

Report of Public Comments

Title:	IDN Variant TLD Program -- Draft Final Report Examining the User Experience Implications of Active Variant TLDs																				
Publication Date:	12 March 2013																				
Prepared By:	IDN Variant Project 6 Team																				
Comment & Reply Comment Period:		Important Information Links																			
Open Date:	18 Jan 2013	Announcement																			
Close Date:	1 March 2013	Public Comment Box																			
Time (UTC):	23:59 UTC	View Comments Submitted																			
Staff Contact:	Steve Sheng	Email:	steve.sheng@icann.org																		
Section I: General Overview and Next Steps																					
<p>To better understand and address the challenges surrounding the activation of variant internationalized top-level domains (IDNs), this draft final report was published for public comment.</p> <p>Based on the input received in this public forum, ICANN will produce a final report that incorporates the community feedback. The final report will be published within 30 days after the closure of public comment.</p>																					
Section II: Contributors																					
<p><i>At the time this report was prepared, a total of 3 community submissions had been posted to the Forum. The contributors, both individuals and organizations/groups, are listed below in chronological order by posting date with initials noted. To the extent that quotations are used in the foregoing narrative (Section III), such citations will reference the contributor's initials.</i></p>																					
ORGANIZATIONS AND GROUPS:																					
<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 50%;">Name</th> <th style="width: 30%;">Submitted by</th> <th style="width: 20%;">Initials</th> </tr> </thead> <tbody> <tr><td> </td><td> </td><td> </td></tr> <tr><td> </td><td> </td><td> </td></tr> </tbody> </table>				Name	Submitted by	Initials															
Name	Submitted by	Initials																			
INDIVIDUALS:																					
<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 45%;">Name</th> <th style="width: 30%;">Affiliation (if provided)</th> <th style="width: 25%;">Initials</th> </tr> </thead> <tbody> <tr> <td>Chris Dillon</td> <td> </td> <td>CD</td> </tr> <tr> <td>Abdulaziz Al-Zoman</td> <td>SaudiNIC</td> <td>AA</td> </tr> <tr> <td>John C. Klensin</td> <td> </td> <td>JCK</td> </tr> <tr> <td>Kal Feher</td> <td> </td> <td>KF</td> </tr> <tr> <td>Joseph Yee</td> <td> </td> <td>JY</td> </tr> </tbody> </table>				Name	Affiliation (if provided)	Initials	Chris Dillon		CD	Abdulaziz Al-Zoman	SaudiNIC	AA	John C. Klensin		JCK	Kal Feher		KF	Joseph Yee		JY
Name	Affiliation (if provided)	Initials																			
Chris Dillon		CD																			
Abdulaziz Al-Zoman	SaudiNIC	AA																			
John C. Klensin		JCK																			
Kal Feher		KF																			
Joseph Yee		JY																			
Section III: Summary of Comments																					
<p><i>General Disclaimer: This section is intended to broadly and comprehensively summarize the comments</i></p>																					

submitted to this Forum, but not to address every specific position stated by each contributor. Staff recommends that readers interested in specific aspects of any of the summarized comments, or the full context of others, refer directly to the specific contributions at the link referenced above (View Comments Submitted).

GENERAL COMMENTS

(CD)

Chris praised the Chinese model of using a well-formed table to facilitating automatic delegation of variants and stressed that this model should be replicated across other languages/scripts. He noted that the automatic delegation of three variants (one simplified, one traditional, and one user choice) may go contrary to this report recommendation that variants not be approved automatically; however, he noted that this process does promote consistency.

He also praised, regarding the Chinese model, consistency in the application of the LGR across the DNS tree.

(AA)

The report should limit discussions and recommendations to apply *only* to IDN TLDs. For example, the report makes policy recommendations concerning ccTLDs, which falls outside of the stated scope of the report. In addition, the report makes recommendations that impact other levels of the DNS.

The existing DNS system (without variants) does allow registries a degree of freedom (e.g. number of DNS servers, domain name life cycle, prices, renewals) that has not deferred registrants and the technical community from using and accepting these differences.

The report has succeeded in frightening readers of the many potential dangers of activated variants. However, the report does not note the dangers of variants if they are *not* supported and/or activated. The need for variants is crucial.

It could be understood from the report that the Arab states that got Arabic IDN TLDs have violated or worked against the recommendations outlined in RFC 5564 that was developed and endorsed by the Arab League. This is of course not the case. The RFC 5564 provides the general and basic rules that were developed to cover some linguistic issues that are related to the development of the IDN protocols. It was agreed among the Arab team (who developed the RFC) that local registries could amend it with their local needs.

The accepted characters for writing an IDN label should be based on languages rather than script tables. Many of the risks outlined in this report can be avoided through adopting a language table for each registry. Users write and read a language and not a script.

It is not necessary for the end user to know or to be able to type all of the variants. Users need only to

be able to write one form of the variants to reach a website. It is the responsibility of the registry to make this transparent to the user.

(JCK)

John praised the report for clarity and relative austerity but also noted that a number of key points were overlooked and that the report suffered from difficulties regarding IDN scope, concepts, and general aspects of the systems of which IDNs and variants are a part. Major issues include:

