Report on Chinese Variants in Internationalized Top-Level Domains

This report considers the issues relating to the Chinese (Han) script variants being represented as multiple different labels in the Domain Name System, with respect to how they should function in use as internationalized top-level domain names.

1 Introduction

The delegation and management of IDN variant TLDs is an important issue. Language communities that use variant characters are affected by decisions about how variants are managed and implemented in new IDN TLDs. This is of concern in both IDN gTLD and IDN ccTLD implementations.

To develop potential solutions for the delegation of IDN variant TLDs, the ICANN Board, in its 2010 meeting in Norway, directed the CEO to "develop an issues report identifying what needs to be done with the evaluation, possible delegation, allocation and operation of IDN gTLDs containing variant characters, as part of the new gTLD process in order to facilitate the development of workable approaches to the deployment of gTLDs containing variant characters IDNs."

After consultation with the community, ICANN outlines its approach to move forward by first undertaking work to identify issues associated with the beneficial and safe delegation of IDN variant TLDs through six case studies (Arabic, Chinese, Cyrillic, Devanagari, Greek, and Latin).

Subsequently, the Chinese case study team was formed to complete this task. It comprises thirteen experts in linguistic, DNS and IDNA, security, policy, and registry/registrar operations. This report is the Chinese case study team's report to ICANN.

The report is organized as follows: Firstly, we provide the reader with some context about the Han script. In section 2, we briefly introduce the Han script and its use in Chinese, Japanese and Korean. In sections 3 and 4, we define the terms used in the document as well as the scope of this paper. In sections 5 and 6, we enumerate different types of Chinese variant, and the associated user expectations. In sections 7, 8 and 9, we explore issues related to Chinese variants in the internationalized top-level domains. Specifically, section 7 discusses issues related how Chinese variants are represented in IDN tables at the top-level, and section 8 uses a domain lifecycle approach and compiles issues related to the evaluation, allocation, delegation and operation of Chinese Variant TLDs. Finally, to ensure important issues are not missed, section 9 examines the impact of Chinese variant TLDs on different stakeholders.

The report has three appendixes. In Appendix A, we list the relevant Chinese team members, their statements of interest as well as acknowledgements. In Appendix B, we list the blocked code points. In Appendix C, we provide a brief overview of RFC 3743 and RFC 4713.

2 Introductions to the Han Script

The Han script consists of a character set including tens of thousands of characters, of which several thousand characters are in common use for each language using it. It is used to write Chinese languages, ancient and modern, in simplified and traditional forms. It is also used to write Japanese (Kanji) in addition to two other scripts (Hiragana and Katakana) and Korean

(Hanja) in addition to the Hangul script.

Chinese Hanzi, Japanese Kanji and Korean Hanja are often referred to as ideographs. An ideograph is a *graphic* symbol that represents an idea. Since 1990, tens of thousands of Chinese Hanzi, Japanese Kanji and Korean Hanja have been merged into "CJK Unified Ideographs" and their Extension in ISO/IEC 10646 and Unicode.

2.1 Chinese Hanzi

Chinese Hanzi 漢字/汉字 (U+6F22[1] U+5B57/U+6C49 U+5B57) originated from pictographs. They are pictures that evolved into ideographs over a period of several thousand years. For instance, the ideograph for "hill" 山 (U+5C71) still bears some resemblance to the three peaks of a hill.

Not all Hanzi are pictographs. There are other classifications such as compound ideographs and phonetic ideographs. For example, 'endurance' 忍 (U+5FCD) is a pierced 'knife' 刀 (U+5200) above the 'heart' 心 (U+5FC3).

Hence, almost every Hanzi is associated with some meaning by itself, which is very different from most other scripts that are based on alphabets.

Since the Hanzi unification in the Qin dynasty (221-207 B.C.), the most important change in the Chinese Hanzi occurred in the middle of 20th century when more than two thousand Simplified Hanzi were introduced as official forms in Mainland China.

