

## **IDN Implementation Guidelines (IDNG) Working Group (WG)**

### **Notes from Meeting on 1 June, 2017**

#### Meeting Attendees (in alphabetical order)

WG members:

1. Edmon Chung
2. Dennis Tanaka
3. Kal Feher
4. Mats Dufberg

Staff:

5. Sarmad Hussain

#### Meeting Notes

The WG members continued to discuss the public comments received, based on the PC summary circulated to the WG members.

1. **Comments O-X1 - O-X3.** The WG decided to park these points and come back to them later, after reviewing other comments.
2. **Comment IAB2.** It was discussed that this remark supports the WG recommendations and no further action is needed.
3. **Comment IAB3.** WG discussed it was not clear what IAB meant as the comment spans six different guidelines. It was suggested that IAB3 is asking to make a general statement, but it may be better to not include general and guiding statements because such statements are not actionable for the contracted parties which have to comply with these guidelines. It was suggested to follow up with IAB for further clarification. It was also summarized that the comment is generally supportive of the recommendations by the WG.
4. **Comment IAB4.** It was suggested that the last paragraph be combined with the previous one. It was brought up the recommendation 13 is very long and needs to be reviewed. Other option would be to delete the last paragraph. WG was reminded that the extra text was included to ensure that a conservative view is taken for registry side activation. It was shared that if the reference text is to be included, it is better that it point to specific sections in documents.

WG agreed to split the recommendation, e.g., into 13a. and 13b. The example could be moved to the footnote or possibly an appendix, with Arabic example added, in addition to the Chinese. The WG will review the second paragraph at a later point, when addressing the relevant comment.

5. **Comment IAB5.** It was pointed that the RFCs in the IDNA2008 protocol say that registries should have policy that may restrict use of certain characters. It was suggested that a pointer to the relevant RFC be added. WG was not clear what is meant by “fully understand”. Did it mean that registries needed to have a linguist on board for that purpose? The statement likely refers to

section 3.2. of RFC 5894, that character repertoire be changed after understanding the implications, which WG agreed is a reasonable goal. It was also suggested to look at Inclusion principle from RFC 6912, in this context, section 4.2. These statements will be collected and all members are to review the statements and then discuss this further next time.

6. **Comment Gol1.** WG noted the comment and also noted that WG has made specific recommendations for whole-script confusables.
7. **Comment Gol2.** WG noted it is a broad comment. WG noted its response that this is addressed, i.e. in recommendation 17 second part. It was suggested to divided 17 into two parts.

Action Items

S. No.	Action Items	Owner
1	<i>Reach out to IAB to get clarification on IAB3.</i>	EC
2	<i>Based on IAB4, split recommendation 13 into two parts. Add Arabic languages as an example and review the second paragraph when the relevant comment is addressed.</i>	
3	<i>For IAB5, all are requested to review the statements from the RFCs. The relevant sections be identified and circulated to the WG for further discussion.</i>	All, SH
4	<i>Revise 17 by splitting it into separate recommendations.</i>	

**Notes from Relevant RFCs 5890, 5891,5894 and 6912**

RFC 5890.

**2.3.2.3. Internationalized Domain Name and Internationalized Label**

An "internationalized domain name" (IDN) is a domain name that contains at least one A-label or U-label, but that otherwise may contain any mixture of NR-LDH labels, A-labels, or U-labels. Just as has been the case with ASCII names, some DNS zone administrators may impose restrictions, beyond those imposed by DNS or IDNA, on the characters or strings that may be registered as labels in their zones. Because of the diversity of characters that can be used in a U-label and the confusion they might cause, such restrictions are mandatory for IDN registries and zones even though the particular restrictions are not part of these specifications (the issue is discussed in more detail in [Section 4.3](#) of the Protocol document [[RFC5891](#)]). Because these restrictions, commonly known as "registry restrictions", only affect what can be registered and not lookup processing, they have no effect on the syntax or semantics of DNS protocol messages; a query for a name that matches no records will yield the same response regardless of the reason why it is not in the zone. Clients issuing queries or interpreting responses cannot be assumed to have any knowledge of zone-specific restrictions or

conventions. See the section on registration policy in the Rationale document [[RFC5894](#)] for additional discussion.

#### 4.4. Visually Similar Characters

To help prevent confusion between characters that are visually similar (sometimes called "confusables"), it is suggested that implementations provide visual indications where a domain name contains multiple scripts, especially when the scripts contain characters that are easily confused visually, such as an omicron in Greek mixed with Latin text. Such mechanisms can also be used to show when a name contains a mixture of Simplified Chinese characters with Traditional ones that have Simplified forms, or to distinguish zero and one from uppercase "O" and lowercase "L". DNS zone administrators may impose restrictions (subject to the limitations identified elsewhere in these documents) that try to minimize characters that have similar appearance or similar interpretations.

If multiple characters appear in a label and the label consists only of characters in one script, individual characters that might be confused with others if compared separately may be unambiguous and non-confusing. On the other hand, that observation makes labels containing characters from more than one script (often called "mixed-script labels") even more risky -- users will tend to see what they expect to see and context is a powerful reinforcement to perception. At the same time, while the risks associated with mixed-script labels are clear, simply prohibiting them will not eliminate problems, especially where closely related scripts are involved. For example, there are many strings that are entirely in Greek or Cyrillic scripts that can be confused with each other or with Latin script strings.

It is worth noting that there are no comprehensive technical solutions to the problems of confusable characters. One can reduce the extent of the problems in various ways, but probably never eliminate it. Some specific suggestions about identification and handling of confusable characters appear in a Unicode Consortium publication [[Unicode-UTR36](#)].

RFC 5891

#### 4.3. Registry Restrictions

In addition to the rules and tests above, there are many reasons why a registry could reject a label. Registries at all levels of the DNS, not just the top level, are expected to establish policies about label registrations. Policies are likely to be informed by the local languages and the scripts that are used to write them and may depend on many factors including what characters are in the label (for example, a label may be rejected based on other labels already registered). See the Rationale document [[RFC5894](#)], [Section 3.2](#), for further discussion and recommendations about registry policies.

### 3.2. Registration Policy

While these recommendations cannot and should not define registry policies, registries should develop and apply additional restrictions as needed to reduce confusion and other problems. For example, it is generally believed that labels containing characters from more than one script are a bad practice although there may be some important exceptions to that principle. Some registries may choose to restrict registrations to characters drawn from a very small number of scripts. For many scripts, the use of variant techniques such as those as described in the JET specification for the CJK script [[RFC3743](#)] and its generalization [[RFC4290](#)], and illustrated for Chinese by the tables provided by the Chinese Domain Name Consortium [[RFC4713](#)] may be helpful in reducing problems that might be perceived by users.

In general, users will benefit if registries only permit characters from scripts that are well-understood by the registry or its advisers. If a registry decides to reduce opportunities for confusion by constructing policies that disallow characters used in historic writing systems or characters whose use is restricted to specialized, highly technical contexts, some relevant information may be found in [Section 2.4](#) (Specific Character Adjustments) of Unicode Identifier and Pattern Syntax [[Unicode-UAX31](#)], especially Table 4 (Candidate Characters for Exclusion from Identifiers), and [Section 3.1](#) (General Security Profile for Identifiers) in Unicode Security Mechanisms [[Unicode-UTS39](#)].

### 4.2. Inclusion Principle

Just as IDNA2008 starts from the principle that the Unicode range is excluded, and then adds code points according to derived properties of the code points, so a public zone should only permit inclusion of a code point if it is known to be "safe" in terms of usability and confusability within the context of that zone. The default treatment of a code point should be that it is excluded.