1. This report, like the LGR report, goes behind variant issues to addressing a number of general IDNs issues including the character repertoire that should be allowed across the DNS tree. This expansion of scope is desirable and it is also desirable that ICANN ensure that the topics and recommendations in this report be reviewed in a broader and more appropriate context. In addition, a number of recommendations for specific actions are relevant for IDNs, regardless if any variants are activated. It would have been helpful if these recommendations were highlighted.
2. This report takes a DNS-centric approach to addressing variant issues. In fact, there may be other options that pose less risk to the DNS.
3. This report notes that the LGR is just one method of identifying variants. In fact, we've been dealing with variant issues for some time, such as when companies register multiple domains and rely on redirection. So we've survived reasonably well for more than two decades without mechanisms for identifying/bundling/provisioning labels. This report assumes that these mechanisms are required but does not make a persuasive case that these systems and mechanisms will add enough marginal value to security, stability, or usability to justify the considerable costs involved. Second, if bundling is required for IDN variants, raises the question that retroactive bundling of related names (e.g., ti.com and texasinstruments.com) may also be appropriate.
4. The issue of FQDNs is not fully addressed (though implied) in this report. If discipline is not enforced regarding the mixing of variant labels across the FQDN, we are looking at an administrative nightmare.
5. If a limit is placed on the number of delegated and/or activated variants, users cannot easily predict which string will work, which naturally goes against the predictability principle. Also, the report argues, for good reason, that variants should not be allocated, much less delegated, unless they are important for a positive user experience. However, if a variant does not meet this test, what are the implications on the primary label? If this leads to exceptions, predictability may further suffer.
6. Delegated variants are a hack to compensate for user inability to guess exactly how a string is written or for difficulty within input. So if we assume that end users must be fully educated about variants, then variant delegation is inherently not necessary. Conversely, if software could query the variant bundle database in real time, there would not be a need for DNS bundling. Granted, this model is not trivial to deploy, but it would be preferable.
7. This report gives mixed signals regarding decisions on how variants are assigned a state (blocked, delegated, activated).
8. The characterization of the "technical community" is inaccurate. The community values a high-

quality and predictable user experience and is not solely concerned with security and stability.

9. **High-Level Conclusion:** This is all just too complicated and, as this report (even without the additional issues identified in this review) clearly shows, creates significant additional risks and costs without being able to offer the payoff of a predictable and consistent experience that will meet the needs of end users. Adjustment of expectations to match reality and considerable simplification appear to be in order. Most obviously, that might be to simplify and redefine the LGR activity to an effort to define a repertoire for the root and make general repertoire recommendations for SLDs (and perhaps below), to ask the team that produced the present report to redo it to focus exclusively on root IDN issues rather than IDN variant ones, and to deal with requests for related names in the root zone — especially those not associated with historical ccTLDs — on an exception basis with the applicant having to demonstrate to the community that the relationship is important enough (and will solve a specific problem that can be identified) to justify the various “challenges” and other difficulties outlined in this report. The LGR report creates a process that will make it very difficult to get global agreement on labels that can be activated. The recommendations of the present report set a very high threshold for actual activation of more than one variant in a bundle and justify those recommendations by identifying a wide range of risks and problems. So, net and with the understanding that other options might be possible, ICANN seems to have two options:

- Follow roughly the outline above with the understanding that it will result in very few variants being delegated (other than what would be the primary name in a bundle) although no one would be prevented from applying for additional names subject to the usual reviews and challenge procedures. While I believe it would be better to see if a simplified version of the LGR process (perhaps even skipping the generation panels for the scripts for which we have significant experience and confidence) can produce a satisfactory root repertoire and to get a revised and rethought version of this report as a set of guidelines about issue to be considered, ICANN could, in principle, solicit the required additional information and start considering the variant requests that are implicit in the applications now in the queue at its convenience.
- Follow the general model of the LGR report, including multiple panels and rules for generating and establishing the status of character variants; create mechanisms (so far undefined and undiscussed) for generating or otherwise establishing candidate variants that cannot be produced from character-based generation rules; and implement the many recommendations of the present report (perhaps informed by the comments in this review). With the understanding that the two or three years (at least) those processes would be likely to take would increase the risks from the contradictions discussed above, expect to end up with few, if any, activated variants that would not be activated under the above model.

(KF)

Requiring registries to implement the same LGRs as those used in the root, for their relevant script, is onerous technically and not justified by any reasoning within the report.

While the use of common LGRs is advantageous to the consumers (and those that write software for them) of IDN domain names, this advantage is already well understood by registries that have deployed IDNs. There is no discussion within the paper as to why, when faced with these advantages, some registries have opted not to use the same LGRs, while others have chosen to use LGRs already implemented.

If the option of voluntarily using the same LGR exists, as it does today, then there should be a strong reason to require through additional regulation that TLDs use common LGRs. Such a reason has not been put forward in the report.

SPECIFIC COMMENTS

(AA)

Executive Summary

There are no significant differences within the Arabic language communities regarding how IDN variants are managed. All are based on RFC 5564. Local registries may amend it according to their local needs. The differences reside across languages that share the Arabic script.

Section 1.4

Report states that different implementations the life cycles of variants could cause confusion. However, each registry already has its own implementation of life cycle with no use confusion.

Report notes that TLDs are defined at the script level while SLDs may be defined at the language level, resulting in different implementations. We recommend using language tables to determine accepted characters instead of using a script table.

The report notes that the user experience with IDN TLDs will vary by the technology platforms being used. However, this issue is not unique to variants. Case sensitivity with ASCII passwords has long been an issue.

Section 2.1.1

The report notes differences in IDN ccTLD implementations (number of variants activated) across Arabic registries. These differences are normal and will not cause any confusion provided that variants that not activated remain blocked. These differences are determined by different business needs of each registry.

The solution used of SaudiNIC has many advantages and addresses many of the problems outlined in this report. It is not very complex and it minimizes the number of variants activated. SaudiNIC has two sets of variants – at the language level (similar to other Arabic countries) and at the script level. Currently, up to 10 variants can be activated at the language level. Script-level variants are ready but have not be rolled out to registrants until IDN TLD variants are approved by ICANN.

Section 2.1.5

Report notes that SaudiNIC intends to implement positional variants using ZWNJ. However, SaudiNIC has a very strong position against using ZWNJ in domain names.