Hanzi simplification was conducted in several stages that are summarized in three tables published by the National Language Committee PRC in 1964 and 1986. In a nutshell, these three tables are:

- Table 1 defines 350 simplified Hanzi which are to be used by themselves, and can never serve as "simplified components." Some examples are 'treasure' 寶 (U+5BF6) is simplified to 宝 (U+5B9D); 'factory' 廠 (U+5EE0) is simplified to 厂 (U+5382).
- Table 2 lists 132 simplified Hanzi that can be used either by themselves as well as a component to further simplify other Hanzi. It also lists 14 components (known as "radicals") that cannot be used by themselves, but can be used to simplify other complex Hanzi. Some examples of the first category include 'horse' 馬 (U+99AC) is simplified to 马 (U+9A6C), 'bird' 鳥 (U+9CE5) is simplified to 鸟 (U+9E1F), 'dragon' 龍 (U+9F8D) is simplified to 龙 (U+9F99). Examples for the second category (simplified radicals) include 言 (U+8A01) is simplified to 讠 (U+8BA0) ('the idea of talking').
- Table 3 lists 1,753 Hanzi which are simplified based on the simplified components (or radicals) defined in Table 2. For example, using the examples in table 2, 'drive' 駕 (U+99D5) is simplified to 驾 (U+9A7E), 'duck' 鴨 (U+9D28) is simplified to 鸭 (U+9E2D), 'deaf' 聾 (U+807E) is simplified to 聋 (U+804B), 'lesson' 課 (U+8AB2) is simplified to 课 (U+8BFE).

---

[1] In this report, the Unicode characters are identified by their positions, or code points. The notation U+6F22, for example, indicates the character at the position 6F22 (hexadecimal) in the Unicode table.

- In addition to the above manners, simplified Hanzi take the simpler form variants.

It is very important to note that *not* all Hanzi were given a new simplified form, because some of the as these un-simplified (or traditional) Hanzi were already very "simple" or not frequently used. Thus, although the 'Comprehensive Table of Simplified Hanzi' is non-exhaustive, there are only 2244 simplified Hanzi, and very few new simplified Hanzi have been added since 1986.

As a result, the Chinese language has two writing systems: Simplified Chinese (SC) and Traditional Chinese (TC). Both systems use the same script but are expressed using different subsets under Unicode definition of the same Han script.[2] The two writing systems use SC and TC respectively while sharing a large common "unchanged" Hanzi subset that occupies around 60% in contemporary use. The common "unchanged" Hanzi subset enables a primarily simplified Chinese user to understand texts written in traditional Chinese with little difficulty and vice versa. The Hanzi in SC and TC have the same meaning and the same pronunciation and are typical variants.

Today, Simplified Chinese is used in Mainland China and Singapore as the official form and is increasingly used in other areas, whereas the Traditional Chinese is the official form used in Taiwan, Hong Kong, and Macau. Singapore, Malaysia and many other overseas Chinese communities are using TC or SC or even a mixture of SC and TC in their daily communications.

Finally, although most relationships between Simplified Chinese and Traditional Chinese are 1-to-1, this is not always the case. The Chinese Domain Name Consortium, when developing the Chinese IDN character table, identified 184 instances where the simplified form of multiple traditional Chinese characters is the same. For example, traditional characters 發 (U+767C) and 髮 (U+9AEE) have the same simplified form 发 (U+53D1), and a handful of SC to TC mappings are one to many, depending on the context.

2.2 Japanese Kanji

Japanese 'Kanji' 漢字 (U+6F22 U+5B57) were imported from China and used as ideographic characters. According to JIS X 0208, there are more than 6,000 Kanji characters used in Japan, 2,000 of which are regularly used, as defined by Jōyō Kanji (常用漢字 U+5E38 U+7528 U+6F22 U+5B57). Among those 2,000 Kanji characters, some of them are in a simplified form (called "new character form" 新字体 U+65B0 U+5B57 U+4F53) derived from the traditional imported form (called "old character form" 旧字体 U+65E7 U+5B57 U+4F53), and they are recognized as variants in general context. However, to express proper nouns such as names of persons and places, the old character form and the new character form are recognized as different independent characters.

