet, and somebody apparently thought that ASCII and Latin Script were exactly the same. We firmly believe this must be corrected now and respectfully request that the discussion on at very least “e” and “é” with</p>



	<p>acute as variants, but more generally, diacritics and base characters as variants be reopened. Point Québec wants the Latin Script committee to express a request to modify the report to include in the Latin GP script that the letter é (No 47 : code 00E9) as a variant of the ASCII «e» without an accent but more generally, diacritics and base characters as variants be reopened. &gt;&gt; &gt;&gt; Thank you for considering our request.</p> <p><b>Summary</b></p> <p>Point Québec wants the Latin Script committee to express a request to modify the report to include in the Latin GP script that the letter é (No 47: code 00E9) as a variant of the ASCII «e» without an accent</p> <p>Point Québec wants to inform the Latin Script committee that the GeoTld «.Québec» with the acute accent should be considered as a variant of «.Quebec» without an accent. It was not possible to successfully apply for the name with an accent in 2012, unfortunately. We strongly believe that, from a worldwide French community perspective, the letter «é» (No 47: code 00E9) is a simple variant of the ASCII «e» without an accent.</p>
Latin GP Analysis	<p><b>[1]</b> See Comment 1. Also look at <a href="#">Final Report of the GNSO New gTLD Subsequent Procedures</a>, being considered by the ICANN board, for various recommendations about variants.</p>
Conclusion	<p>No requirement to update the proposal.</p>

No. 5	Submitted by Armenian Generation Panel
Comment	<p><b>Additional variant code points [7]</b></p> <p>Armenian GP had studied Latin GP proposal and has no objection to it. Some cross-script variants defined in Latin LGR will produce more variants in Armenian LGR. Armenian GP will incorporate these variants in the updated version of Armenian LGR.</p> <p>At the same time, some of the similarity of the Armenian and Latin script (see below) is not reflected in the Latin GP proposal.</p> <p>Latin script Armenian script</p> <p>U+0071 q U+563 զ</p> <p>U+0064 d U+56A ժ</p> <p>U+0068 h U+570 հ</p> <p>U+006E n U+575 ն</p> <p>We are ready to discuss them with the Latin GP.</p>

<p>Latin GP Analysis</p>	<p>The Latin GP agrees that there are certain similarities between the Latin and the Armenian script. For this reason, it has dedicated a full section (Section D.9.3.) to compare all characters that could be considered remotely similar.</p> <p>When you take a look at the table in that section, you will find all four of your cases mentioned there.</p> <ul style="list-style-type: none"> <li>• U+0563: The Latin GP compared this character both with U+0071 and U+0105. For both comparisons it was decided that they should not be variants due to them not being similar enough.</li> <li>• U+056A: The Latin GP compared this character with U+0064, U+0257 and U+0111. It was decided that none of them should be variants due to them not being similar enough.</li> <li>• U+0570: The Latin GP compared this character with U+0068 and found that they indeed are almost indistinguishable for certain fonts. We therefore decided to include a variant relationship for this pair.</li> <li>• U+0575: The Latin GP assumed this is a typo and you actually meant U+0578.</li> <li>• U+0578: The Latin GP compared this character with U+006E and found that they indeed are almost indistinguishable for certain fonts. We therefore decided to include a variant relationship for this pair.</li> </ul>
<p>Conclusion</p>	<p>No requirement to update the proposal.</p>

<p>No. 6</p>	<p>Submitted by Bill Jouris</p>
<p>Comment</p>	<p><b>Repertoire [4]</b></p> <p>The whole point of the IDN project is to make sure that users worldwide can create domain names in their own language and their own script. That includes not just very widespread scripts, but also very local ones. Of the 450+ living languages which use the Latin script, the Latin GP has chosen to include less than half in establishing the repertoire. (The omission is made worse by the fact that the vast majority of the languages which are not included are from sub-Saharan Africa, while none of the Panel members are from that region nor speak the languages used there.)</p> <p>The criteria used were to include “official languages” plus languages which are not so designated but have at least 1 million users. However, classification as an official language depends on local politics, the resources available to the local government to conduct business in additional languages, and vagaries of colonial boundaries drawn by diplomats thousands of miles away without references to the peoples on the ground. And the threshold is both arbitrary and subject to the ever-changing populations of native speakers. Due to the criteria used, we have the ludicrous situation where a language like Hawaiian (25,000 native speakers) is included, while a language like Obolo (300,000+ native speakers) is not.</p>