- Our position is well known to some of the authors! And documented in the Arabic Script Case Study
- report (<http://www.icann.org/en/topics/new-gtlds/arabic-vip-issues-report-07oct11-en.pdf>)
- Allowing ZWNJ in domain names is a very serious problem and threat to our users.
- ZWNJ causes some security, stability, usability, and reachability problems, and hence mistrust of IDNs.
- Here are some of the arguments that the Arabic Script Case Study report outlined against allowing ZWNJ
 - At Script level, the ZWNJ is considered by UNICODE to be an invisible join control character and listed in the "Unicode Security Considerations" document, which warns that incorrect usage can expose programs or systems to possible security attacks. This is especially relevant for IDNs.
 - The ZWNJ in some cases is not visible to all users (e.g., U+0637, U+0638, U+069F, U+06BE, and U+06FF). A comprehensive analysis of Unicode Arabic Script Code Charts is needed to find any additional cases. This process should be repeated as the Unicode gets updated.
 - The ZWNJ concept and behavior are not known to many Arabic script users, who do not use it or know how to type it.
 - ZWNJ is not conveniently available on the keyboard, where typing it requires multiple simultaneous key-presses, which is complicated for users. ZWNJ is also inconsistently placed on keyboards across various operating systems. In addition, it is not available on many keyboards, making it difficult for people to use ZWNJ when, for example, they travel.
 - The users may not be able to type a domain name as they may think it is a <space> not ZWNJ, which may lead to reachability and usability problems, and therefore, mistrust of IDNs.
 - Based on the ZWNJ Contextual Rule (RFC 5892 Appendix A.1) for handling CONTEXTJ labels under the current IDNA2008, the Rule implementation does not totally resolve the non-visibility problem particularly in some cases as discussed above (e.g., U+0637, U+0638, U+069F, U+06BE, and U+06FF).
 - It is not one with the general category of {Ll, Ll, Lm, Mn}, as per the requirement defined by the gTLD Applicant Guidebook (v 2011-09-19, Module 2, page 2-13, Section 2.2.1.3.2, Part II, Item 2.1.3.).
 - Root policy should be more conservative than labels for other levels.
 - Use of ZWNJ may cause additional bidirectional display issues.
 - Hyphen can be used instead of ZWNJ to break a string into ligatures

Section 5

A number of these "challenges" listed here are not caused by variants. Some of the challenges already

exist in current (IDN or ASCII) domain names. For example, an entity that registers multiple domain names to protect against domain squatting would face some of the same challenges. Also, search engine optimization issues are not unique to variants.

Section 5.1

This section frightens readers away from supporting variants. Furthermore, it does not highlight the risks (problems and security issues) if variants are not supported or activated. The need for variants is crucial.

Report notes that “Variants of IDNs are a new concept.” Yet variants are already supported regarding upper/lowercase ASCII domains. Furthermore, even if variants are a new concept, it should be encouraged and fully supported.

Section 5.1.1

IANA is the repository for language tables used by registries.

Section 5.1.2

Reiterates the need to determine accepted code points by language and not script.

Sections 5.1.5 – 5.1.14

Notes that these examples are hypothetical and, in some cases, are not specifically variant issues. Some of these examples already exist with ASCII domains.

Section 5.2.4

The issue of not supporting the user’s language via registration interface is a business issue and will be solved by competition between registrars.

Section 5.2.6

It is not correct that variants are not part of the registration data. SaudiNIC displays the originally registered domain name when you search for one of the variants regardless if it is activated or blocked.

Section 6

You should not recommend activating a minimum number of variants. Instead, the recommendation should be based on the community’s need and user protection. Minimizing is not a goal if there is a user need. The numbers of activations are not the issue here.

Section 6.1.1 and 6.1.3

It is unacceptable to require a separate approval process for the approval of a variant TLD. Once the initial string has been approved, we need a light approval process. Why GAC? Public?

Section 6.1.4

It is not practical to require consistent rules across the DNS tree if the root is more conservatively

managed than the second level.

(JCK)

Section 1

Would be useful to know how many of the IDN fast track IDN ccTLDs are associated with variants. Also valuable to know the number of gTLD IDNs being applied for.

Section 1.2

Constraining the report to variant labels generated by the LGR is a grave disservice to the community.

“There are ongoing discussions” is not true or should not be.

Section 1.4

Script-based distinctions in the root are feasible for some scripts and not for others.

Section 2.1.1

Last bullet “five variants at a time” is ambiguous.

Section 2.1.4

Would be helpful to note a Latin string example to illustrate issue regarding multiplicative explosions of FQDNs.

Section 2.1.5

Provide more information on ccTLDs and variants in the third level (see TWNIC).

Section 3.3

It’s worth noting that leading users to expect a degree of security may lead to a false sense of security

Sections 3.4, 3.6, 3.7

Regarding the closely related principles of Predictability and Consistency, we have a dual role that we need to balance. Part of it is to design our systems to accommodate the expectations of the users — to provide them with predictable and consistent experiences— and the report seems to be designed around that part. But the other part is to help those users understand what they should expect and predict. We need to balance the two, not assume that the entire burden can or should be shifted entirely in one direction or the other.

Section 4.1

Include the use of identifiers that use DNS elements as part of this list of functions.

Section 4.2: Registrants and Registrars

To mitigate the risk of registrants who seek to use variants to attack end users , an ICANN mechanism that identified related names but did not immediately reserve or withhold them would probably be of great value to this group of would-be registrants and the registrars who support them.

Section 5.1.6

Statements like “user may not be able to input all of the variants” are, with rare exceptions, probably false. It would be more accurate —and would put a better perspective on the issues— to say that users would not find it acceptably convenient or would need to learn special skills in order to do so.

Section 5.1.10

This section is confusing. Users don’t typically search for an FQDN.

Section 5.2.1 and elsewhere

Should explain why the notion of a “primary” name is important. Unless there are aliases involved, the DNS cannot know the difference. They may be convenient for some management and provisioning models and databases, but the topic needs more explanation before one worries extensively about, e.g., redesignating the primary.

Section 5.2.7, 5.2.8

Until and unless ICANN succeeds in adopting enforceable policies that make registrations that are intended to take advantage of confusion (or even referral commissions based on typographical errors or the like) invalid, reports like this one need to consider those cases as legitimate as the cases of those who want to protect their trademarks.

Section 5.3.3

This section or a separate one should discuss the very important issue of certificate revocation, especially the need to do so on an emergency basis. The existing text examines the question of obtaining the certificates, which may be burdensome but is easy in principle — all of the needed names are known. But revocation of a cluster of certificates linked to a variant cluster requires being able to retrieve all of the names and credentials associated with that cluster and take action on them.