The simplification of old and new forms in Japanese was performed independently from those of Chinese; therefore, many of the mappings between old and new forms in Japanese are different from mappings between Simplified Chinese and Traditional Chinese forms.

Finally, in Japanese, all three scripts (Kanji, and the syllabaries Hiragana and Katakana) are used as main scripts. The Latin alphabet and Arabic numerals are also used as required. Both the Hiragana and Katakana scripts consist of syllabic characters. Although there are corresponding

---

[2] Note, some linguistics regard simplified Chinese and traditional Chinese as two scripts. For example, see <www.cjk.org/cjk/reference/cjkvar.pdf>

Hiragana and Katakana for each syllable, because of differences in usage, they are not recognized as interchangeable variants. Hiragana are typically used for grammatical endings and particles and for Japanese words which cannot be written in Kanji. Katakana are typically used to represent words borrowed from English and other foreign languages, and for onomatopoeia.

During the development of RFC 3743 by the Joint Engineering Team (JET), experts from Japanese Internet community had discussions and decided that, domain names are fundamental Internet identifiers, and can be thought of as having the same status as proper names. Therefore, for Japanese in IDN, it is appropriate to distinguish old and new forms as different independent characters instead of handling them as variants. This understanding has been reflected in the IANA IDN table developed by the .JP registry JPRS, in which no variants are identified for Kanji.

2.3 Hanja in Republic of Korea[3]

Chinese ideographs were imported into the Korean language as written forms. These Korean ideographs, known as Hanja 漢字/ 한자 (U+6F22 U+5B57/U+D55C U+C790), were widely used until recently, when Hangul 한글 (U+D55C U+AE00) (or Hangeul by ROK's Romanization rule) become more popular.

Hangul is a systemic script designed by a 15th-century ruler and linguistic expert, King Sejong 世宗 (U+4E16 U+5B97). It is based on the pronunciation of the Korean language. A Korean syllable is composed of Jamo 字母/자모 (U+5B57 U+6BCD/U+C790 U+BAA8) elements that represent different sounds. Hence, unlike Han ideographs, individual Hangul syllables do not have any meaning. Each Hanja ideograph can be represented by a Hangul syllable.

Today, Hanja is no longer widely used in ROK. A law[4] enacted on April 14th, 2011 orders all ROK official government documents to be written only in Hangul. Hanja or other scripts can only be written within parentheses if allowed by presidential decree. Though many Korean vocabularies are derived from Hanja, they are usually written in Hangul. Modern Korean is rarely written with mixed scripts like Japanese Kanji, Hiragana, and Katakana.

Members working on this issue report have not been able to confirm that Hanja or Han script variants do not matter to ROK speakers. However, staff from the Korea Internet & Security Agency, registry operator for ccTLD .KR, stated that its IDN policy does not allow Hanja as reflected by the language table it submitted to IANA[5], and they have no intention of allowing the use of Hanja in their domestic market.

2.4 Chinese Hanzi and Japanese Kanji in a TLD Context

As discussed in Section 2.2 above, in the context of a Japanese environment, there is no need for the consideration of IDN variants for Japanese Kanji domain names. However, unlike the case of ccTLDs like .JP where only Japanese IDN registrations are expected, both Chinese and Japanese IDN gTLDs can be expected in the root zone. Because of the unified approach to Han

---

[3] The Chinese team do not have representatives from The Democratic People's Republic of Korea (DPPK), and therefore would not be able to determine its use of Hanja. We note that DPPR has not participated in IDN related developments in the chinese domain name community.

[4] "Framework Act on the Korean Language" (2011) Available at <http://law.go.kr/LSW/lsInfoP.do?lsiSeq=112364#0000>

[5] See http://www.iana.org/domains/idn-tables/tables/kr_ko-kr_1.0.html

ideographs in Unicode, there is an overlap of the character repertoire between the Japanese and Chinese IDN scripts. More specifically, Kanji characters used in Japanese can also be found in Chinese. This means that it is possible that a Japanese domain name consisting only of Kanji characters can appear to be like a Chinese domain name for a Chinese user.

