

# IDN Variant TLD Program - Risk Analysis Studies

May 2013

## Executive Summary

In early 2013, a member of the ICANN community raised concerns about risks inherent in the delegation of IDNs — particularly IDN variant TLDs — with emphasis on the risks of delegating IDN TLDs before the completion of ICANN’s ongoing IDN Variant TLD Project or wider experience of IDN deployment in other domains.

As part of its on-going work to assess risks and risk mitigation strategies, ICANN commissioned an expert study to evaluate these concerns with specific reference to the risks associated with the Root Zone LGR.

The expert study identified ten risks, rated on probability of occurrence (high, medium, low and near zero), severity if it does happen (high, medium, low) and the type of risk represented (security and stability, user experience, and legal).

A key conclusion from the expert study is the following: “None of these risks are very likely to come to pass, but some are moderately likely. All risks that have a high severity are of low risk of happening.”

This summary presents highlights of both the original paper and the subsequent study. The full papers are available as [Appendix 1](#) and [Appendix 2](#).

## The Risk Paper

The concerns raised to ICANN were conveyed in a paper entitled “IDN TLDs: An endeavour fraught with risk” that presents “an assessment of the risks to users of the Domain Name System, and to ICANN, associated with the introduction of IDN gTLDs and, with them, IDN Variant gTLDs.”

The paper seeks “to identify the risks for users of the domain name system associated with proceeding with IDN gTLDs at this time, and to suggest that a rigorous examination of the risks be undertaken before proceeding to delegate IDN gTLDs.”

After outlining several key assumptions, the paper enumerates various specific risk areas and provides some suggestions on how they might be addressed. The risks are seen as affecting the consistency and predictability of user experience as well as the interoperability of the DNS. From there, the author Dennis Jennings proceeds to “examine the differences between ASCII domain names and IDN domain names and to make the requirements for IDNs explicit” and does this by “examining each of the components of the IDN ‘Label Generation Rules’ or LGR, which is the name given by ICANN’s IDN Variant TLD Program to the combination of the code point repertoire, variant code point rules, and IDN variant label disposition rules for the DNS Root Zone.” Together with “a standard way of representing these rules (e.g. IDN tables), storing and making them available to users, and the software tools that ICANN will need to process the IDN LGR,” these form the technical underpinnings for managing the delegation of IDN TLDs.

A key concern of the paper is that the work to populate this IDN LGR for the Root Zone has not begun<sup>1</sup> and until it is completed “no ICANN community-wide agreed specifications of which code points are variants of one another will be available.” The lack of a defined IDN LGR, as well as the lack of any agreement among TLD operators on consistent handling of IDNs, are some of the key sources of risks Jennings identified. The paper sees a further risk of “chaotic inconsistency and unpredictability for users” should the IDN LGR allow the allocation of any but the most restricted number of IDN TLD variants defined by the IDN LGR.

The paper then raises risks of incompatibilities between already delegated IDN TLDs such as the ccTLDs in the IDN ccTLD Fast Track or the applications for IDN TLDs in the new gTLD Program and those IDNs and variants permitted under the IDN LGR once it is developed. It warns that “ICANN has recently given priority to the processing and delegation of IDN TLD applications in the new gTLD Program. ICANN is coming under increasing pressure from elements of the Internet community to give priority to the delegation of the variants that have been requested as part of the applications for IDN TLDs in the new gTLD Program.”

The paper concludes with a number of recommendations. In part, the recommendations include pausing the delegation of IDN TLDs until the risks are further examined and the IDN LGR is reviewed, as well as recommendations to communicate the potential risks of IDN TLDs to the ICANN community and IDN TLD applicants.

## The Expert Study

To better understand the risks identified in the paper, ICANN commissioned an expert study to review its findings. The paper outlining the study’s findings is entitled “Expert Review of ‘*IDN TLDs: An endeavour fraught with risk*’ and the staff response”, with a goal to “evaluate what risks arise from delegating IDNs (or any name implicated by the eventual IDN Label Generation Rules – “LGR”) prior to the development of the LGR.” The study presents a detailed matrix of risks identified together with their estimated probability and severity. According to the terms of reference<sup>2</sup> for the study, the focus is on the risks due to incompatibilities between existing and applied for IDN TLDs and those permitted or denied under the final IDN LGR process<sup>3</sup> for the Root Zone.

A series of twelve questions were posed to the expert reviewer, and the analysis paper responds to them in a systematic question-answer format. The questions examine issues such as “are there risks ... that come from proceeding with delegation of IDN TLDs without the deterministic rules the LGR are supposed to provide?”, and whether there are risks to “proceed” and “not to proceed” with ICANN’s announced plans relating to IDN Variant TLDs.

---

<sup>1</sup> In the time since this paper was written, the ICANN board approved the IDN LGR procedure, which has now progressed into its implementation phase, with an expected completion date of mid 2014.

<sup>2</sup> The Terms of Reference contain some references to a Staff Response to the risk paper. They are part of the staff work on analyzing the risks and are not critical for the understanding of the evaluation contained in the expert study.

<sup>3</sup> The risks of user confusion associated with the likelihood that IDN TLD operators will have completely different rules for the code points permitted, and the associated IDN variant rules, for the second (and lower) level domain labels sold by them to third parties, have not been covered by this expert study. These second level domain label rules are a matter for the IDN TLD Registries, and are outside ICANN's remit.

A small sampling of observations made in the review include that the risk of not detecting variants because of the absence of the IDN LGR “is high, [but] the consequences appear to be low”; on amending the project plan based on newly identified concerns that “it would appear to be somewhat improvident to continue with an old plan of action in the face of new understanding” but that “failing to cleave to announced plans could affect (negatively) some applicants”.

A fundamental assumption made by the study in these evaluations is that implementation of the IDN LGR procedure begun by ICANN will be carried out according to the principles laid out in the procedure as approved by the ICANN Board. In particular, any conclusions on risks by the Expert Study depend on the LGR to be “conservative” in its implementation.

“The LGR process requires, in its guiding principles, an overarching conservatism. If we assume that the LGR process actually produces a very conservative LGR in the end, then the LGR will include a much reduced subset of Unicode, will minimize the number of allocatable variants, and will maximize the number of blocking variants.”

The complete set of assumptions and observations made in the study can be found in the full report.

## Expert Study Conclusion

The study concludes that “there are three classes of risk: risks related to the code point repertoire, risks related to variants arising from those code points, and risks to ICANN’s reputation and future status as a responsible steward of the root zone.

Based on these categories of risks, the study identifies ten “particular risks that are apparent after contemplating the answers to the questions in the terms of reference” and evaluates the probability of these risks coming to pass. The ten risks are evaluated based on their probability of happening (high, medium, low and near zero), severity if they do happen (high, medium, low) and the type of risk they represent (security and stability, user experience, and legal). The expert review concludes that “none of these risks are very likely to come to pass, but some are moderately likely. Some of the risks are of low consequence, whereas others would have severe negative effects if they were to happen. According to the analysis [presented in the study], all risks that have a high severity are of low risk of happening.”

Because the intent of the study was to describe and evaluate the risks, the study does not contain any particular recommendations.

## Final Remarks

After reviewing both the risk paper and the expert study, and discussing the papers with the authors and the program team, the IDN Board Variant Working Group recommended to the ICANN Board to approve the procedure for creating the IDN Label Generation Rules for the Root Zone. The ICANN Board [approved](#) this important milestone in the IDN Variant TLD Program and work continues on the next set of projects, namely the implementation of the IDN Root LGR procedure.

## **Appendix 1**

### **IDN TLDs: An endeavour fraught with risk**

# IDN TLDs: An endeavour fraught with risk:

---

Prepared by: Dennis Jennings

The work to develop this paper has been supported, in part, by ICANN, but the content does not necessarily reflect the views of ICANN staff or any other members or groups in the ICANN community.

## **QUICK Colloquial Style Board Level Summary:**

**This paper suggests that ICANN and the ICANN community may have put the “Cart before the Horse” in proceeding with IDN gTLDs before the appropriate definitions, standards, rules, etc. for variants have been developed and agreed by the ICANN community.**

**The paper outlines several serious risks to the usability of IDN Domain Names and suggests that the ICANN Board needs to ensure that the “Horse” is firmly ahead of the “Cart”.**

**The paper recommends that ICANN get external DNS and Unicode experts to review the current IDN TLD applications. Depending on the outcome of the reviews, the paper suggests possible approaches to minimise the risks.**

**The paper also suggests that it may be prudent for ICANN to pause all activities leading to the introduction of IDN gTLDs, and possibly additional IDN ccTLDs, until a comprehensive set of solutions has been developed and implemented.**

## **1 Introduction:**

This paper presents an assessment of the risks to users of the Domain Name System, and to ICANN, associated with the introduction of IDN gTLDs and, with them, IDN Variant gTLDs.