Section 6.1.7

Despite the comments about training materials and user education above, I am concerned that this section calls for a major expansion in ICANN’s scope and mission, something that should not be taken on simply on the basis of a few paragraphs on page 45 of a 56 page report.

Section 6.4.1

This is one of the issues that is necessitated by IDNs with or without variants, but, if the community is serious about IDNs, especially at the TLD and SLD levels, it is no longer acceptable for generally-available and heavily used basic DNS management and diagnostic tools to be IDNA-unaware and to operate only in terms of A-labels as a result. Such tools are much less complicated than the other ones listed in this section and include, e.g., basic command line DNS name query and reporting tools. I believe a recommendation should be added to this section for ICANN to facilitate the development or upgrading of those tools.

Section 6.4.3

I find this section confusing.

(KF)

Section 6

I suggest that recommendation 6 have the word "require" replaced with "request". Full suggested recommendation:

- ICANN should request IDN TLD registries with variants to apply the relevant (script) subset of the root zone LGR and state life cycle for variants across second-level domain labels. Deviations should be justified.

(JY)

The recommendations, especially recommendations for ICANN& registries & registrars, are clear and specific. However, many of these recommendations are short term focused and registration focused. Majority of challenges fall into technical communities. And for these areas of challenges, the recommendations for technical communities are not as specific and clear as recommendations for ICANN, registries, and registrars; the traceability matrix should be extended to include the recommendations; The process should address registration process as well as longer term operations around variants.

Section IV: Analysis of Comments

General Disclaimer: This section is intended to provide an analysis and evaluation of the comments received along with explanations regarding the basis for any recommendations provided within the analysis.

We thank the community for the in-depth and constructive feedback. The comments have helped us improve the report and our understanding of issues yet to be resolved.

We recognize that the report notes challenges with IDN variants across the DNS tree (as opposed to only at the top level) and that may appear to be beyond the scope of the report. However, when one views a URL through the eyes of a typical web user (which is the goal of the report), distinctions between, say, the TLD and SLD become irrelevant. As the focus of this report was first and foremost to analyze potential usability issues, we were compelled to consider all aspects of the DNS tree.

We give analysis on general comments presented in these public comments and carefully consider the detailed comments to update the relevant portions of the report. However, due to specific nature of detailed comments provided, their analysis is not discussed here.

(CD)

Authors thank the reviewer for pointing to the relevant practice in CDN community.

Three variants (just about as low a number as it gets) are allowed; two automatically from a single table used across the script, and one user-defined.

Note that the "automatic" bit of this goes against 6.1.1.1. ("The approval of a variant TLD must not be automatic, but initiated upon the request of an applicant ...) but does promote consistency.

Limit of Chinese variants should be reduced to two (TC and SC) to make it consistent. The third user defined variant actually introduces inconsistency as it can be arbitrarily different across registrants (applicant in case of TLDs) and therefore unpredictable. The other mechanism to introduce consistency is to activate all labels (which is practically very difficult as it creates too many possibilities; if 10 for a TLD label, then 100 for SLD.TLD and 1000 for 3LD.SLD.TLD).

Consistency is desired but is still elusive for other reasons. User sees a fully qualified domain name. For user experience to be totally consistent, it means that TLDs should be consistently activated + SLDs should be consistently activated (+ 3LDs should be consistently activated + ...). Further, just activating a domain label does not necessarily lead to content (for the end user). This is not straight forward to enforce in the chain which involves ICANN-Registry-Registrant all behaving the same way.

Further, there are limits as well to how broadly the model can be applied to other scripts, because though in Chinese the variants are normally visually distinct, for many other scripts the variants are visually identical or similar.

The primary label and all the variants form an indivisible set. This should address many of the issues in 5.1. Internationally, it may be desirable for the set to be referred to using the A-Label of the primary label in case old software cannot handle U-Labels.

Yes, the report also supports the concept of primary label and formation of variant labels forming an indivisible set (for allocation). Please note that though Chinese script defines a preferred label, the choice of primary label is user dependent in other scripts, for example, for the same variant label set for Arabic script, variant label may be different for Arabic language applicant vs. Persian language applicant. Further, not all variants in the set are activated. Thus, issues in 5.1 still apply.

As far as I can see there is no intention to use different LGRs for generic and country code 2LDs and TLDs. This approach reduces the work necessary in 6.1.6.1. for ICANN to evaluate second level LGRs. If it is ever necessary to have separate LGRs at the second level for a script, a table used at all levels with a short list of characters such as the hyphen also available at the second and below would be better than maintaining two tables.

Such consistency could help, but may not be enforceable for second and lower levels, unless community itself agrees to it. This is because LGRs for TLD are to be defined at script level but ccTLDs or gTLDs may choose to serve a language community so may choose to use a subset of the top-level

LGR (in addition to the code-points for hyphen, digits, etc.). Thus, LGR for TLD and SLD may be different.

A well-made table could have columns for e.g. Traditional and Simplified Chinese code points variants (or e.g. Arabic and Urdu variants) which could be used to do a lot of the generation of variant labels automatically. This could be useful if, for example, an applicant forgot to apply for a Traditional Chinese variant code point.

Recommendation 6.1.2 and 6.2.2 ask to provide the full set of variants, as rightly pointed out that users may need the information for various reasons. However, no specific format for LGR is recommended, as that is being finalized by another sub-project within the IDN Variant Issues Project.

(AA)

Authors greatly appreciate the input and insights from the perspective of a ccTLD registry.

One major concern that I have is that the report discussions and hence some of its recommendation are OUT OF SCOPE of the project goals and objectives. Regardless of the unquestionable decent intention and the goal to cover other areas beyond TLD level with some nonbinding recommendations they are, nevertheless, considered out of scope.