For example, in the Japanese environment, 'association' "学会.jp" (xn--6oqq31a.jp) and "學會.jp" (xn--n9so2y.jp) (or "学会.日本" (xn--6oqq31a.xn--wgv71a) and "學會.日本" (xn--n9so2y.xn--wgv71a)) would be considered two distinct second-level domain registrations and allowed to co-exist without any problems because the domain name itself already provided a contextual indicator that it is intended to be a Japanese domain name. However in the case of the root zone, ".学会" and ".學會" would be viewed by Chinese users around the world as variants TLDs.

We explore this issue in detail in section 6.3.

3 Chinese specific terminology

To ensure the various working groups were speaking the same vocabulary, the group was advised to adhere to specific terminology[6] where applicable. The following terminology was selected to augment those terms as required for this specific case study.

Han Script Variant

Characters with different visual forms but with the same pronunciations and with the same meanings as the corresponding official forms in the given language contexts. For more details please refer to the section 5 of this report.

Official Form

In different country/region, the government specifies "official forms" for a set of general use Hanzi. In Mainland China, they are called normalized Hanzi (规范字 U+89C4 U+8303 U+5B57), and in Taiwan, they are called orthographic Hanzi (正體字 U+6B63 U+9AD4 U+5B57).

CJK Characters

"CJK characters" are characters commonly used in the Chinese, Japanese, or Korean languages, including but not limited to those defined in the Unicode Standard as ASCII (U+0020 to U+007F), Han ideographs (U+3400 to U+9FAF and U+20000 to U+2A6DF and U+2A700 to U+2B73F and U+2B740 to U+2B81F), Bopomofo (U+3100 to U+312F and U+31A0 to U+31BF), Kana (U+3040 to U+30FF), Jamo (U+1100 to 11FF and U+3130 to U+318F), Hangul (U+AC00 to U+D7AF and U+3130 to U+318F), Kangxi Radicals(U+2F00 to U+2FDF), CJK Radicals Supplement (U+2E80 to U+2EFF), and the respective compatibility forms.

CJK Unified Ideograph

An ideograph is a graphic symbol that represents an idea. Chinese Hanzi, Japanese Kanji and Korean Hanja are often referred to as ideographs. Since 1990, tens of thousands of Chinese Hanzi, Japanese Kanji and Korean Hanja have been merged into CJK Unified Ideographs and their Extension in ISO/IEC 10646 and Unicode. In this document, if not

---

6 https://community.icann.org/download/attachments/16842778/Draft+Definitions.pdf

otherwise specified, the term “ideograph” means a CJK Unified Ideograph.

Language Variant Table

The key mechanism used by current domain registries for calculating Chinese variant labels is a three-column table, called a Language Variant Table, designated for each language permitted to be registered in the zone. Those columns are known, respectively, as “Valid Code Point,” “Preferred Variant,” and “Character Variant,” and are defined separately below. In this document, “LVT” and “Variant Table” are used as short forms for “Language Variant Table.” IANA maintains the list of variant tables for the Chinese Script. Appendix C provides some examples.

Valid Code Point

In a Language Variant Table, a “Valid Code Point” is an entry on the list of code points that is permitted to registered for that language. Any other code points, or any string containing them, will be rejected by this specification. The Valid Code Point list appears as the first column of the Language Variant Table.

Preferred Variant

In a Language Variant Table, a “Preferred Variant” is an entry on the list of code points corresponding to each Valid Code Point and providing possible substitutions for it. These substitutions are "preferred" in the sense that the variant labels generated using them are normally registered in the zone file, or “activated.” The Preferred Code Points appear in column two of the Language Variant Table. “Preferred Code Point” is used interchangeably with this term.

Character Variant

In a Language Variant Table, a “Character Variant” is an entry on the second list of code points corresponding to each Valid Code Point and providing possible substitutions for it. Unlike the Preferred Variants, substitutions based on Character Variants are normally reserved but not actually registered (or “activated”). Character Variants appear in column 3 of the Language Variant Table. The term “Code Point Variant” is used interchangeably with this term.

