Refutation of listed arguments for 3-character minimal length of IDN gTLDs

The Explanatory Memorandum "Discussions about the 3-char String Requirement" of contains a series of mistaken arguments in support of requiring 3 "distinct characters" for each gTLD string. (Page 5 on http://icann.org/en/topics/new-gtlds/three-character-30may09-en.pdf)

I submit a refutation of each of these. For easer reference, I copy each mistaken argument as quoted text followed by the refutation.

As all the arguments brought forward to date are false, the proper way forward is to adapt the Draft Applicant Guidebook. One solution is to make the minimum length dependent on script. ICANN should also allow the respective language communities to request a minimum string length based on considerations specific to the underlying languages or script.

REFUTATION OF ARGUMENT 1: "Fairness of treatment"

The following statement is false:

> "In addition to comments received from the CJK community, ICANN received arguments from the European region that certain single or two character combinations in European languages represent a word or a meaning and in some cases also geographic identifiers. These arguments were made to counter arguments for allowing less than 3 character strings per the proposal above. If less than 3 character strings are allowed for CJK based languages then they should also be allowed for other languages, for fairness of treatment."

Proposing "equal" treatment in terms of uncomparable measurements grossly unfair. By analogy, imagine a food rationing system where a set amount of food were alloted for to each building, irrespective of how many people lived there.

To make it even clearer, imagine for a moment that the DNS had always been in Chinese ideographs, and that we Westerners were asking for ASCII. How would we feel if we were told that the domains should be at least 3 separate words? Or if we were told that all characters used should be composed of a least 5 separate brush strokes? Or that a TLD should require at least 6 keystrokes? (Most Chinese and Japanese two-letter words require over 7 keystrokes to type.)

The fact that a small number of word exists in European languages that are composed of just characters (or one) does no change this. In any non-ideographic language, there are very few words composed of just 2 characters. By contrast, most words in CJK that expressing a generally understood concept are composed of just two ideographs, and at least 1000 often used Chinese and Japanese words are just one character. As for Korean, words _appear_ to a Westerner as one or two characters, but in reality these are syllable blocks composed of multiple Jamo characters.

REFUTATION OF ARGUMENT 2: "Statements concerning the Chinese words and number of characters"

The following argument is false:

> "Other arguments state that few Chinese characters are words, most Chinese words are two or more characters. If one separates out the specific phonetic implications (one character equals one phoneme), characterizing Chinese characters as syllables would be much more accurate. In addition, some opinions are that most Chinese words consist of more than one character."

In Chinese almost all characters correspond to one syllable, but this does not change the fact that each ideograph has a meaning as a word. Most ideographs can be used alone and in combination. It is only logical that there are more word combinations than single words: in this respect all languages share the same feature. In Japanese, each ideograph has at least 2 different readings and generally at least one of them is pronounced as two or more syllables. In Korean, ideographs are rarely used today as most writing is in Hangul, a phonetic script. The latter indeed appears as blocks of syllables. However, each block that appears as a Hangul "character" to a Westerner is actually a block of two or more individual phonetic characters.

The overwhelming majority of words one can find in a Chinese, Japanese or Korean dictionary are composed of what appears as one or two "distinct characters" to a Westerner. These are actually compound words, i.e. combinations of full words. It is impossible to see how one could justify the imposition of a Western-style "abbreviation" by requiring more characters than the full word!

REFUTATION OF ARGUMENT 3: "Trial of a few gTLDs with less than 3 characters"

The discussion document reports the following suggestion:

> "Some suggestions have been made that ICANN perform a trial implementation of a certain small number of gTLDs that have less than 3 characters. This would then be used to inform the development of the process for allocating such strings more widely."

There are no stability issues to address with such a trial. Requiring such a trial would be a deliberate and malevolent act of discrimination.