We recognize that some of our recommendations impact ccTLDs and we are aware that such domains fall under the control of their respective governments. However, our focus is first and foremost on end user experience, who may not make distinctions between IDN TLDs and IDN ccTLDs and, in fact, may expect variants to behave similarly across these domains. By “casting a wide net” our goal was to make all impacted registries aware of the many complex and interrelated usability challenges related to IDN variants. The suggestions in the report are not binding on ccTLDs but are recommended for consideration and adoption.

Another observation is that the report and its recommendations diminish the language community contributions and authorities over IDN issues while they are closer to the user’s needs. But, on the contrary, it expands and broadens ICANN responsibilities beyond its mandate and roles while ICANN is far away from user’s needs and requirements.

The report upholds the final dependence on the linguistic community. LGR development is strongly based on community input, which is initiated by a linguistic community panel. Community focus in LGR is further emphasized in Recommendation 6.1.3. Applications for each variant label also needs justification coming from community (e.g. 6.1.1 Point 1- (iii): “the need for the variant (e.g., motivated by *linguistic*, security, usability and/or other considerations”). Though one could tell by looking at a label which script it belongs to (in some cases even that may be an issue, e.g. for some labels in Latin and Cyrillic scripts), it is not easily possible to determine the language of a label. For example, the word کتاب (kitaab, “book”) is a word of many languages using Arabic script.

The existing DNS system (without variants) provides some freedom to registries in

defining many directions in managing their registries (e.g., database fields, number and type of contacts, number of dns servers, domain name life cycle, prices, renewals, ... etc). This has not deferred registrants and technical community from using and accepting these differences.

These possibilities have existed for labels without variants. The report does not limit these possibilities for variants, but it does indicate that in some (user experience relevant) contexts, variants should behave similarly (without placing restrictions on implementation models) as they are expected to represent the same label.

Section 5 listed several challenges related to active variant TLDs. Even though they appear accurate and true from the first look, but they are not caused by variants. Some of the challenges already exist in the current (IDN or ASCII) domain names. For example, if an entity has multiple domain names due to protection against phishing or domain squatting would face some of the same challenges outlined in section 5. Some of the challenges outlined in Section 5, e.g., search engine optimization, is complex and dynamic and is not related to only variants.

Authors agree that some of the challenges identified may not be exclusively applicable to variants. However, these challenges are still relevant and in some cases further exacerbated by variants. For example, phishing is applicable to similar strings, but as variants may be especially susceptible to it as they are visually identical or similar in some scripts, there is need for defining a variant set. Variants will have also have impact on search engine result set ranking and optimization, which may have impact on user experience.

The report (particularly section 5) has succeeded beyond any doubt to frighten and horrify the reader from variants. But it fails to illustrate the problems and security issues if variants are not supported or activated. The need for variants is quite important and crucial. We should not be frighten or consider it as a complex just because it is new or because we are not familiar with.

The report assumes that the case of variants has already been made in the earlier phase of the Variant Issues Project. Thus the current report focuses on implications of activating them. However, the point is well made and authors will include specific text in the section on Scope in the report to clarify this point explicitly.

It could be understood from the report that the Arab states that got Arabic IDN TLDs have violated or worked against the recommendations outlined in RFC 5564 that was developed and endorsed by the Arab League. This is of course not the case. The RFC 5564 provides the general and basic rules that were developed to cover some linguistic issues that are related to the development of the IDN protocols. It was agreed among the Arab team (who developed the RFC) that local registries could amend it with their local needs.

The report will be updated to incorporate this perspective.

In many occasions we recommended determining the accepted characters (code points) for writing an IDN label should be based on languages rather than script table. Many of the confusions and risks outlined in this report can be avoided through adopting language table for each registry. Users write and read a language and not a script.

This comment is more relevant to IDN VIP sub-project 2.1, which focuses on the the process of developing label generation tables. The motivation for script table is discussed earlier. A particular user will certainly be looking at a label from the perspective of a language. However, due to the global nature of the internet it is not possible to predict the language of the user who is accessing a label. Users of many different languages may access the same label. Thus a script level analysis has to be undertaken.

It is not necessary for the end user to know or to be able to type all the variants. He/she needs to be able to write one form of the variants (using his/her keyboard) and reach the ultimate resources (e.g., web site). It is the responsibility of the registry to make this happen transparent to the user.

Authors agree that ultimately for the best end-user experience, variants should be “transparent to the user”. To make this happen at least the registrants, registries and registrars (also user roles in the current study) need to be aware and be able to input and output variants explicitly, as this may have operational and financial implications.

Authors thank for the detailed specific comments. They are considered in final revision of the report. Detailed comments specific to Saudi-NIC are noted below.

- *SaudiNIC has 2 sets of variants*
 - o At the language level (similar to other Arabic countries)*
 - o At the script level*
- *Currently we provide variants at the language level (up to 10 variants can be activated) and this was provided since we provided IDN registration.*
- *The script-level variants is ready but not rolled out to the registrant until we get the IDN TLD variant from ICANN.*

Authors have noted that Saudi-NIC is offering variants and will update the report accordingly.

SaudiNIC has a very strong position against using ZWNJ in domain names.

This is addressed in the final version of the report.

(JCK)

The comments presented are both valuable and thought provoking. Authors will work to resolve the specific comments regarding confusing sentences, omissions, or inaccuracies.

The general comments are discussed here.

... this report is not limited to variant issues but instead addresses a large number of issues with IDNs generally, including the character repertoires that should be allowed at the top and lower levels, and even some topics that are applicable to more traditional all ASCII domain names. From my point of view, that expansion of scope beyond what the community probably anticipated is actually very desirable because it permits looking at the systems involved rather than isolated corners of them that cannot be put together with the whole.

Authors appreciate the support of the reviewer in this context. The reason for going into other levels has been, as noted earlier in this document, the focus on how activation of variants at TLD level impacts the user experience. This context of user experience reaches further than just the TLD level. Scope of cause is limited to TLD, but the report considers the effect in the larger context.

As noted below, there are some areas in which I do not believe the process went far enough even with the scope expansion.

Suggestions in the context will be duly considered.