Zone Variant

A “Zone Variant” is either a Preferred or Character Variant Label that is actually to be entered (registered) into the DNS, that is, into the zone file for the relevant zone. Zone Variants are also referred to as Zone Variant Labels, active labels, or Activated Labels.

Internationalised Domain Label (IDL)

The term “Internationalized Domain Label” or “IDL” will be used instead of the more general term “IDN” or its equivalents. This is the string of characters of the domain name being applied for and has been validated as suitable for inclusion in the DNS zone file. In the case of an IDN TLD, the IDL is simply the string of characters of the TLD being applied for and has passed the evaluation.

IDL Package

An “IDL Package” is a collection of IDLs as determined by the guidelines in RFC 3743. All labels in the package are "reserved", meaning they cannot be registered by anyone other than the holder of the Package. These reserved IDLs may be "activated", meaning they are actually entered into a zone file as a "Zone Variant". The IDL Package also contains the language tag. The IDL and its variant labels form a single, atomic unit, however, not all

labels in the package are active.

Language Tag

Language tags, as defined in RFC 5646, are used to help identify languages, whether spoken, written, signed, or otherwise signaled, for the purpose of communication. This includes constructed and artificial languages but excludes languages not intended primarily for human communication, such as programming languages.

4 Scope of Work

TLD is the last label of the domain such as ".com" & ".org", which is the label before the last dot (“.”) in the whole domain. ICANN manages the top-level domains. Registrars are responsible for managing the second label, which is referred to as the second-level domain (SLD) and is the level at which most end users apply for registration.[7] Although TLD management strongly affects second- and lower-levels, this report focuses on TLD issues. This report still considers issues in lower levels and considers consistency between the top-level and lower-level registrations to be important to stability.

An applicant for an IDN TLD will abide by the policies and variant tables developed or certified by ICANN. However, operators of zones lower down the DNS tree have no ICANN imposed requirement to abide by the same variant tables and policies of those used at the top-level. For example, a character that is considered to be a variant of another character when handling (for allocation, delegation, and so on) an IDN TLD may not be considered as a variant when handling SLD and beyond.

Since inconsistent handling of domain labels confuses users (ordinary internet users, server administrators, and so on), service developers (including web browser developers), as well as domain registrants, not only administrators of TLDs but holders of other zones are encouraged to follow the same policies and variant tables (Assuredly, with some modifications if necessary).

From an Internet protocol perspective, the U-label can be a combination of a wide array of characters; and, significantly, a combination of code points with different script properties. The U-label is merely just a sequence of code points in “Normalization Form C” (NFC). The sequence does not need to form any word defined by any dictionaries. In essence, the label is not necessarily language/script-dependent. At the top level, the label usually has a very short sequence of code points associated with only one script. Hence, the label has no language context at the top level. Having a loose relationship between label and language at Internet protocol specification is very important. Language and meaning of words change over time without any predictable pattern and frequency, and separating language rules from TLD compositions and DNS resolutions allows new words and characters to emerge while also maintaining a domain name's stability when meaning of words change or characters are removed.

This report discusses variant issues in Han script, but given the fact that variant issues have rarely been raised by the Korean and Japanese language communities since the IDN service was launched, this report primarily focuses on Chinese issues. Our rationale is as follows:

- Although there may be variant issues in Japanese Kanji and Korean Hanja, these issues

[7] The last label applied for gTLD but not always applied for ccTLD since many manage the second last label as well, like “.ac.kr” and “co.jp”

are limited in scope as compared with the Chinese. For example, insofar as Chinese characters are likely to be used in Korea at all, Koreans mostly use traditional forms. Korean Hanja includes very few character variants. Japanese Kanji have new and old forms, but Japanese has defined "Jōyō Kanji" where the old forms are not part of the definition. There are some character variants in Kanji, but they are categorized as "Rarely Used Old Hanzi" (生僻字 U+751F U+50FB U+5B57), such as "囡" (U+5700) (a variation of "国" (U+56FD)). Japan has been using simplified character forms for many years, much longer than those forms have been official in China, and does not alternate them with traditional forms.