The objective in writing this paper is to try to identify the risks for users of the domain name system associated with proceeding with IDN gTLDs at this time, and to suggest that a rigorous examination of the risks be undertaken before proceeding to delegate IDN gTLDs.

An underlying assumption is that it is desirable to introduce IDN Variant TLDs<sup>1</sup>, and that the decision to proceed to introduce them is in principle correct, but that they must to be introduced with due care and attention, and the risks outlined here carefully addressed, so that that IDN Variant TLDs will be usable and achieve significant benefits for Internet users.

A key assumption is that the draft report “Examining the User Experience Implications of Active Variant TLDs”<sup>2</sup> developed by ICANN’s IDN Variant TLD Program accurately depicts the impact of the introduction of IDN Variant TLDs and variant domain names on various types of users of the Internet and the serious risks to those users.

A key assumption is that the approach being taken by ICANN’s IDN Variant TLD Program to develop a “Procedure to Develop and Maintain the Label Generation Rules (LGR) for the Root Zone in Respect of IDNA Labels”<sup>3</sup> is correct and is implementable,

---

<sup>1</sup> A number of DNS experts do not agree with this assumption.

<sup>2</sup> ICANN IDN Variant TLD Program - User Experience Report <http://www.icann.org/en/resources/idn/idn-variant-active-ux-23oct12-en.pdf>

<sup>3</sup> ICANN IDN Variant TLD Program - Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA Labels - (<http://www.icann.org/en/resources/idn/draft-lgr-procedure-07dec12-en.pdf>)

# IDN TLDs: An endeavour fraught with risk:

---

and that the proposed ICANN Integration Panel and the ICANN community based script / writing system Generation Panels will succeed in developing the LGR for the Root Zone.

(It is important to note that this paper only addresses the question of IDNs at the TLD level. However, IDN TLDs should not be considered in isolation, and the impact for users of fully qualified IDN domain names cannot be ignored. The statements in this paper about adherence to a single simple set of rules to ensure that a user's experience is consistent and predictable across all IDN TLD environments, apply – with even greater force - to fully qualified IDN domain names (IDN FQDNs) of the form - *IDN-level4.IDN-level3.IDN-level2.IDNTLD*. However, that is a topic for a separate paper).

## 2 IDN Issues

### 2.1 IDNs vs. ASCII Domain Names

Internationalised Top-Level Domain Names (IDN TLDs) are a major development of the use of the Domain Name System (DNS), but introduce complexities for the users (whether an end user, system configurator, application developer or other user) that are new and mostly unprecedented<sup>4</sup>.

For over 25 years TLDs have been restricted to a very limited set of ASCII code points, with no ambiguity as to what is a valid / permitted code point, no variant code points, and a totally consistent and predictable set of rules for users. Those restricted rules and the nature of ASCII (really "Basic, undecorated Latin" characters) have brought significant advantages, notably that the relationship between upper and lower case is unambiguous, and the category – whether ccTLD or gTLD – of each top-level name was clear from its length. The enforcement of rules, while they constrained what could be registered, made things predictable for users.

Introducing internationalised domain names challenges the assumptions that users have for using domain names. Many of the assumptions that users may have developed, based on experience with ASCII domain names, are not valid when dealing with IDNs. The character repertoire is vastly expanded. A character written in one way in one place or for one language may be written in a different way in a different place or for a different language. There are differences in keyboards and differences in writing, spelling, and even the script in which a particular language is written.

However, the promise of IDN TLDs has also led to the expectation that IDN variants will just (somehow) “work”: Many users appear to expect that the introduction of IDN TLDs will be accompanied by changes to the IDNA / DNS environment that will automatically resolve differences in keyboards and differences in writing, spelling, and even the script in which a particular language is written. There are expectations that the DNS should adapt to the users rather than the users changing behaviour toward what names are in the DNS<sup>5</sup>. There are expectations for what are really language-specific rules about what strings match, and that multiple domain name variants can be used that "mean the same" or even "are the same".

---

<sup>4</sup> The ICANN IDN Variant TLD Program - User Experience Report  
<http://www.icann.org/en/resources/idn/idn-variant-active-ux-23oct12-en.pdf>

<sup>5</sup> There seem to be people who still believe that DNAME or some other to-be-developed DNS technology will magically provide a matching function that will make all of the variant problems go away, and some believe that this has been promised by ICANN. In particular, because DNAME specifies only a relationship between a single name and a sub-tree, it is not applicable to a situation in which a particular script requires variant treatment for both the TLD names and names in that TLD (much less names further down the tree).

## **IDN TLDs: An endeavour fraught with risk:**

---

These expectations, and the reasons for them, are clear, but are wildly optimistic. Developing approaches to accommodate any of them must be understood as a trade-off against user-perceived consistency, predictability, and stability.

### **2.2 Consistency and Predictability**

Adherence to a single simple set of rules for the generation of ASCII TLDs has enabled the DNS to scale for users of ASCII domain names and has ensured that a user's experience is consistent and predictable across all ASCII TLD environments. The key for users has been consistency and predictability.

Consistency and predictability will also be the key for users of the IDN environment. Without such consistency and predictability, a user's experience will vary from one TLD to another, even within the same Unicode script and writing system. A user's experience may differ from one application to another or one machine to another, even with the same TLD.

Without such consistency and predictability, IDNs in the DNS will not be usable in any meaningful way, and IDNs will fail to deliver any of the expected benefits to users, or any of the expected commercial or other benefits to Registries and Registrars, and to the users of the services.

Let's explore this a little bit more.

If IDN TLD applicants are permitted to define all the rules for their TLD, including the code point repertoire and the variant rules for the TLD in the Root Zone, the result will create total chaos. There will be conflicting rules about which IDN strings are variants of one another, and conflicting claims about which strings are in an IDN TLD variant set. Apart from the contention issues that will arise, this would destroy any possibility of consistency and predictability for users of IDNs.

It is therefore an obvious requirement that there be a single set of rules for the generation of IDN TLD labels in the DNS Root Zone.

However, let's further imagine that the single set of rules for the Root Zone results in a large number of variants of any particular IDN TLD string, and that the rules permit any number of the label variants to be delegated. This will cause serious management problems if the Registry is trying to maintain consistency across all the variants - there will be errors, and each error will contribute to an overall user perception of unpredictability of the Internet, not just the DNS. Since all variants are in some sense "the same" or are indistinguishable one from the other, this will also create huge problems for all users

It is reasonable to re-state that such a scenario could not deliver consistency and predictability to users of the DNS, and in that scenario IDNs in the DNS will not be usable in any meaningful way, and that IDN TLDs will fail to deliver any of the expected benefits to users, or most of the expected commercial or other anticipated benefits to Registries, Registrars and the users of their services.

### **2.3 Interoperability**

The key concept is Interoperability – the ability of systems and applications to interoperate (or exchange data and communicate and work together) – based on the (voluntary) adoption of a set of open standards. This permits new systems and applications to be developed, and, by adhering to the open standards, to interoperate with all existing standard systems and applications. It has been the basis for the growth of the Internet to date; it's the mechanism for enabling diverse users and applications to work in stable, predictable ways.

## **IDN TLDs: An endeavour fraught with risk:**

---

This point has a further consequence when we look at DNS as a protocol and a set of naming conventions that must support interoperability not just for one community of users or set of applications, but globally: the DNS root is a unique shared resource, and decisions about what it allows (or does not allow) have implications throughout the namespace. Since mistakes here have wide-ranging implications, support for interoperability for IDNs in the current and future global Internet demands consideration of possible impacts, even by side effect, on the stability and predictability of the system overall, and a strongly conservative preference towards what's introduced in it. Many things that are simple and obvious to a particular user community in a particular context such as a ccTLD namespace are not so simple or obvious without that assumed context – and names in the root do not and cannot carry such an assumption of context.

This requirement for the root to support interoperability for all users suggests that, in addition to IDNA 2008, a set of open standards for IDN TLDs needs to be developed and adopted that copes with variants, and delivers interoperability and the required consistency and predictability. Furthermore, the scope of such standards needs to include the issues raised in the “user experience” work, as well as acknowledging the limitations in the DNS protocol which make it impossible (or at least very, very dangerous) to attempt to invent new properties of “sameness” or proceed as if they already existed anywhere in the DNS. These open standards should ideally be agreed and available before any deployment of IDNs is contemplated.

### **2.4 Risks**

It is important to stress that such additional open standards to support expected use by the user community should have been required before any IDN TLDs were deployed, because, unlike in other previous Internet standards situations, non-standard implementation cannot easily - if at all - be undone. Certainly, no further IDN TLDs should be deployed until such open standards are available. Once a TLD has been delegated and brought into use, it is almost impossible to remove it. This is quite unlike earlier standards situations – for example users of proprietary communications standards could always drop existing protocols and adopt the new standard ones (e.g. network (IP), transport (TCP), Web (HTTP, HTML), Registry-Registrar communications (EPP), or other) to achieve the desired interoperability.

