

## **DENIC's Comments on ICANN's Guidelines for the Implementation of Internationalized Domain Names, Version 2.0**

DENIC, the registry for the Top Level Domain .de, takes the opportunity to comment on ICANN's Draft Guidelines for the Implementation of Internationalized Domain Names, Version 2.0, albeit DENIC is obviously not subject to any ICANN guidelines on IDN in the first place.

However, it is of high importance not only for the registries within ICANN's sphere of influence but also for the whole global Internet community and, with respect to ICANN's legitimacy and reputation, for ICANN itself that ICANN does not attempt to set rules that are, to say the very least, questionable.

Regardless of the actual substance of the Draft IDN Guidelines and the assessment of that substance, the simple fact that such guidelines are being issued by ICANN require one very elementary statement in the beginning:

It is unclear why the issuance of IDN guidelines dealing with linguistic aspects and aiming at avoiding "the deceptive use of visually confusable characters from different scripts in individual IDN labels" would fall under ICANN's mission as defined in Article I Section 1 of the ICANN bylaws: According to these bylaws, ICANN's mission is "to coordinate, at the overall level, the global Internet's systems of unique identifiers, and in particular to ensure [its] [...] stable and secure operation". The issues dealt with in the IDN guidelines as well as the goal of dealing with them are clearly not matters of the technical "operation" of the Domain Name System (DNS) but attempt to protect users of said system against criminal activities, linguistic confusion, and a certain degree of carelessness of some of them. If this was regarded as being in the remit of ICANN's policy-making there were effectively no limits to such policy-making at all.

Furthermore, while it is a thankworthy improvement that, different from version 1.0, the Draft IDN Guidelines do now, in the beginning, explicitly state that only gTLDs are subject to these guidelines, DENIC regrets that the subsequent text still addresses "top level domain registries" in general.

Another improvement in comparison to version 1.0 of the guidelines is the – yet neither complete nor consequent – shift from putting the emphasis on languages and "language tags" to focussing on scripts. That is because making languages the foundation of any policy is highly problematic in that not even linguists would, due to the flexibility and perpetual change of languages, be able to deal

with such policies. A more detailed explanation of the downsides of language-related Internet policies is given in Appendix G to the Unicode Technical Report (UTR) #36 ("Unicode Security Considerations").

For this very reason, it is regrettable that the Draft IDN Guidelines still mention "language tags" two times and keep referring to languages. At the same time, it becomes clear from this that the guidelines, even though the authors seem to have realized that referring to languages is problematic, do not entirely abandon such a concept but rather merely try to smoothen its disadvantages. It is obvious that such an approach will rather create, not solve problems and severely impairs the implementability of the guidelines.

In general, the rules laid out in the Draft IDN Guidelines tend to be stricter than those in version 1.0. Naturally, this means that some domain names were compliant with version 1.0 but will not be compliant with version 2.0 anymore, and some of these domains might already have been registered. One would expect that the guidelines would include some indication as to how to deal with this situation.

With respect to the different concrete rules contained in the Draft IDN Guidelines, DENIC further comments as follows:

#### Section 3. (a)

The reference to UTR #23 ("The Unicode Character Property Model") as an authority for script designators is infelicitous; instead, reference should be made to UTR #24 ("Script Names"). Additionally, leaving the choice of use between ISO 15924 and UTR #24 is not a very good idea. UTR #24 itself states: "ISO 15924 is aimed primarily at the bibliographic identification of scripts; consequently it occasionally identifies varieties of scripts that may be useful for book catalog[u]ing, but which are not considered distinct scripts in the Unicode Standard.". In light of this, UTR #24, not ISO 15924, would be the appropriate reference.

It is superfluous, if not counterproductive to require associating script designators with domain names, since script, naturally, is an attribute of each of the code points composing a label anyway. Requiring registrants/registrar to send script designators along with domain names and the registry to store them is an error-prone guideline in that it results in duplication of information that generally should be avoided.

#### Section 3. (c)

Simply stating that "all code points in a single label must be taken from the same script" betrays that not enough thought has been given to the underlying

issue in that it ignores that there are code points that are being used in different scripts without being part of them. This applies, for instance, to the code points that belong to the "Common" and "Inherited" scripts such as the hyphen-minus which is part of the former. Consequently, the requirement that all code points in one label have to stem from the same script would rule out the usage of the hyphen-minus in labels that otherwise consist of Latin script (like in "example-online.net"). If some rule was felt necessary in this instance, it would be more appropriate to give a definition of an algorithm that determines whether a label is mixing scripts or not (as, for example, in Appendix D to UTR #36).