- As discussed in sections 2.2 and 2.3, current top-level domain operators from Japan and Korea either do not include variants in their IDN tables, or, in the case of Korea, simply do not allow Hanja characters at all in its IDN Table. Although these organizations do not fully represent their respective language communities, given the limited resources, and limited time available for this case study team to consider the issues, we respect these communities' choices and focus our effort on Chinese variant issues.

- The Chinese working group understands that Chinese variants *might*, however, have implications for Japanese and Korean users. Therefore, during the formation of the team, we sought to recruit experts from Japan and Korea. As a result, two well-known IDN experts, one from Japan (Yoshiro Yoneya) and one from Korea (Yangwoo Ko) are on the team. Where appropriate they identified potential implications for the Japanese and Korean community.

Finally, Unicode is the only character representation and encoding standard used and discussed throughout this report. Anything that is not included in the Unicode is considered out of scope for this report.

5 Chinese Variants

5.1 Types of Chinese Variants

Chinese (character) variants are:

> "characters with different visual forms but with the same pronunciations and with the same meanings as the corresponding official forms in the given language contexts.[8]"

Most importantly, variant characters do not share the same Unicode code points. Different spellings, presentation in multiple scripts, translations of a string, and transliterations of a string are not considered variants.

In Unicode, there are different code points presenting same characters but with different visual widths (i.e. half and full); however, this is not considered as issue because the full width and half

---

[8] During drafting of this definition, Chinese team members collected comments from linguistic experts in Mainland China, Taiwan, and Japan. We would like to acknowledge the following experts: Li Yuming, Wang Tiekun, Wang Ning, Zhang Shuyan, Eiji Matsuoka, C.C.Hsu, Wang Xiaoming, Shi Jianqiao, Zhang Guoqiang, Wang Cuiye.

width form encode Katakana, Hangul and alphabetic scripts only. Han script encodes in full width form only.

Homographic differences, where two characters are spelled the same but are pronounced differently and have different meanings, are not considered variants. In CJK, many common vocabularies use more than one word (or ideographic icons) and one of the word will have the homographic traits: for example, 会 (U+4F1A) can be used in 'meeting' 会议 (U+4F1A U+8BAE), and it can be also used in 'accountant' 会计 (U+4F1A U+8BA1). The character 行 (U+884C) can be used in 'walk' 行走 (U+884C U+8D70), as well as in 'bank' 银行 (U+94F6 U+884C). These would not cause confusion to native speakers, and are not generally considered as variants in the Chinese language.

In the Chinese language, there are two types of variant under the conditions described above.

The first type is created by regional variations in the standard writing system. As mentioned in section 2, there are now two common writing systems: Simplified Chinese and Traditional Chinese. Both writing systems use different subsets of the same Unicode Han script, and they are not mutually exclusive to each other. Although some academics, historians, linguists and regional policies may not consider differences between writing systems technically as variants, and even many end users will try to maintain their writing in one system only, there is no doubt that words presented under both writing systems signified the same words to native Chinese speakers. The study team considers this type of variant as the most important type of variant for ICANN and Internet communities to resolve. We explain this in detail in the section on end user expectations.

The second type is the generic variant. Several Chinese characters have another form that are slightly different visually, but are treated the same and have universal interchangeability. This interchangeability relationship is much stronger than the relationship between Traditional and Simplified form. Many of them are different in glyph and writing style. Some of them aren't considered as variants because they share the same code point in Unicode (for example: when applying different language tags), but some have 2 distinct code points. Some examples include "户/戶" (U+6237 / U+6236) and "黄/黃" (U+9EC4 / U+9EC3) which have two distinct code points for each pair. The code point difference is the result of the "Source Separation Rule" from Unicode, which states that if two ideographs are distinct in a primary source standard, then they are not unified.[9]

6 Consideration of User Expectations

As in section 5, variants are defined as "characters with *different* visual forms but with the *same* pronunciations and with the *same* meanings as the corresponding official forms in the given language contexts". In the Chinese IDN language table produced by CDNC (Chinese Domain Name Consortium), around 40% of characters have variant forms. In this section, we formulate the expectations coming from end users, and then summarize problems that may emerge if the Chinese variant TLDs were not delegated properly.