REFUTATION OF ARGUMENT 4: "Translations of TLDs"

The following argument is fundamentally misguided:

> "Comments have been received that in relation to translation of existing TLDs, there has never been a model for "translating" TLDs. Therefore, the two-letter ISO codes or other TLDs that are “abbreviations” cannot be translated into IDN strings of less than 3 characters as a meaningful representation of that definition. They are not standardized abbrevia-

> tions and abbreviations are not a standard concept across languages and cultures. The ccTLDs in particular are a standardized coding system, chosen as codes for a number of reasons including recognizability and distinctiveness of undecorated Latin character."

There is no reason why IDN TLDs should be "translations" of Western (or any other) expressions, and there is no need for them to be "abbreviations" in the Western sense.

The proposed Chinese TLD .公司 (gongsi) means "business" or "corporation". It is present in most registered company names in China. It is neither a translation of .com, nor a translation of .biz. And it should not be. It is a full word, more than an abbreviation. If two characters give us a complete word (remember, they require 6 keystrokes to type), why would anyone pretend that it should be lengthened to 3 characters?

About language-to-language translations of TLDs (or other names), I might add that they tend to be either hilarious or confusing. They could hardly be of serious use. A generalized translation-based approach to TLDs is utterly impossible.

REFUTATION OF ARGUMENT 5: "ICANN's ccTLD delegation function"

The following concern does not exist:

> "Currently the IANA delegation function relies on the scarce availability of 2-character ASCII combinations and all of these (when entered in the ISO3166-1 list) are treated as ccTLDs. Discussions in the ICANN community as well as at ISO MA meetings in the past few years have focused on the feasibility of expanding the ISO3166-1 list to contain 2-character combinations of other scripts, representing country and territory names. This would require a multi-year ISO led development (which might occur after the ccNSO IDN PDP). The outcome of the Fast Track Process will inform the ongoing discussion about whether or how to expand the ISO3166-1 list and associated ICANN ccTLD delegation function, as well as the long-term ccNSO PDP for IDN ccTLDs. Delegation of single and two-character labels now, might jeopardize the future shape of the ccTLD delegation mechanism."

It is natural to extend the ISO-3166-1 list with corresponding country *names* and *well-known abbreviations* in all languages. But names and abbreviations are not *codes*. It would be a very bad idea to create more than list of standard codes for the same things.

The ISO3166-1 list is of codes is based on the basic 25-letter Latin alphabet. This is the only character set present in all the computer systems and all education systems around the world. This means that ISO3166-1, as a list of codes, is addressable in all systems around the world. There is no need for a translation of a code into news codes. There is a need for translation from code to *names* in specific languages. The ISO-3166-1 list is being updated in this respect and will eventually contain country names and abbreviation (not codes) in all major languages and in the respective scripts.

It must also be stated clearly that it is not possible to "translate" a code from one script to another. Translation is possible from language to language and from code to language. For instance, it is possible to translate the name "USA" from English to "EEUU" in Spanish or to 美国 (měiguó = "beautiful country") in Chinese or "米国" (beikoku - "rice country") in Japanese. The code "US" can thus also be translated into each of these language-specific names. But it there is not way to create, for instance, a "code" that translates "US" into ideographic script: any such attempt would lead to confusion.

There can thus only be mapping of ISO-3166-1 from code to all names in all major languages (as opposed to a mapping from script to script). Names cannot be restricted to 2 "letters". It has already been shown that IDN ccTLDs cannot be limited to 2 letters. For instance, the ideographic name of Singapore, 新加坡 (Xīnjiāpō) cannot be shortened to 2 characters.

The mapping codes to names and well-known abbreviations in major languages is a compilation effort. It is not standardization, as the country names exist independently of the standards or documentation authority. This compilation effort is already protected in the gTLD process as country names may only be used with the approval of the relevant government.

As a result, there is no imaginable standardization or compilation effort whatsoever that could possibly be "jeopardized" by allowing 2-character or 1-character ideographic gTLDs.

Werner Staub