Besides, the "all code points from one script" rule gets subsequently undermined by opening room for nearly limitless exceptions when the guidelines state: "Exception to this is permissible for languages with established orthographies and conventions that require the commingled use of multiple scripts". Apparently, ICANN has realized that there are far from a few legitimate cases for mixing scripts in a single domain label, but still doesn't want to abandon the "all code points from one script" rule altogether, nor is able to give clear instructions on which are acceptable exceptions. This, once again, underlines that the "all code points from one script" rule itself is ill-conceived and its validity being overridden.

Similarly, the following rule that "visually confusable characters from different scripts must not appear in a single label" immediately gets devalued by an incomprehensible and broad exception when it is stated that this rule does not apply when "there are overriding legitimate linguistic reasons". In fact, anybody could claim their reasons are overriding and it remains unclear who would finally make the decision whether they really are. Furthermore, a clear and workable definition of "visually confusable" code points is missing and with that, visual confusability is already a matter of font size and type.

Seemingly, the reason for banning the mixing of scripts as well as the usage of visually confusable code points is to avoid spoofing. However, it would be naïve to think this was a solution to the problem.

Indeed limitations to mixing scripts might make a repetition of the paypal.com incident less likely, but still it does not help against in-script spoofing (where all characters in one label belong to the same script) or whole-script spoofing (where a whole label from a single script spoofs another label from a different script). UTR #36 gives further details on this. Also, no rules on the composure of domain names will ever help against conceptually confusable labels that the users regard as materially the same (such as "paypal.com" and "paypal-online.com").

On top of that, the DNS is a decentralized system so that a registry has no control mechanisms at all to enforce a rule to not mix scripts or to not allow visually confusable code points beyond the level of registration. Thus spoofing remains possible on the lower levels.

In light of this, it is imperative to understand that the solution for phishing and scamming cannot be provided by registration policies. In fact, already the attempt to solve such problems by registration policies is likely to give registrants and users a false sense of security which would, in the end, let them act incautiously and with that worsen the problem.

#### Section 4. (c)

In comparison to version 1.0 of the IDN Guidelines, Section 4. (c) now is less strict and classifies punctuation characters as permissible code points, as long as they have "grammatical significance in the language with which the IDN registration is associated". This new rule, however, brings up two serious issues: Firstly, it is incoherent with the previous position of moving away from language association with domain names. From that, the question derives: What if a domain name has no language association, since it is now an optional attribute? Secondly, it again opens a door for exceptions with a confusing term like "grammatical significance". Providing two examples for that is simply not sufficient.

#### Section 4. (d)

Under version 1.0 of the Guidelines, spacing characters were forbidden, which was a clear and practicable rule. Now section 4. (d) instead bans "other characters with well-established functions as protocol elements". With this, more room for interpretation and confusion is opened: What is an established and what a well-established function? And which protocols are we talking about? And who decides in the end, applying which criteria?

#### Section 4. in general

To become coherent with the "inclusion-based" approach recommended in the Draft IDN Guidelines, section 4. should not attempt to ban the usage of certain yet not clearly defined code points and, at the same time, create equally vague exceptions from that ban. Rather, reference should be made to whitelists of permissible characters as provided by appropriate and competent technical bodies, such as the syntax rules defined in Unicode Standard Annex #31 ("Identifier and Pattern syntax") or, focussing on security, the "IDN Security Profile for Identifiers" (Appendix A to UTR #36).

#### Section 5.

Section 5. defines the scope of an IDN registration in terms of both, its Unicode and ASCII-encoded representations (even though an ASCII-encoded representation is always Unicode, too). However, the text then states that the "registry should treat the ASCII-encoded form as the primary registered name". In this instance, it remains unclear what the classification as "primary registered name", presumably opposed to a "secondary registered name", is supposed to mean and which (technical) consequences it should result in. Besides, it is

obvious that registrants and users will always see the IDN form as the actual domain name and often not even know that there is an ACE form behind it. Accordingly, in the event of changes of the IDNA standard, registrants will not be impressed if being told that they still hold the same ACE form which just, unfortunately, now translates into a slightly different IDN.

Administrative details

RFC 2234 has been obsoleted by RFC 4234.

Frankfurt, 21 October 2005

DENIC eG