The Latin GP needs to go back and spend the few weeks required to include those languages. (I recognize that this will also require some additional work to evaluate possible variants involving the additional codepoints from the additional languages. So be it.)

### **Variants [1] [8]**

One of the foundational [principles](#) for the Internationalization of Domain Names project is the

**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.

Further, the [IDN Variant TLD Implementation](#) document states that

“The variant management mechanism should “promote a good user experience”, which means that it should “avoid including variant TLDs in a manner that would create user vulnerabilities or a probability of confusion”.

But the Panel seems to have lost track of that in establishing variants.

The Latin script makes use of some 20 diacritic marks, which are used to modify one or more underlying letters. However, most languages use only a small subset of these. While the members of the Panel are, of course, familiar with all of the diacritic marks, the typical user will only be acquainted with half of them or less. This increases the potential for confusion, as it is difficult to recognize as different a diacritic which one does not even realize exists. For example, suppose the user has never encountered the Dot Above diacritic, and is expecting an Acute (or Grave) diacritic. When presented with a Domain Name which uses a Dot Above, he is unlikely to realize that it is not the Domain Name featuring an Acute that he was expecting.

Evaluation of possible variants involved the Panel members looking at pairs of glyphs side by side (a luxury that an end user would not have during normal use), while knowing that they were looking at two different glyphs. They then decided how likely it was that, in looking at a Domain Name, they would realize that the two were different.

The Panel initially took a simple majority rule approach (4 of the 7 Panel members). But the Integration Panel quickly advised the Panel that identifying variants was not a matter for a vote. What the Panel chose to do in response was to raise the bar, and only recognize as variants those that 5 of the 7 expert members found indistinguishable. What the Panel should have done at that point was to recognize that typical users have far less expertise and experience than the Panel members. And that our experts were doing a side-by-side comparison (a luxury the user would not typically have) of two glyphs that we already knew were different. And so if 3, or even just 2, members found the glyphs confusable in those circumstances, the typical naïve user looking at a domain name would also.

The criteria used has resulted in the Panel having an official position that some pairs of glyphs, which a *majority* of the members could not distinguish, nonetheless are different enough that a “reasonably careful user” would somehow magically see the difference. It is simply not possible to reconcile this with the principles above.

The Panel needs to go back and apply a more reasonable standard for variants. Fortunately, the spreadsheets used for the existing ratings are still available. Thus the revision should be quite straightforward, if tedious.

### **Underlining [2]**

Also, there is the issue of Underlining. The Panel asserts that the underlining added to domain names is interrupted by blank pixels on either side of any below-the-line diacritics, and thus a “reasonably careful user” will be able to find them. Even assuming that said user could spot diacritics that he has never seen and so doesn’t know to look for, there is another problem. While some browsers (Chrome, Firefox) do put in blank pixels, others (Internet Explorer) do not. Further, some widely used word processors (Word, PowerPoint, Excel) do not. It may be noteworthy that the software that ICANN uses in the Public.

Comment feature also does NOT insert blank pixels. Thus it seems mistaken to say that users will generally be able to see below-the-line diacritics. In fact, any codepoint involving a below-the-line diacritic which is not connected to the basic letter should be considered a variant of that letter.