The risks to users associated with the introduction of further IDN TLDs are therefore very significant, and the utmost caution and the most conservative possible approach are required.

ICANN – as the responsible steward of the DNS Root Zone – has responsibility to ensure that these risks are addressed, and, to the maximum extent possible, eliminated. As a consequence, the risks to ICANN itself are possibly existential, certainly reputational, and most obviously financial. Applicants for IDN TLDs are likely to be (or claim to be) unaware of the risks to them and to their users and will most likely sue ICANN for loss and compensation for damages if it becomes clear that ICANN did not appropriately deal with the risks involved or advise them of the risks.

## **3 The IDN label Generation Ruleset (LGR)**

It is helpful when considering the introduction of IDNs to examine the differences between ASCII domain names and IDN domain names and to make the requirements for IDNs explicit.



## **IDN TLDs: An endeavour fraught with risk:**

---

These requirements include the specifications of the code point repertoire, the code point variant rules that determine the set of variant labels associated with a requested label, and the rules for which of the variant labels may be delegated.

The name given by ICANN's IDN Variant TLD Program to this combination of the code point repertoire, variant code point rules, and variant label disposition rules, is the "Label Generation Ruleset" or LGR for the DNS Root Zone. In addition, a standard way of representing these rules (e.g. IDN tables), storing and making them available to users, and the software tools that ICANN will need to process the LGR, will need to be developed.

Examining each of the components of the LGR:

### **3.1 The Code Point Repertoire**

#### **3.1.1 ASCII**

RFC 1123 says that ASCII labels in the root zone "will be alphabetic", using 26 code points, all of which are valid and permitted.

#### **3.1.2 IDNs**

The latest version of Unicode contains a repertoire of more than 110,000 characters, of which more than 70,000 are Chinese, Japanese and Korean (CJK) characters. (Many of the characters – e.g. punctuation, currency, mathematical, drawing, geometric, etc. – are of course explicitly prohibited in IDNA 2008). The repertoire is divided into 100 scripts.

The IDNA2008 "protocol valid" ("PVALID") category provides a selection of code points for IDNs, mostly by formalizing the elimination of things that are not letters or digits. IDNA also requires that zones define more specific and limited repertoires (and perhaps other naming rules) for use in those zones. There is, however, no general specification yet of which of the remaining codes points are to be permitted in IDNs. This will be addressed for the Root Zone by the implementation of the LGR Procedure.

### **3.2 The Variant Code Point Rules and Variant Labels**

#### **3.2.1 ASCII**

No variants

#### **3.2.2 IDN TLDs**

The definitions of variants may be different from one script / writing system community to another. The definition used in the Integrated Issues Report<sup>6</sup> produced by the ICANN Variant Issues Project is "exhibiting some relationship such that one of the variants may, under some circumstances, somehow be conflated with another". No ICANN community-wide agreed specifications of which code points are variants of one another will be available until the LGR for the Root Zone has been developed. The assertions about code point repertoire, code point variants, and variant labels made in ccTLD Fast Track delegations and in some applications in the current gTLD round may or may not be consistent with those that will be defined under the LGR.

The development of the LGR for the Root Zone by the relevant script / writing systems communities (as envisioned in the Procedure) will provide the mechanism for those communities to define and agree the set of variant TLD labels associated with any string of permitted code points (a TLD Variant Set).

---

<sup>6</sup> <http://www.icann.org/en/topics/idn/idn-vip-integrated-issues-final-clean-20feb12-en.pdf>

# IDN TLDs: An endeavour fraught with risk:

---

## 3.3 The Disposition of IDN TLD Labels in an IDN Variant Set

In the work on the Procedure the word used to describe the permitted actions on each of the labels in a label variant set is the “Disposition” of the label. To date, the dispositions “blocked” and “allocatable” have been considered. This may lead to the need for arbitrary selection from a set of “allocatable” labels. Dispositions of “preferred” or “required” may also be required to ensure that there are consistent and predictable rules for which labels in the a variant set are to be delegated.

(In the Procedure the “Disposition” rules are considerably more sophisticated than implied here, and introduce the notion of “tags” to differentiate between different sets of disposition rules for various sub-repertoires of a script. The details are beyond the scope of this paper).

In principle, it is possible to envision a situation where every variant in a TLD Variant Set may be delegated to the applicant Registry, with the TLD registry required to provide some synchronisation or other coordination process to ensure that labels at the second level are appropriately delegated. In practice, to minimise confusion and complexity, a consistent and predictable process is required to select the minimal subset of a TLD Variant Set for delegation.

In all cases, unless absolutely necessary, that subset should be a single IDN variant TLD (i.e. only one label is delegated). In the case of CJK TLDs, because of the demographic split between Simplified and Traditional Chinese character usage, that subset should be a limited to a single pair of simplified and traditional TLDs. A similar exception may be required for Arabic (or other script) TLDs. The current version of the Procedure does not ensure this outcome. This poses a political risk for ICANN with certain communities if IDNs do not function as expected.

The development of the Root Zone LGR (using the Procedure) must ensure this outcome and must identify the single IDN TLD (or Chinese (or other) pair of IDN TLDs) to be delegated, and the identification process must be consistent and predictable. All other variants in a TLD Variant Set must be blocked.

Any other approach will most likely lead to chaotic inconsistency and unpredictability for users.

ICANN has the authority to ensure that these outcomes are achieved at the TLD level.

## 3.4 Status of the Label Generation Ruleset (LGR)

The ICANN IDN Variant TLD Program is developing the procedure (the Procedure) to be followed by ICANN and the various script / writing system communities around the globe that will enable the development of the LGR<sup>7</sup>. At this time, the work on the Procedure has not been completed, and on completion must still follow ICANN processes for approval and implementation. Implementation of the Procedure is itself only the beginning of the process that will generate actual and specific rules. We are still a long way from having the necessary LGR for the Root Zone.

## 4 Status of IDN TLDs

At this time

- There is no defined and agreed LGR for the Root Zone.

---

<sup>7</sup> The ICANN IDN Variant TLD Program - Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA Labels (<http://www.icann.org/en/resources/idn/draft-lgr-procedure-07dec12-en.pdf>)

## IDN TLDs: An endeavour fraught with risk:

---

But

- ICANN has delegated a number of IDN ccTLDs in the IDN ccTLD Fast Track. With hindsight, this is already a risk, since these prior delegations may present challenges for future policy or standards, or for consistency in naming. There is obviously risk that these TLDs will not be consistent with whatever IDN standards and rules ultimately emerge. ICANN shares responsibility for IDN ccTLDs with the applicable governments<sup>8</sup>.
- ICANN has permitted applications for IDN TLDs in the new gTLD Program, now underway. However, ICANN's IDN Variant TLD Program did not have time to complete its work so that ICANN could ensure that the development of the necessary LGR for the Root Zone would be completed in advance of the processing of these applications. It would be very sensible for ICANN to ensure that the launch of applied-for IDN TLDs includes the outcomes of the LGR for the Root Zone. ICANN bears sole responsibility for gTLDs.
- ICANN has recently given priority to the processing and delegation of IDN TLD applications in the new gTLD Program. This decision makes it much harder for ICANN to ensure that the necessary Root Zone LGR is developed and used when processing the applications. In addition, delegating a single variant when possibly two are required (or delegating the incorrect variant according to the rules of the future LGR) would be irresponsible and would exacerbate the risks. (This decision by the ICANN Board appears to have been taken without any consideration of the ICANN IDN Variant TLD Program work, or any consultation with the ICANN Board's working group that oversees that Program (the BV-WG)).
- ICANN is coming under increasing pressure from elements of the Internet community to give priority to the delegation of the variants that have been requested as part of the applications for IDN TLDs in the new gTLD Program. From the point of view of the applicants – many of who are undoubtedly unaware of the risk to users described here – this is perfectly understandable. The risks to the users of the DNS would be greatly exacerbated if ICANN were to succumb to that pressure. Without addressing and resolving the issues raised here, that would not be responsible.