However, it is possible, indeed likely, that some groups of stakeholders in the broader community have believed that they can ignore this work because it is strictly about “IDN variants” and they have concluded that they need not be concerned with that topic. It would be therefore be desirable for ICANN to ensure that there is broad understanding of the topics and recommendations addressed by this report so they can be reviewed in a broader and more appropriate context.

Authors have passed this information to ICANN staff.

Equally important, there are a number of recommendations in the report for specific actions or activities which, if they are needed at all, would be needed for IDNs even if no variants were ever allocated or activated. It would have been much more helpful if the report identified those so that the reader could understand what, as its title suggests, the actual implications and effects of active variant TLDs are given IDN TLDs (or new TLDs generally, see item 3 below), rather than lumping everything the writers identified as useful together independent of specific cause or motivation.

The section is currently structured in a way that the recommendations to different stakeholders including ICANN, Registries, Registrars and Technical Community are grouped separately. Authors will consider how further re-organization is possible (especially vis-à-vis IDN and IDN variant) within these sets of recommendations to facilitate the reader.

The report makes the very fundamental assumption that the best (or only) way to deal with the

“linguistic” and user expectation issues it identifies is by some Domain Name System mechanism identified with the term “variant”. That is not the case and, interestingly, the report provides foundations for some of the other options. By assuming that “variants” are the preferred or only solution to many of the expectations outlined in the report, the report tends to lead the ICANN community into a syndrome that is often described in such metaphorical terms as “if all you have is a hammer, everything looks like a nail”. However, especially if ICANN were to engage in some of the user education and applications support activities recommended in the report, there may be other options that pose less risk to the DNS and to Internet stability and security than trying to use DNS mechanisms to address all such issues.

There could be alternate possibilities to variants, e.g. not have them at all. However, exploring the impact of these alternate options is not in the scope of current work. The report is focused on the impact of activating variants at TLD level on the user experience. Authors will make it clearer in the section on Scope in the final version of the report

How the variants are implemented in the DNS is also not addressed in the report. Separate work needs to be undertaken to address this issue. This is indeed called for in Recommendation 6.1.11 Point 1.

The report correctly notes that relationship among labels that might reasonably be considered to be “variants” can arise from many sources of which the character-based variants that are the subject of label generation rule work are only one, one that is very important for some scripts and not for others. It then largely ignores those other cases, focusing almost exclusively on LGR produced variants.

We have limited the scope of this report to variants identified by means of the LGR as the latter is based on what is agreed by the community following the six script case studies and the integrated issues report. In our analysis we have considered other “variant” mechanisms currently employed, where sony.com would take the users to an English page vs. sony.jp would take the users to a Japanese page. Such mechanisms are relevant in the context of Chinese script as well, with users being directed to content in TC or SC based on the domain name variant. Variants which are visually identical or similar are also taken into account. We even recognize that no two individuals may have the same understanding of what a variant is, and, even if they did, may have different views on the value of a given variant.

However, the LGR gives us a baseline to focus the work. We add this discussion in the report.

It seems to me that there are two important lessons from that history, neither of which is on the radar of the report. First, we have survived reasonably well for more than two decades without specific mechanisms for bundling labels, ... That has been managed by allowing registrants to apply for (and pay for) the names they want, defending those names by preemptive registrations, challenge and objection mechanisms, and dispute resolution procedures or legal action when needed.

Yes. This indeed has worked for ASCII domain names. However, this has been a very small set of characters which is almost universally understood by users, courts of law and other stakeholders involved. This may not be the case as the repertoire expands considerably (though IDNs have been around for a few years now at lower levels, but in practice these have not been taken up by the user community for reasons which are beyond the scope of this work). These characters are best understood by the communities which use them. Thus, managing these repertoires by preemptively engaging the communities greatly assists the users and helps prevent the dispute process. We will find a solution through existing mechanisms even if variants are not formalized through the LGR (perhaps), but that may be a more difficult path for the users (including registrants) as there will be many more disputes due to many more character sets and they will be less likely to be addressed as these character sets are not universally understood.

The report assumes that special mechanisms for variations are required; it does not make a persuasive case that the systems and mechanisms it recommends in each of its many categories will add enough marginal value to security, stability, or usability to justify the considerable costs in resources and changed procedures that those recommendations imply. My own guess is that the answer may turn out to be different for different recommendations, with some being justified and others not. Differentiating among them requires a more nuanced examination than the report attempts about what problems are being solved and which ones would cease to exist or become minimal under different scenarios.

These aspects have also been discussed internally, especially in the context of some specific recommendations. The suggestion is noted and authors will try to address this by adding relevant discussion in the report.

Second, the arguments in the report (some only implicit) for special handling of (label) variants suggest that, if bundling techniques are necessary for IDN variants or character-based ones, then they are probably desirable, retroactively, for some or all of the historical registrations of related names in the same zone. The fact that we did not know for several decades that we had a problem that needed a solution does not force a conclusion that there was actually no such problem. However, if that is correct and there is as much reason to manage “ti.com”, “texasinstruments.com”, and perhaps “national.com” together as there is to manage, e.g., a Simplified Chinese and Traditional Chinese pair of labels together, then much of the discussion in the report is not tied to IDNs at all. In addition, examples of related names in separate zones — inevitable if variants are allocated in the root— should cause ICANN to think about whether cross-zone bundling (and perhaps blocking) is appropriate

As discussed, defining variant sets is a way to manage labels. There are other alternatives possible. Similarly, other (arbitrary) sets of labels (like variant labels) may also use the mechanism being defined for variant sets. Such innovation is certainly possible.

While DNS registration and similar activities operate on labels, many uses of the DNS,

including URLs, email addresses, and most DNS-based security identifiers, depend on fullyqualified domain names. The report does not make the distinction clearly even though it is possible to deduce what is intended from the context of most sections. For the purposes of the report, the difference is important for at least one major context: with most character variant scenarios that involve actual delegation of more than one variant in a bundle, the number of variant labels within a single zone will be bounded by the bundle generated by the LGR or equivalent mechanism and whatever limits are set on the maximum number of labels to be delegated, but the number of FQDNs that must be managed —and, to the extent to which the end user actually needs to understand variant relationships, understood by that user— is multiplicative with the number of labels in the FQDN that need variant treatment.