### **Other**

There are also a couple of areas where the Latin GP was handed a decision from above, but which should be noted:

### **Capitals [3]**

Because domain names are strictly lower case, the various Generation Panels were directed to ignore capital letters when identifying variants. But this ignores the reality that users have decades of experience which teaches them that, in a domain name, upper and lower case letters are completely interchangeable. (The Greek GP has written about this as well.) For example, .COM would be entirely interchangeable with .com. Therefore, if a user encounters a Cyrillic TLD of .com, he will automatically assume he is seeing .com in upper case. What he will NOT do is recognize, as the IDN project apparently expects, that the third letter doesn’t look like a lower case M, and therefore this is obviously different from .com.

One rationalization for ignoring capital letters anyway is that the size of the letters will indicate that they are not capitals. Note, however, that the Root Zone LGR includes this: “In typical user interface fonts, even code points like “s” and “s” (U+0D1F) may look indistinguishable.” So, we have a type case for cross-

	<p>script variants where, in the same font size, on letter is TWICE the size of the other. So such argument cannot stand.</p> <p>The Latin GP cannot change this policy. But the IDN project should be asked to rethink it.</p> <p><b>Numerals [9]</b></p> <p>Numerals are allowed in Second Level Domain Names, but not in Top Level Domain Names. Because the GPs are focused entirely on TLDs, they necessarily ignored numerals. But there are some potential variants here. For example:</p> <ul style="list-style-type: none"> <li>Numeral One vs Latin Small Letter L</li> <li>Numeral Three vs Latin Small Letter Ezh</li> <li>Numeral Five vs Malayalam Letter Tta</li> <li>Numeral Zero vs Malayalam Letter Ttha</li> </ul>
Latin GP Analysis	<p>[1], [2], [3], [4], see Comment 1.</p> <p><b>[8], IDN Variant TLD Implementation Framework</b></p> <p>The reference to the IDN Variant TLD Implementation Framework is not applicable in the context of the making of the RZ-LGR.</p> <p><b>[9], Numerals</b></p> <p>Digits are not included in the MSR (Maximal Starting Repertoire) and are prohibited for top-level labels and are therefore deemed out of scope for analysis.</p>
Conclusion	Besides updating the proposal as per [4] (see Comment 1), no other requirement to update the proposal.

No. 7	Submitted by Khmer Script Generation Panel
Comment	<p>On behalf of Khmer GP, I would like to congratulate to Latin GP for the Proposal for Latin Script Root Zone Label Generation Rules.</p> <p>We have no technical changes suggestion for the proposal.</p>
Latin GP Analysis	The GP thanks the Khmer script GP for the comment.
Conclusion	No requirement to update the proposal.

No. 8	Submitted by Eduardo Diaz, NARALO
Comment	<p>Appendix E - Confusables - I went through this document and found out that there are many characters that can be easily be abused in domain names to harm end-users. For example Latin Small Letter U with Acute (00FA) ú - VS.- Greek Small Letter Upsilon with Tono - (03CD) - ú are almost indistinguishable.</p> <p>I recommend to not include any of the characters presented in Appendix-E until further assessment with the community is done on these. <b>[10]</b></p>
Latin GP Analysis	The Latin Small Letter U with Acute (U+00FA, ú) and Greek Small Letter Upsilon with Tonos (U+03CD, ú) are defined as variant code points. The Appendix E will be updated accordingly.
Conclusion	<a href="#">Appendix E</a> will be updated to remove the cases already defined as variants in the LGR. Appendix E will only contain the confusable cases.

No. 9	Submitted by Japanese Script Generation Panel
Comment	<p>I appreciate the opportunity to comment on the Latin Script RZ-LGR proposal.</p> <p>First of all, on behalf of Japanese Generation Panel, let me congratulate the completion of the Latin Script RZ-LGR proposal.</p> <p>As the proposal document says, the Latin Script is used by many countries, territories, cultures, languages and so on. As Chair of a Generation Panel with similar situation, although our script may be loaded with less diversity, I can imagine the difficulties the Latin Script GP has experienced in coordinating and dealing with such diversity.</p> <p>I believe the proposal developed by the Latin Script GP is more than reasonable through thoughtful consideration.</p>
Latin GP Analysis	The GP thanks the Japanese script GP for the comment.
Conclusion	No requirement to update the proposal.