## 5 Recommendations to the ICANN Board

### 5.1 ICANN's Responsibilities

*[ICANN] ... is a nonprofit public benefit corporation and is not organized for the private gain of any person. ... in recognition of the fact that the Internet is an international network of networks, owned by no single nation, individual or organization, ... and promoting the global public interest in the operational stability of the Internet by (i) coordinating the assignment of Internet technical parameters as needed to maintain universal connectivity on the Internet; ... (iii) performing and overseeing functions related to the coordination of the Internet domain name system ("DNS"), including the development of policies for determining the circumstances under which new top-level domains are added to the DNS root system; ... and (v) engaging in any other related lawful activity in furtherance of items (i) through (iv).*

---

<sup>8</sup> The IDN ccTLDs are, at least in principle, all linked to governments and it can be assumed that governments are responsible custodians of their own name spaces. (In relation to IDN FQDNs, governments have another advantage: if they make a rule as to how 2<sup>nd</sup> (and even 3<sup>rd</sup> and beyond) level domains are to be administered within "their" TLDs, and do so as a matter of law, they really can enforce those rules).

# IDN TLDs: An endeavour fraught with risk:

---

ICANN – as the responsible steward of the DNS Root Zone – needs to address the risks that the introduction of IDNs will create for the users of the DNS.

## 5.2 The Ideal Situation

On the assumption that the concerns about the introduction of IDNs and the risks to users of the DNS, as outlined above, are indeed valid, the ideal solution is for ICANN to pause all activities leading to the introduction of IDN gTLDs, and possibly additional IDN ccTLDs, until a comprehensive set of solutions to the issues have been developed and implemented.

Pausing the introduction of IDN gTLDs, while painful, is viable because there is still a rational stopping point until the first IDN gTLD is delegated. The pause would be temporary until appropriate solutions were developed.

## 5.3 Possible Practical Recommendations

Perhaps more realistically, the following are the actions that the ICANN Board can take to deal with the issues presented in this paper:

### 5.3.1 Advise the ICANN Community

- Formally advise the ICANN community that recent work on the introduction of IDN TLDs has confirmed serious but as yet un-quantified risks to users of the DNS<sup>9</sup>.
- Formally brief the applicants for IDN TLDs under the new gTLD Program of the risks they will have to accept<sup>10</sup> should they insist that ICANN proceed with the evaluation and possible delegation of their requested IDN gTLDs.
- Formally advise the community that consideration needs to be given to pausing the processing and delegation of IDN gTLDs (or IDN gTLD with variants) until adequate general rules and understandings are in place.

### 5.3.2 Consult the DNS Technical Experts

ICANN should explicitly and formally consult the recognised DNS experts, including the full range of independent outside voices, consultants on DNS and related technical areas who are already familiar with ICANN processes and constraints, and Board liaisons from the RSSAC, SSAC, and the IETF – and others especially DNS experts with in-depth script / writing system expertise – and consult the recognised Unicode experts – to confirm (or contradict) my concerns about the risks to users of the DNS.

### 5.3.3 Review the Root Zone LGRs (Root Zone IDN Tables) Provided with the TLD Applications

In addition, it is recommended that ICANN should immediately organise an external panel of DNS and Unicode experts to review<sup>11</sup> the Root Zone LGRs (also called the TLD IDN tables) that have been provided by the applicants for the IDN TLDs under the new

---

<sup>9</sup> There are obvious risks to ICANN in publishing such an advisory notice, but the requirements of openness and transparency demand that this be done. In any case, there is widespread and growing appreciation of these risks in the ICANN community.

<sup>10</sup> It could be argued that this is covered in the terms associated with the new gTLD Application Guidebook and the new gTLD Application process, but it may not be prudent for ICANN to rely on those terms.

<sup>11</sup> It is possible that pre-delegation testing (PDT) of the new gTLD registries will address some of these issues, but I have no information on this PDT work. It might be valuable to add this IDN TLD table validation with respect to a root zone LGR to the statement of work for the contractor commissioned to undertake the pre-delegation testing (PDT).

## **IDN TLDs: An endeavour fraught with risk:**

---

gTLD Program. (Accelerating the formation of the ICANN Integration Panel, as envisioned in the Procedure, could provide the required expert panel<sup>12</sup>).

### **Review of all submitted tables:**

If all these tables can all be reconciled into a single coherent consistent LGR for the Root Zone – code points, variants, variant rules, disposition rules, etc. – then, while this result does not necessarily eliminate the risks, it might be considered safe to proceed with the processing of the applications.

Given that some of the applications will have been prepared by people that are intimately familiar with the LGR issues and have prior experience with the existing IDN ccTLDs, whilst others are likely to have been prepared by people with absolutely no such experience and expertise, it is unlikely that all the submitted tables can so be reconciled.

Indeed, a number of experts have advised that anybody familiar with the issues would quickly come to the conclusion that the IDN tables as submitted cannot trivially be reconciled into a single coherent consistent LGR for the Root Zone<sup>13</sup>. Perhaps this is not surprising, as the new IDN gTLD applications were invited and submitted before the work in ICANN's IDN Variant TLD Program on the LGR Procedure had started.

If these tables cannot all be reconciled into a single coherent consistent LGR for the Root Zone, then that should be considered to be a serious program risk. In that case ICANN should ensure that the IDN Variant TLD Program completes the Root Zone LGR Procedure, and the LGR then developed by the relevant script / writing systems communities. The applicants for the new IDN gTLDs should be encouraged to participate in the development of the LGR for the Root Zone.

### **Review of submitted tables in a particular script / writing system:**

If all the applicant tables cannot be reconciled, ICANN should then ask the panel of DNS and Unicode experts to explore whether the applicants' Root Zone IDN tables for IDNs in a single script (e.g. the Han or Devanagari script) could be reconciled into a single coherent consistent partial LGR for the Root Zone (without of course having a negative impact on, or creating potential for confusion with, related or similar scripts).

If that can be done to the satisfaction of the DNS and Unicode experts, then, while this result does not eliminate the risks, it might be considered safe to proceed with the processing of those applications.

If these tables cannot all be reconciled into a coherent consistent partial LGR for the Root Zone, then this should be considered to be a serious program risk. In that case ICANN should adopt the recommendations above on the need to complete of the IDN Variant TLD Program and the development of the LGR for the Root Zone.

### **Discriminating between applications:**

If the applicant tables in a particular script / writing system cannot be reconciled, then it might possibly be safe to proceed with some individual applications.

Given the likely disparity between the level of expertise and experience of the people behind the applications, and the resultant quality of the submission in term of the proposed IDN TLD tables for the Root Zone, it may be necessary and appropriate for ICANN's application review process to discriminate between the applications on the

---

<sup>12</sup> Establishing the interim Integration Panel (ref the Procedure document) with a mandate both to do these evaluations and also to get started on a first-draft repertoire would be an efficient approach.

<sup>13</sup> For example, approximately a dozen submissions of IDN tables have repertoires that contain the full list of EGYPTIAN HIEROGLYPHICS, which would massively violate the principles laid down in the draft procedure for creating a root zone LGR.

## **IDN TLDs: An endeavour fraught with risk:**

---

basis of the quality of the submitted IDN TLD tables, and the likelihood – in the opinion of the external DNS and Unicode expert panel – that these might be consistent with the future LGR for the Root Zone when developed by the ICANN community.

Such an approach might minimise the risks – while not of course eliminating them – of proceeding with the delegation of some of the applications for IDN TLDs.

### **However:**

If none of these reviews identify applications that might proceed, then careful consideration needs to be given to pausing the processing and delegation of IDN gTLDs (or IDN gTLD with variants) until adequate general rules and understandings are in place.

### **5.4 Enforcement of the LGR Rules**

One of the important aspects of the development of the LGR for the Root Zone, as outlined in the Procedure, is that the LGR will be developed by the relevant script / writing system communities participating in the LGR Generation Panels (moderated by the ICANN appointed expert Integration Panel – to ensure that the more technically risky suggestions are eliminated). This is critically important for subsequent enforcement efforts by ICANN: the Root Zone LGR rules will be the community's rules not ICANN's rules<sup>14</sup>.

It is absolutely necessary that ICANN enforce the LGR rules effectively and in a completely non-discriminatory manner. When ICANN says, "No, that is not permitted by your community rules", it must mean "No" and must stick to "No", notwithstanding whatever political or other pressure are applied. If the pressure becomes intense, ICANN can refer the LGR back to the relevant script / writing system Generation Panel for review and perhaps modification if consensus can be achieved (except that very limited change may be possible, given the facts dictated by the IDN TLDs (and associated variant sets) that have already been delegated).

---

I believe that it is desirable to introduce IDN Variant TLDs, and I would like to see them introduced. But, they must to be introduced with the utmost care, and the risks outlined here addressed, so that that IDN Variant TLDs will be usable and achieve significant benefits for Internet users.

Dennis Jennings  
17 January 2013

### **Postscript:**

I am very grateful to a number of DNS and Unicode experts who have reviewed various drafts of this paper, and provided comments, corrections, and criticisms and have suggested amendments that I have incorporated as best I can – including especially Cary Karp, Lyman Chapin, Asmus Freytag, Patrik Fältström, John Klensin, and Andrew Sullivan. However, the paper does not necessarily reflect their views. The opinions expressed, and of course all errors, whether of commission or omission, are mine.