The multiplicative effect is discussed in the section which considers existing SLD practices (Section 2.1.2). The context is assumed, but authors will clarify this in the recommendations, where necessary, for the readers. Further, restricting second level to same rules applied at top level (e.g. Chinese SLD to TC for TLD in TC) may help restrict the possible variants. Authors will discuss this possibility further.

...Consequently, from that particular user experience perspective, either all plausible variants in a bundle should be placed in the DNS or at most one of them should be. Anything else leads to a bad experience. The report argues convincingly that delegating all plausible variants is a bad idea from both user experience and security, stability, and other perspectives, so we have a contradiction.

Authors had very early on realized that the usability principles identified for good user experience cannot all be fulfilled simultaneously, so a compromise is needed. The question the authors have had to address is where the right balance between the two extremes is: exactly-one label to all labels activated. Thus, the solution proposed should not be viewed as a contradiction. The current solution proposed is biased towards minimality (not excluding the possibility of more or even all labels activated, under the conditions stipulated). This is a conservative approach (motivated by manageability and security). Exactly one label may not give a good user experience, especially in cases where two labels are visually identical (e.g. if the user types the example discussed above: کتاب which has two visually identical labels and find what it is typing is not resolving and cannot figure out why? In such cases more than one label could give better user experience).

... However, if failure to allocate the (additional) variant has that negative an impact on the primary name, then the primary name should not be delegated unless the additional variant(s) can be allocated and delegated at the same time: if the primary name is adequately usable without the additional variants, then the variants are not necessary. Turning that into a policy guideline would be fully consistent with the report as I read it. However, it implies that any application that is now in the queue that requests that a variant be delegated is in a difficult position: either the variant is unimportant, in which case it should not be delegated (ever) or the variant is important, in which case the requested primary label should be withheld until ICANN is ready to consider allocating and delegating variants. One could go ahead and delegate the primary string on an exception basis, but the contradiction lies in the application

itself (and willingness to accept one delegation without the other(s)), not in the system. And every exception or special case is going to make the overall user experience less predictable.

Authors agree that labels and their variants should be delegated simultaneously for a good user experience. But there are practical considerations (where variant allocation process is still being defined but applications for labels have been acquired and are being assessed for delegation).

Several places in the report suggest that end users will need to understand, be educated about, and adapt to variants and that the applications that support those users will need to be modified to be variant-sensitive. From one perspective (and to be a bit more blunt than the report), delegated variants are a DNS hack to compensate for end user inability to guess exactly how a string is written or for difficulties in entering some forms but not others. If the users or their application software have to understand that variants are special, how a bundle is composed, etc., then the hack fails and the variant mechanism doesn't accomplish anything very useful.

Though this is largely true for end users who want to use the IDN variants. However, if they need to register labels they (as potential registrants) would need to know the entire set of variants possible to decide which and how many variant labels to apply for activation. We will clarify in this further in the report.

Conversely, assume that the introduction of top-level variants really does require ICANN (and TLD registries and registrars) to mount massive end-user educational efforts, that ICANN (for the root) and TLD registries (for those zones) keep "bundle" databases available for real-time access and searchable by any of the names in the bundle, and that end-user interface software systems be modified to be variant-aware. If those conditions hold, we should be moving a further step beyond "variants in the DNS are the answer, almost independent of the question". For example, if an application could query the bundle database at lookup time and use it to determine which member of the variant set should actually be used in a DNS query, then a very large fraction of the problems mentioned in the report would simply disappear: there would be no issues associated with equivalent names in the DNS because there would be only one name in the DNS and there would be no issues associated with the user not being able to access the right variant (e.g., because it was not delegated) because all of the variants in the bundle could be considered without adverse consequences for the DNS. Of course, that approach would not be trivial: a new database with performance, robustness, and security/integrity features at least equivalent to the DNS would need to be designed and deployed. But it would provide a far better and more predictable user experience than DNS-based variants, would isolate tasks in ways that would make many things easier for registrants, registrars, and registries than what is recommended in the report, and might not take significant longer to deploy than the systems the report calls for (and far less time than trying to modify the DNS to support, e.g., a modified query model or even new required RR types).

Authors suggest that at this time, "The repository must be accessible at least for informational

use” (see Recommendation 6.1.2 for further details). Developing a database for operational use is a discussion relevant to how the variants are implemented. The report considers the discussion around implementation of variants out of its scope. However, this discussion is important and necessary to be taken up by ICANN separately, as suggested in Recommendation 6.1.11, especially Point 1.

This report claims to be agnostic about the decisions as to how a variant that is identified and made part of a bundle is assigned to a state (see Section 6.1.4 and the Integrated Issues Report (IIR)) but then proceeds to focus on activated and delegated variants as its title suggests. The decision to separate choices of state from the report has important consequences in terms of questions that are not raised and issues that are not addressed. In particular, suppose a decision were made that the disposition of all variant bundles would permit a maximum of one label to be activated and that other labels would be blocked (or permanently withheld). That would turn the variant picture into a proactive and low-overhead mechanism for keeping excessively similar names out of the DNS and would dramatically change much of the analysis in Section 5 of the report

The report suggests that all variants generated using the LGR against a label are automatically assigned the withheld state (on approval of the primary label) and then considered for allocation and subsequently activation through an application process (Recommendation 6.1.1). Thus it is not clear why it is suggested that “[t]his report claims to be agnostic about the decisions as to how a variant that is identified and made part of a bundle is assigned to a state”. Authors will make this process more explicit in the report for more clarity for the readers.

As someone who would almost certainly be considered part of the technical community, I believe the characterization of that community in the report, particularly in Section 1.4, is inaccurate. Many of us consider a high-quality and predictable user experience, one that is sensitive to a range of issues with data entry and character rendering and output, to be fully as important as the sort of narrow view of security and stability issues represented in the report. It is likely that, compared to other groups, a larger fraction of the technical community understands the inherent limitations of the design of the DNS and are consequently more hesitant about the possible consequences of trying to “trick” the DNS into doing things that lie outside or in contradiction to that design, but that is a separate issue. Although at least some of us see many issues with an expansive view of variants (some of which are reflected in this review), those concerns come from efforts to understand and balance the issues, not from the narrow focus the report suggests.

Thanks for the comment. This is understood and appreciated. The report will be updated to make this very clear.

For me, each new report from the Variant Information Project activity seems to reinforce a tentative conclusion This is all just too complicated and, as this report (even without the

additional issues identified in this review) clearly shows, creates significant additional risks and costs without being able to offer the payoff of a predictable and consistent experience that will meet the needs of end users. Adjustment of expectations to match reality and considerable simplification appear to be in order. Most obviously, that might be to simplify and redefine the LGR activity to an effort to define a repertoire for the root and make general repertoire recommendations for SLDs (and perhaps below), to ask the team that produced the present report to redo it to focus exclusively on root IDN issues rather than IDN variant ones, and to deal with requests for related names in the root zone —especially those not associated with historical ccTLDs — on an exception basis with the applicant having to demonstrate to the community that the relationship is important enough (and will solve a specific problem that can be identified) to justify the various “challenges” and other difficulties outlined in this report.

We agree that IDN TLD variants are, from a management perspective, quite complex. From the user’s perspective, variants may not be nearly so complex. Most activated variants may simply redirect to the primary labels, similar to the redirection examples cited with ASCII domains. Or they may fail to resolve. Or primary goal is to avoid any instance in which a variant resolves to a domain owned by a different entity. Our secondary goal is to consider how to improve the user experience of those active variants in light of the fact that registries are already supporting different variant delegation and support models.

As discussed, an ideal case is not achievable among competing principles. However, there are still advantages of the current approach. First, LGR does give a very useful common reference to define variant sets for all levels in the DNS, which would promote consistency. Further, though activation of all variants may not be practical, in practice there will be variants activated due to security or usability needs (as is proposed). A high threshold to get arbitrary variants activated means that variants following similar rules and motivations as those already activated will have a stronger case for approval. So we will get more of the same type activated over time. This does make the user experience more consistent. Further, (like the 80/20 rule) there is a good possibility that a few first variants which are added will contribute significantly toward the consistency (as those will be the most necessary ones for usability and security).

Therefore, there is value in doing this in a community driven coordinate fashion than leaving it arbitrarily on the applicant.

(KF)

The authors thank the reviewer for his feedback regarding the requirement for LGR at second level.

While the use of common LGRs is advantageous to the consumers (and those that write software for them) of IDN domain names, this advantage is already well understood by registries who have deployed IDNs. There is no discussion within the paper as to why, when faced with these advantages, some registries have opted not to use the same LGRs, while others have chosen to use LGRs already

implemented.

There is a section on current SLD practices for ccTLDs in various scripts, including Chinese, Arabic, Devanagari and Latin scripts. The state of practice does indicate that there is use of common LGRs across various scripts and languages. These similarities have been derived by the communities directly without intervention of ICANN. Authors will further look into the gTLDs practices.

If the option of voluntarily using the same LGR exists, as it does today, then there should be a strong reason to require through additional regulation that TLDs use common LGRs. Such a reason has not been put forward in the report. As a result, this appears to be excessive regulation that may stifle innovation in the IDN market for no clear benefit. Today an open market exists and TLDs are free to choose to use common LGRs, with the benefits such commonality delivers, or not to do so. There is no discussion as to how the market's current cooperative approach has failed and how further regulation would assist.

The authors reiterate that though this report recommends requiring registries to implement the same LGR consistently across the DNS tree, it also makes allowances for additional characters that are generally used outside of the root. The report also makes allowances for registries further localizing their LGR provided they document the reasons why such localization is necessary as well as documenting the changes. This allows for innovation but balances the need for consistent user experience.

I suggest that recommendation 6 have the word "require" replaced with "request". Full suggested recommendation:

ICANN should request IDN TLD registries with variants to apply the relevant (script) subset of the root zone LGR and state life cycle for variants across second-level domain labels. Deviations should be justified.

Authors will further consider this option in light of the current discussion and suggestions.

(JY)

Authors thank the reviewer for his earlier feedback on issues and current feedback on recommendations. His earlier suggestions have already been incorporated in the published report.

The observation, if not obvious enough, is that Variant IDNTLD's "success" relies not only on variant management and registration process, but its relationship with other components beyond registration. In my opinion, this is worthy to point out in conclusion and even in executive summary.

This is indeed true and noted for further actions by the authors while finalizing the report.

I use the same matrix table from Section 5.4 (see second page) and add two columns: one for area of focus to resolve the challenges, and another one for where the recommendations applied.

Thank you for your suggestions. We are including additional traceability matrix for mapping various aspects of the report. These will be included in the final version.

The recommendations, especially recommendations for ICANN& registries & registrars, are clear and specific. However, many of these recommendations are short term focused and registration focused. Majority of challenges fall into technical communities. And for these areas of challenges, the recommendations for technical communities are not as specific and clear as recommendations for ICANN, registries, and registrars. ...(What work can be done beyond provide technical specification? What effort can be made to help making specification into implementation?)

The report has tried to focus on aspects which ICANN can undertake and also identify other relevant aspects not directly in the ambit of ICANN. However, in the latter case, it is up to the other stakeholders (not holding any contractual obligations to ICANN) to review and determine the best way forward so the recommendations cannot be made much more concrete.

This report also recommends ICANN, registries, and registrars, in different wordings but same intent, to help end users and tool/software makers (6.1.11, 6.2.5, 6.3.5). There is no doubt some efforts overlapped, but a great opportunity for collaboration to promote same (or similar) work that suit different entities' interests.

Thank you for this useful observation. The authors agree that IDNs and variants present a great opportunity for collaboration among the relevant stakeholders.