---

<sup>14</sup> This could be contrasted with ICANN's visual confusability testing of requested TLD strings, where the community sees that process as opaque, without community endorsement, arbitrary, and with no possibility of appeal. The visual confusability testing is necessary and important, but the characteristic that I want to highlight is that it does not have community buy-in.

## **Appendix 2**

### **Expert Review of "IDN TLDs: An endeavour fraught with risk" and the staff response**

# Expert Review of "IDN TLDs: An endeavour fraught with risk" and the staff response

---

*A report for the Board Variant Working Group<sup>1</sup>*

## Summary

This review attempts to evaluate what risks arise from delegating IDNs (or any name implicated by the eventual Label Generation Rules – “LGR”) prior to the development of the LGR. It concludes that there are three classes of risk: risks related to the code point repertoire, risks related to variants arising from those code points, and risks to ICANN’s reputation and future status as a responsible steward of the root zone. None of these risks are very likely to come to pass, but some are moderately likely. Some of the risks are of low consequence, whereas others would have severe negative effects if they were to happen. According to the analysis below, all risks that have a high severity are of low risk of happening. This is covered in tabular form on page 4ff.

The intent of this report is to describe risks, but not to recommend specific action, so it avoids recommending any particular conclusion about what to do.

### *Note added in publication*

While there are some references in this report and in its Terms of Reference to a Staff Response to the risk paper, they are part of the internal discussions and evaluations that took place during the assessment of these papers and are not critical for the understanding of the evaluation contained herein. The report also mentions a consultant’s review of IDN tables submitted with IDN TLD applications or registered with IANA. Key findings from the IDN tables review are excerpted in Appendix B.

---

<sup>1</sup> Prepared by Andrew Sullivan. I prepared this report under contract with ICANN. The views in this report are my own, and may not represent the views of ICANN staff.



## Introduction

This report reviews two papers that the BV-WG has already seen: Dennis Jennings's "IDN TLDs: An endeavour fraught with risk", and the staff response to that paper. It assumes the reader has read those papers. This review is undertaken under the terms of reference outlined in Appendix A: Review terms of reference and has been prepared for the BV-WG.

This report answers the questions in the terms of reference in order. For the reader's convenience, I offer an overall evaluation first.

This report depends upon familiarity with the current drafts of the "Project 2.1" project team, which is the draft procedure for the development of the Label Generation Rules (or LGR. Henceforth this draft procedure is called the LGR procedure). The present report also depends upon familiarity with the terms common in the LGR procedure and those assumed in the Jennings paper and staff response.

It is important to keep in mind that this report does not concentrate on efforts at risk mitigation (though it assumes their effectiveness in some cases). This is because of the terms of reference for this work, which were aimed at exposing risks, both in kind and in degree. This is not to minimize the extent of efforts to avoid risk in the new gTLD delegation process, and the lack of discussion of mitigation already undertaken is not to suggest that it hasn't happened. The staff response to the Jennings paper contains discussion of those mitigation strategies.

## Overall evaluation<sup>2</sup>

### Risks and their categories

It is useful to think of the risks in this area as falling broadly into three categories:

1. Risks related to the repertoire of code points permitted for use in the root zone.
2. Risks related to variant labels that are in the root zone (or that could be).
3. Risks to ICANN's processes, procedures, or reputation.

The following are particular risks that are apparent after contemplating the answers to the questions in the terms of reference:

1. Repertoire, False Positive: Delegating a TLD that includes a Code Point (CP) that is not later included in the root repertoire.

---

<sup>2</sup> Much of the discussion in this section is due to review of an earlier draft of this report by ICANN staff. I am particularly indebted to Francisco Arias, Dennis Chang, and Steve Sheng for their helpful and elucidating comments.

2. Repertoire, False Negative: Denying a TLD delegation based on concerns about a CP that is later allowed by the root repertoire.
3. Variant, False Negative With Independent Delegation: Delegating two independent TLDs that are later determined to be variants by the root LGR.
4. Variant, False Positive: Denying a delegation based on variant checking (or variant declaration) that is later determined not to be a variant by the root LGR.
5. Variant, False Negative With No Activation: Delegating a TLD without activating a variant that users need (and that the LGR later determines to be a variant).
6. Variant, False Negative With Mistaken Delegation: Allowing immediate delegation of variant TLDs that the LGR later determines to be blocking.
7. ICANN: Applicants taking action against ICANN for allowing declaration of variants in applications that the LGR later determines not to be variants.
8. ICANN: Having to process more IDN TLD applications before the root LGR is ready.
9. ICANN: Succumbing to political pressure to deviate from the root LGR (thus undermining the effectiveness and legitimacy of the LGR).
10. ICANN: A court ordering ICANN to deviate from the root LGR (thus undermining the effectiveness and legitimacy of the LGR).

### The probability of these risks coming to pass

The LGR process requires, in its guiding principles, an overarching conservatism. If we assume that the LGR process actually produces a very conservative LGR in the end, then the LGR will include a much reduced subset of Unicode, will minimize the number of allocatable variants, and will maximize the number of blocking variants. For practical purposes, this would likely mean that Traditional and Simplified Chinese pairs will be allocatable variants, and a small number of Arabic code points will be allocatable variants. All other variants will be blocked.

*It cannot be emphasized enough that there is no way to be sure the LGR process will actually produce such a conservative LGR.* The LGR process has not even been adopted definitively. The Generation Panels and the Integration Panel have not been constituted. Also, there have been participants in the IDN TLD project who have expressed strong opposition to the most conservative vision for the LGR.

For the purposes of the following, I assume the most conservative LGR. This might as well be called a fond hope as a prediction. If the assumption is false, the *likelihood of risks coming to pass rises* in unpredictable ways. For instance, if (contrary to the recommendation of the Latin Variant Issues Report) the LGR contains variants for Latin code points, then all the ASCII-only applications also present a variant problem. Actual problems are, one hopes, forestalled by the String Similarity Panel's work, but in the absence of the LGR the code points will not be evaluated algorithmically. The uncertainty here is, in my reading, one of the key motivators of Jennings's argument. Given the current structure of the LGR process, there is no way to mitigate this.

With that caveat in mind, and given the small number of actual IDN applications in the current gTLD round and the efforts that have been made to find negative effects, we may evaluate the risks as follows:

Risk #	Probability	Severity	Type
1	Low	High	Security & Stability
2	≈ 0	Medium	Legal
3	Low	High	User Experience/Security & Stability
4	≈ 0	Medium	Legal
5	Medium	Medium	User Experience
6	Low	High	User Experience/Security & Stability
7	≈ 0	Low	Legal
8	Medium	Medium	User Experience/Security & Stability/Legal
9	Medium	Medium	User Experience/Security & Stability
10	Low*	Medium*	Legal/User Experience/Security & Stability

### Discussion

Risk 1 is judged to be low because the very existence of a delegation with a particular code point is evidence of demand for it. The risk is above zero, however, and if it were to happen the consequences would be very bad.

Risk 2 is judged to be near zero because none of the current TLD applications use obscure or unusual code points.

Risk 3 is judged to be low (but not zero) because of an independent assessment ICANN solicited of the IDN applications. The consequences would be very bad.

Risk 4 is judged to be near zero because there does not appear to be a great deal of contention among the IDN applications. The legal risk is that the disgruntled applicant will sue ICANN.

Risk 5 is judged to be medium, because there are known to be applications for variant strings that, under the AGB, will need to be blocked until the LGR work is completed. It is only a medium risk because, if applicants are serious about the notion of variants, they will choose to delay delegation of any string in the variant bundle until they may all be delegated. This risk was always part of the model of this round for gTLD applications, in any case, so its severity is also judged to be medium.

Risk 6 is judged to be low, because there are few IDNs in the applied-for strings and we should expect that someone would have objected if something should have been blocked, but was not. This is not near zero only because of the uncertainty introduced by the unknown nature of the LGR. The consequence of this risk coming to pass would be very bad.

Risk 7 is judged to be low mainly because the AGB was always very clear that ICANN was taking no position on whether asserted variants actually were variants. ICANN has the resources to withstand legal action in the event this comes to pass.

Risk 8 is judged to be medium because the LGR process is not yet adopted, and we have no idea how long it will take to render judgement in the event it is adopted. If that takes a short time, then the risk may be close to 0. If the process drags on a long time, this risk goes up. Because the risk affects so many areas, the consequences seem to be of medium severity.

Risk 9 is related to Risk 8, and is judged to be of medium risk for the same reasons. Observers of ICANN's actions in the past can in fairness observe that ICANN does not have a sterling record of standing up to pressure when the entity applying the pressure is influential or well-funded (or both). There is an argument to be made that the risk here should be judged low on the grounds that there are few applications in the current pool. The medium rating takes into consideration new applications (particularly ccTLD applications), however.

Risk 10 is difficult to evaluate. It is here judged to be low, mostly because of the assumption of a very conservative LGR. If the LGR process turns out to be very contentious, the probability of this happening goes up to some unknown degree. Similarly, the initial effects are bad because the LGR starts to have a lot of exceptions. In the event two different courts of competent jurisdiction rule differently on the same string (for instance, courts in two different countries), the effects would appear to be very bad.

### **Risk matrices**

We can place the risks into matrices related to risks arising from the code point repertoire or from variants.

## Repertoire risks

Code Point	DSP* rules string as	Later in root LGR	Later not in root LGR
In current applications	Pass	No risk foreseen	Risk #1
	Fail	Risk #2	No risk foreseen
Not in current applications		No risk foreseen	No risk foreseen

\* DSP = DNS Stability Panel

## Variant risks

Two Code Points are deemed:		Later by the root LGR Procedure		
		Blocking variants	Allocatable variants	Not to be variants
In this new gTLD round	Blocking variants	No risk foreseen	Risk #5	Risk #7
	Allocatable variants	Risk #6	No risk foreseen	Risk #4
	Not to be variants	Risk #3		No risk foreseen

## Questions and answers from the Terms of Reference

**Is it possible to know with any certainty that some IDN labels will never have any allocatable variants prior to the latest Label Generation Rules procedure being completed (at least for some subset of Unicode; for brevity, in what follows this is called "LGR")? If so, to what degree?**

The simple answer to this question is, "No," but partly because the LGR itself cannot answer this.

A key fact about any procedure that will both work and deal with the changing nature of Unicode is that the procedure can never guarantee certainty. The LGR procedure depends very heavily on a presumption of conservatism; this is not a

presumption of infallibility. Nevertheless, there are some cases where it appears very unlikely that there should be any allocatable variants for a particular label.

There are some assumptions in the LGR procedure, and in the prior Variant Issues Project reports, that suggest certain classes of labels are very unlikely to have any allocatable variants. One such class is labels using Devanagari code points, because the primary justification for variants for Devanagari appears to be to prevent confusing labels from being allocated. Of the writing systems studied by the VIP, only Devanagari exhibits this property. Even in the case of Devanagari, there is some risk, but it appears to be lower than other writing systems.

Apart from that case, it does not seem one could make any general claim about any given IDN label having no allocatable variant, unless one were to perform evaluation of the particular code points. For reasons already outlined by the proposed LGR procedure, evaluating just particular code points will leave open a possibility that, when other code points are reviewed, a relationship will be uncovered that results in new variants.

In the case of very well-developed practices for any given set of related writing systems (such as a widespread, well-developed, and mutually consistent set of practices adopted by several different registries using the same code points), there is some reason to suppose that those practices offer a basis for determining whether any particular label could have an allocatable variant. In order to perform this evaluation, it would be necessary to take into account all of the relevant registries' policies, and to evaluate the extent to which those policies were undertaken with appropriate linguistic guidance. The risks in this area go up according to the number of languages that use the writing system.

**Is it possible to know with any certainty, prior to the LGR being completed, the variants of those IDN labels actually in the current round of new gTLD applications? If so, describe the labels and the degree to which one can be certain.**

Because it is impossible to know in advance what the LGR will contain, it is neither possible to be certain whether any label has a variant nor what those variants will be.

Two strong indicators can suggest probable variants. The first is the list of variants that an applicant offers. This list runs the risk of being too broad (with the applicant attempting a "land grab") and too narrow (where the applicant is attempting to avoid controversy). The second is by using all the known available language tables, and calculating variants out of that list. The problem with this approach is that there is strong evidence that operating registries are not keeping up to date the language tables that they have deposited with IANA (see <http://tools.ietf.org/html/draft-levine-tld-variant-06> for evidence of this).

An expert reviewed the applied-for IDN labels (and the existing delegated ones) for variants, using the criteria mentioned above. The expert found 18 collisions<sup>3</sup> among

---

<sup>3</sup> The expert's report says there are 31 collisions, but actual count shows 18.

applications and existing delegated strings; of these, two are “real” variant collisions. The real cases are both pairs of applications by the same applicant, presumably in an effort to simulate variants by obtaining two delegations. (The cases are noted at <http://www.icann.org/en/news/announcements/announcement-01mar13-en.htm>). The String Similarity Review Panel received the resulting report; the resulting collisions are listed in Appendix B: Collision report. It is worth noting that the expert prepared the report on the basis of existing language tables, with the attendant difficulties noted above. The expert in question, however, is indubitably an expert in Unicode, so one might regard the exercise as a sort of cut-down version of the LGR procedure: the language tables submitted represent the “generation panel” output, and the expert represents the integration panel. There remains a risk, however, that the LGR will eventually treat as variants strings that today are not. Such treatment would only present a real problem if the “new” variant under the LGR had been allocated. It is of course impossible to say how likely this is without knowing what the LGR procedure will yield.

**The staff response assumes that that the process of delegating IDN labels can be arrested (perhaps temporarily) after the current applications are processed and until an LGR is completed. What are the risks if that turns out not to be possible, and there is an additional request for an IDN TLD before the rules for the relevant code points are complete?**

Each additional allocation of a label increases the risk that there will be a collision between that label once the LGR is eventually released. This much is self-evident.

An important question is what sort of risk is run when this happens. There are three types of risk.

The first is that a label will be allocated without the associated variant being detected. As long as that variant happens not to be requested by anyone else, this is really just a problem in theory: the contention set is populated incorrectly, but since nobody asks for the variant label in question, it does no real harm. The probability of this coming to pass appears to increase according to the number of new allocations, but the consequences are small.

The second is that a variant label that needs to be allocated is not, because the LGR is not complete. In effect, this is the very condition that caused people to propose variants in the first place. If we believe that users will be confused or otherwise disadvantaged because of this, then the consequences of this are exactly as bad as we think that user confusion is bad. Given the empirical evidence from the current new gTLD round, we should evaluate the risk of this as moderately high: the worked example we have apparently includes variants. It is hard to know how to extrapolate from this single instance into the future.

The final and worst risk is that two labels will be allocated to different operators when they should not have been, because they are variants of one another. This case presents two problems. First, its direct effect is potentially confusing experiences to users, who think they are using the same name but are led to

different Internet resources through that name. The entire premise of any variant system is that the user cannot tell which string is being used; if we believe that premise, the effect of this risk is to make IDNs with possible variants unsuitable for use as any kind of identifier. Second, it undermines the effectiveness of the LGR over the long term, because it will need to be handled as an exception. It is extremely difficult to know how likely this is to happen without knowing what the LGR itself is to be; but the consequences are quite bad, both for user experience and for the LGR itself. One possible bit of evidence is that nobody has objected to any application on these grounds. One would expect complaints if there were possible variant collisions of the sort contemplated here.

A different kind of risk ensues if ICANN simply refuses all new IDN applications until the LGR is ready. In that case, ICANN will be subject to political pressures, both external and internal, to abandon the LGR procedure and to render a decision. Since we already observe this happening, the risk of this is evidently high. The consequence is more activity of the sort represented by this report; the reader can judge whether that would be harmful.

**What are the consequences if IDNs are delegated prior to the LGR being completed, and when the LGR is completed the results do not match the delegations? Consider both the effects of a label not having been allocated when it should have been under the LGR, and the effects of a label having been allocated when it should not have been under the LGR. Is it possible to assign probabilities to these possibilities?**

These consequences are outlined in the response to the previous question. It does not seem possible to assign probabilities to the possibilities, because it is not possible to determine in advance what the LGR will contain and it is not possible to know how long it will take to develop. Indeed, if we knew in advance what the LGR would contain, we would not need to develop it.

**What is the risk that there are variants in the current (or future) TLD applications that are variants of some IDN label?**

Since some variants have been identified in the list of applications, the risk is 100% for current TLD applications. It is hard to predict what the possibility will be for future TLD applications; but if the LGR is in place and working before such applications are dealt with, there is no issue.

**The staff response to item 4 (about the code point repertoire) critically depends on the evaluation of the new gTLD applications by the DNS Stability Panel. The deliberations of that panel are not public. Did it actually include the issues of the code point repertoire in its deliberations? Also, did it include the issue of future constraints on ICANN's action established as precedent by decisions in this gTLD round? If it did not, what risks here might be unanalyzed?**

This issue boils down to a question of whether the DNS Stability Panel considered the issues of the effect of the LGR and the implied future constraints that ICANN will face as the result of precedent established by any label allocation.



Unfortunately, the AGB does not offer complete guidance here. While it says that the Panel may “determine whether any strings raise significant security or stability issues that may require further review,” that determination is about “security or stability of the DNS.” So, the entire discussion hinges on whether the DNS Stability Panel believed that its evaluation of string use was a question of the security or stability *of the DNS*. The answer is by no means obvious: it could be that the actual use of a given label is a security or stability problem on the Internet, but not a security or stability problem for the DNS as such (i.e. narrowly construed as relevant only to the protocol).

Another instruction to the DNS Stability Panel is also interpretable as being strictly limited to the operation of the DNS itself: “The panel will determine whether the string fails to comply with relevant standards or creates a condition that adversely affects the throughput, response time, consistency, or coherence of responses to Internet servers or end systems, and will report on its findings.” This passage suggests that the question is only whether [DNS] responses (towards other systems, like end systems or Internet servers) are affected. It appears to be possible that the panel might have determined that a given string is a bad idea on systemic reasons, but that it did not affect the criteria listed there.

In keeping with the principle of conservatism, we must assume that the DNS Stability Panel restricted itself to the narrowest interpretation, which means that the DNS Stability Panel’s conclusions cannot be used to show that the IDN applications contained no problematic code points. We do not know whether they did as much. The Panel has been contacted, and if its confidentiality obligations permit it may provide further guidance on this question; as of this writing, it has not been able to provide that guidance.

**Regardless of the outcome of (4), are there risks to ICANN or its future secure and stable co-ordination of the root zone that come from proceeding with delegation of IDN TLDs without the deterministic rules the LGR are supposed to provide?**

As long as ICANN’s decisions are subject to political pressure (and they manifestly are), ICANN’s technical function may always be undermined by political considerations. The LGR is intended to provide a mechanism to maximize the influence of independent and disinterested technical advice, and to minimize the influence of non-technical considerations when making narrow technical decisions.

Without deterministic rules, precedents will be established according to other principles. Any such precedent inevitably undermines the authority of the deterministic rules. In this case, there is a danger that such a precedent would not be consistent with the LGR. Moreover, the undermining of one set of deterministic rules may provide an argument for why other deterministic rules might be attacked. In general, therefore, acting without deterministic rules presents an overall risk to ICANN’s ability to co-ordinate the root zone. This risk is in proportion to the chance that a given allocated label turns out to be inconsistent with the eventually

published LGR. Without actually producing the LGR, it is not possible to predict how likely such an event is.

**Are there risks that come from delegating IDNs where there is a claimed variant (but refusing to delegate or allocate the variant)? What if the variant later turns out to be denied by the LGR?**

The argument in favor of any allocated variant is that for some class of users, the variant is in all practical respects the same label as some other label. To the extent someone claims a variant should be allocated along with an applied-for label, therefore, the original applied-for label ought not to be allocated without the variant.

Assuming the LGR procedure works as planned, there is no negative effect to the Internet if both the original application and the claimed variant are not allocated (or at least not delegated) until the LGR can be used to determine their disposition. There may, of course, be negative effects to the applicant and its business model.

There is a risk to ICANN in attempting to hold names in a claimed variant set until such time as the LGR is complete, and that risk is that ICANN will come under insuperable political pressure to do something before the LGR is ready.

**Are there risks that labels that are delegated are variants of each other and not detected as such because of the absence of the LGR?**

Yes. The LGR is conceived to work code point by code point. As a result, it tends to generate at least some variants that are obscure, and which will probably be overlooked until the LGR is complete. These variants, however, are much more likely to be blocked than allocated. A relationship that resulted in variants that ought both to be allocated (or delegated) would almost certainly attract notice. So, while the risk is high, the consequences appear to be low.

**Are there risks to ICANN in attempting now to deviate from its previously-announced plans, even if that deviation is inspired by learning something new about IDNs or variants?**

Obviously there are. Commercial entities have by now spent considerable effort and money on their TLD applications. Failing to cleave to announced plans could affect (negatively) some applicants, who would naturally be inclined to initiate legal action to attempt to retrieve some of their investment.

**Are there risks to ICANN in attempting not to deviate from its previously-announced plans, even if that deviation is indicated by learning something new about IDNs or variants?**

If ICANN is unable to deviate from prior announcements even if it learns of technical considerations of which it was previously unaware, that suggests that ICANN needs to be infallible when making announcements in the first place. Since it is unlikely that ICANN will be infallible, continuing on a course of action when holding newly-discovered information that militates against that course of action risks bringing ICANN into disrepute as a responsible, secure and stable operator. In the present

case, since ICANN's point in establishing the Variant Issues Project was to uncover issues it had not previously understood, it would appear to be somewhat improvident to continue with an old plan of action in the face of new understanding.

**Are there risks not directly touched on in the Jennings paper or the staff response, or in previous decisions, that should be considered? For instance, most of the discussion about allocation of two variants has assumed well-intentioned actors and lack of external interference; what if opponents to ICANN or to another applicant attack the system and prevail in a lawsuit, thereby creating legally-mandated exceptions to the LGR? What if this happens in more than one jurisdiction, with different results?**

If the LGR itself is a failure – because, for example, courts of competent jurisdiction require different treatments of the same label – ICANN's capability to operate IDN labels will be undermined. That fact itself suggests that a comprehensive, deterministic set of rules is an important tool for ICANN to minimize risk to itself.

Similarly, attempts by bad actors making TLD applications are more likely to be foiled by a comprehensive set of rules than a set of rules narrowly tailored to the applications actually in hand. It is not clear what the motivation for such a bad actor would be, however, given the cost of applying for a new TLD.

One of the great difficulties in making any decision on this matter is the uncertainty of what the LGR will eventually contain. The LGR procedure requires an overarching conservatism, but even under those terms there is room for variation among the panels. If the LGR procedure yields a comparatively small number of code points for allocation, and a large number for blocking, then the chances of any string that is actually delegated conflicts with any other string that is actually delegated seems low. Users may be surprised that some things don't work, but they will not treat two different active names as the same thing when they are different. The greater the number of allocatable code points the LGR produces, the greater the risk that there could be a collision.

This discussion is arranged around IDNs, but if the LGR indicates variant relationships between IDNs and standard LDH-labels (in effect, if the LGR permits any variants in Latin script), the conclusions above apply to all new TLD applications. In that case, the risk is proportional to the likelihood that a given LDH-label is related to a U-label (most likely by virtue of the omission of diacritics in the LDH-label). The consequences of allocation in this case are exactly the same as the consequences of any other allocation of one part of a variant bundle without the other parts.

## Appendix A: Review terms of reference

Terms of reference for reviewing "IDN TLDs: An endeavour fraught with risk" and the staff response

The ICANN Board of Directors, and in particular the Board Variant Working Group (BV-WG), have received a paper written by Dennis Jennings, "IDN TLDs: An endeavour fraught with risk". In that paper, Dr Jennings argues that the current new gTLD round is premature in its consideration of Internationalized Domain Name labels to be placed in the root zone.

ICANN staff prepared a response to Dr Jennings's paper, arguing in effect that the risks are all already known, and that they have either been mitigated effectively, or that they are in the process of being mitigated. The BV-WG has asked that an external reviewer consider the arguments on both sides, and present an evaluation. These are terms of reference for such a reviewer.

The purpose of this review is not to determine a "winner" of the debate, but to highlight risks that the ICANN Board will take if it continues to pursue its current stated goal of prioritizing the delegation of new IDN TLDs.

Following are questions that the review should consider. It is possible that some of them will be unanswerable; to the extent they are, that may itself present a risk worth investigating. The review must be conducted with all deliberate speed, because it is in effect a review of risks of action that has already been initiated. This scope of work depends on terms in common use in the Variant Issues Project and follow-on root zone IDN work. The reviewer is expected to be familiar with those projects, as well as the Jennings paper and staff response.

The review should include, to the extent knowable, an evaluation of the degree of risk and evaluation of the magnitude of risk: how likely is something to come to pass, and how bad would it be if that thing did come to pass?

1. Is it possible to know with any certainty that some IDN labels will never have any allocatable variants prior to the latest Label Generation Rules procedure being completed (at least for some subset of Unicode; for brevity, in what follows this is called "LGR")? If so, to what degree?
  
2. Is it possible to know with any certainty, prior to the LGR being completed, the variants of those IDN labels actually in the current round of new gTLD applications? If so, describe the labels and the degree to which one can be certain.

3. The staff response assumes that that the process of delegating IDN labels can be arrested (perhaps temporarily) after the current applications are processed and until an LGR is completed. What are the risks if that turns out not to be possible, and there is an additional request for an IDN TLD before the rules for the relevant code points are complete?
  
4. What are the consequences if IDNs are delegated prior to the LGR being completed, and when the LGR is completed the results do not match the delegations? Consider both the effects of a label not having been allocated when it should have been under the LGR, and the effects of a label having been allocated when it should not have been under the LGR. Is it possible to assign probabilities to these possibilities?
  
5. What is the risk that there are variants in the current (or future) TLD applications that are variants of some IDN label?
  
6. The staff response to item 4 (about the code point repertoire) critically depends on the evaluation of the new gTLD applications by the DNS Stability Panel. The deliberations of that panel are not public. Did it actually include the issues of the code point repertoire in its deliberations? Also, did it include the issue of future constraints on ICANN's action established as precedent by decisions in this gTLD round? If it did not, what risks here might be unanalyzed?
  
7. Regardless of the outcome of (4), are there risks to ICANN or its future secure and stable co-ordination of the root zone that come from proceeding with delegation of IDN TLDs without the deterministic rules the LGR are supposed to provide?
  
8. Are there risks that come from delegating IDNs where there is a claimed variant (but refusing to delegate or allocate the variant)? What if the variant later turns out to be denied by the LGR?
  
9. Are there risks that labels that are delegated are variants of each other and not detected as such because of the absence of the LGR?
  
10. Are there risks to ICANN in attempting now to deviate from its previously-announced plans, even if that deviation is inspired by learning something new about IDNs or variants?

11. Are there risks to ICANN in attempting not to deviate from its previously-announced plans, even if that deviation is indicated by learning something new about IDNs or variants?

12. Are there risks not directly touched on in the Jennings paper or the staff response, or in previous decisions, that should be considered? For instance, most of the discussion about allocation of two variants has assumed well-intentioned actors and lack of external interference; what if opponents to ICANN or to another applicant attack the system and prevail in a lawsuit, thereby creating legally-mandated exceptions to the LGR? What if this happens in more than one jurisdiction, with different results?

## Appendix B: Collision report

```

CheckCollision: 38 labels have no variants.
CheckCollisions: 9418 collisions detected.
=====
Reporting conflict for application: ccTLD-03, String <U+4E2D,
U+56FD>
..... conflicting with application: ccTLD-04, String <U+4E2D,
U+570B>
Reporting conflict for application: ccTLD-04, String <U+4E2D,
U+570B>
..... conflicting with application: ccTLD-03, String <U+4E2D,
U+56FD>
Reporting conflict for application: ccTLD-15, String <U+0627,
U+06CC, U+0631, U+0627, U+0646>
..... conflicting with application: ccTLD-16, String <U+0627,
U+064A, U+0631, U+0627, U+0646>
Reporting conflict for application: ccTLD-16, String <U+0627,
U+064A, U+0631, U+0627, U+0646>
..... conflicting with application: ccTLD-15, String <U+0627,
U+06CC, U+0631, U+0627, U+0646>
Reporting conflict for application: ccTLD-23, String <U+067E,
U+0627, U+06A9, U+0633, U+062A, U+0627, U+0646>
..... conflicting with application: ccTLD-24, String <U+067E,
U+0627, U+0643, U+0633, U+062A, U+0627, U+0646>
Reporting conflict for application: ccTLD-24, String <U+067E,
U+0627, U+0643, U+0633, U+062A, U+0627, U+0646>
..... conflicting with application: ccTLD-23, String <U+067E,
U+0627, U+06A9, U+0633, U+062A, U+0627, U+0646>
Reporting conflict for application: ccTLD-28, String <U+0627,
U+0644, U+0633, U+0639, U+0648, U+062F, U+064A, U+0629>
..... conflicting with application: ccTLD-29, String <U+0627,
U+0644, U+0633, U+0639, U+0648, U+062F, U+06CC, U+0629>
..... conflicting with application: ccTLD-30, String <U+0627,
U+0644, U+0633, U+0639, U+0648, U+062F, U+06CC, U+06C3>
..... conflicting with application: ccTLD-31, String <U+0627,
U+0644, U+0633, U+0639, U+0648, U+062F, U+064A, U+0647>
Reporting conflict for application: ccTLD-29, String <U+0627,
U+0644, U+0633, U+0639, U+0648, U+062F, U+06CC, U+0629>
..... conflicting with application: ccTLD-28, String <U+0627,
U+0644, U+0633, U+0639, U+0648, U+062F, U+064A, U+0629>
Reporting conflict for application: ccTLD-30, String <U+0627,
U+0644, U+0633, U+0639, U+0648, U+062F, U+06CC, U+06C3>
..... conflicting with application: ccTLD-28, String <U+0627,
U+0644, U+0633, U+0639, U+0648, U+062F, U+064A, U+0629>
Reporting conflict for application: ccTLD-31, String <U+0627,
U+0644, U+0633, U+0639, U+0648, U+062F, U+064A, U+0647>
..... conflicting with application: ccTLD-28, String <U+0627,
U+0644, U+0633, U+0639, U+0648, U+062F, U+064A, U+0629>
Reporting conflict for application: ccTLD-39, String <U+53F0,
U+7063>
..... conflicting with application: ccTLD-40, String <U+53F0,
U+6E7E>

```

..... conflicting with application: ccTLD-41, String <U+81FA,  
U+7063>  
Reporting conflict for application: ccTLD-40, String <U+53F0,  
U+6E7E>  
..... conflicting with application: ccTLD-39, String <U+53F0,  
U+7063>  
Reporting conflict for application: ccTLD-41, String <U+81FA,  
U+7063>  
..... conflicting with application: ccTLD-39, String <U+53F0,  
U+7063>  
Reporting conflict for application: 1-1120-42188, String <U+7F51,  
U+7AD9>  
..... conflicting with application: 1-2101-67873, String <U+7F51,  
U+7AD9>  
Reporting conflict for application: 1-1121-17301, String <U+5E7F,  
U+4E1C>  
..... conflicting with application: 1-1309-35206, String <U+5E7F,  
U+4E1C>  
Reporting conflict for application: 1-1159-3507, String <U+7F51,  
U+5740>  
..... conflicting with application: 1-994-51307, String <U+7F51,  
U+5740>  
Reporting conflict for application: 1-1254-85868, String <U+70B9,  
U+770B>  
..... conflicting with application: 1-1254-86222, String <U+9EDE,  
U+770B>  
Reporting conflict for application: 1-1254-86222, String <U+9EDE,  
U+770B>  
..... conflicting with application: 1-1254-85868, String <U+70B9,  
U+770B>  
Reporting conflict for application: 1-1309-35206, String <U+5E7F,  
U+4E1C>  
..... conflicting with application: 1-1121-17301, String <U+5E7F,  
U+4E1C>  
Reporting conflict for application: 1-1313-58483, String <U+5FAE,  
U+535A>  
..... conflicting with application: 1-950-28485, String <U+5FAE,  
U+535A>  
Reporting conflict for application: 1-1491-83816, String <U+5A31,  
U+4E50>  
..... conflicting with application: 1-963-13166, String <U+5A31,  
U+4E50>  
Reporting conflict for application: 1-2101-67873, String <U+7F51,  
U+7AD9>  
..... conflicting with application: 1-1120-42188, String <U+7F51,  
U+7AD9>  
Reporting conflict for application: 1-2102-26509, String <U+7F51,  
U+5E97>  
..... conflicting with application: 1-858-36255, String <U+7F51,  
U+5E97>  
Reporting conflict for application: 1-858-36255, String <U+7F51,  
U+5E97>  
..... conflicting with application: 1-2102-26509, String <U+7F51,  
U+5E97>



Reporting conflict for application: 1-868-7047, String <U+4FE1,  
U+606F>  
..... conflicting with application: 1-995-44061, String <U+4FE1,  
U+606F>  
Reporting conflict for application: 1-940-43388, String <U+76DB,  
U+8D38, U+996D, U+5E97>  
..... conflicting with application: 1-940-75591, String <U+76DB,  
U+8CBF, U+98EF, U+5E97>  
Reporting conflict for application: 1-940-75591, String <U+76DB,  
U+8CBF, U+98EF, U+5E97>  
..... conflicting with application: 1-940-43388, String <U+76DB,  
U+8D38, U+996D, U+5E97>  
Reporting conflict for application: 1-950-28485, String <U+5FAE,  
U+535A>  
..... conflicting with application: 1-1313-58483, String <U+5FAE,  
U+535A>  
Reporting conflict for application: 1-963-13166, String <U+5A31,  
U+4E50>  
..... conflicting with application: 1-1491-83816, String <U+5A31,  
U+4E50>  
Reporting conflict for application: 1-994-51307, String <U+7F51,  
U+5740>  
..... conflicting with application: 1-1159-3507, String <U+7F51,  
U+5740>  
Reporting conflict for application: 1-995-44061, String <U+4FE1,  
U+606F>  
..... conflicting with application: 1-868-7047, String <U+4FE1,  
U+606F>  
Conflicts reported for 31 application  
=====