6.1 The Chinese Internet Users

The Chinese language community has embraced the Internet at an unprecedented rate. The

---

[9] *Unicode 6.0 Chapter 12 Unification Rules – R1 (pp 401)*

Internet users have surged from 620,000 in 1997 to 485 million by the end of June 2011. The number of Internet users who use Chinese has increased by 24.2% in 2011 alone.

During this process, the composition of Internet users has undergone a dramatic change as well. In 1998, 89% of Internet users in China had college-level English comprehension, while in 2011 only 22% have received such education. This means more and more Chinese Internet users are *only* familiar with the Chinese language environment.

6.2 Chinese User Expectations

Because Chinese variants have the same pronunciation and the same meaning as its official form, Chinese users regard them as interchangeable. Thus a variant IDN, derived from an IDN by replacing some characters with their variants, should match the original IDN.

For example, as mentioned earlier, Simplified Chinese is widely used in Mainland China and Singapore, while Traditional Chinese is commonly used in Hong Kong, Macao, Taiwan and other Chinese communities in Southeast Asia and other parts of the world in their daily communications. In a world where Simplified Chinese and Traditional Chinese are recognized as interchangeable, Chinese Internet users also expect to be able to access Chinese information with either Simplified Chinese characters or Traditional Chinese characters as they have experienced in their offline life regardless what the Unicode code point of the Chinese characters and their corresponding variants are on the DNS.

If Chinese variants are not permitted to be delegated at the top level, the coherent recognition of Chinese characters will be seriously jeopardized. *Delegating only one variant label of a Chinese gTLD* to an applicant will deprive the registry operator, Chinese domain name registrants, and Chinese users worldwide of the ability to use the other variant label, and, hence, impede their use of the Internet when navigating using Chinese domain names. So, if Chinese gTLD is delegated, either both SC and TC forms are delegated or neither is delegated to avoid the possible Internet user confusions.

If variant TLDs were allowed but they can be operated by two independent registries, this could cause a bad user experience, and potentially serious security issues. A user in Mainland China entering a domain name in SC could be directed to the service of one registrant, while another user in Taiwan entering what they perceive to be the same domain name in TC is directed to another service of a different registrant. This may create confusion for end users and would most likely create bad user experiences; it could possibly even invite phishing attacks. Hence, ensuring that both the SC and TC labels of a domain name are registered to the same registrant will avoid confusion to the end user.

From the experience of CNNIC, for SLD under synchronized TLD ".中国/中國" (.xn--fiqs8s / .xn--fiqz9s), over 10% of the DNS queries are for the traditional Chinese SLD labels. If the TC form is not delegated, then the user populations that make up the 10% of Chinese IDN queries will not be able to perform a lookup for a domain name represented in traditional form.

Finally, from a cultural perspective, permitting only one of the names may place some users of the same language, belonging to the same culture, at a significant disadvantage; worse still, it could lead to segregation and fragmentation of populations that are part of that language and cultural group.

6.3 Kanji-only TLDs

It is reasonable to expect that Kanji TLDs would be intended mainly for Japanese users. As

explained in Section 2.2 above, Japanese users do not generally consider and/or expect the need for IDN variants. Therefore, the team believes no variants need to be delegated for Kanji TLDs.

However, as explained in Section 2.4 above, in the global context of TLDs in the root zone, Chinese users may come across Kanji TLDs. Considerations of Chinese users are already described in Sections 6.1 and 6.2 above. More specifically, in the case of Kanji-only gTLDs, if a Chinese user cannot access a variant, at most s/he may conclude that the domain is not intended for her/him, however, problems could arise if the variant domain goes to a different entity. Abusive registrants might also exploit such situations for malicious purposes.

In consideration of the context of the root zone, and taking a conservative approach, when TLDs constructed using only Kanji (i.e. as available for Japanese TLDs but without Hiragana and Katakana characters), it is therefore appropriate to consider the Chinese Hanzi variant table to compile and reserve all variants based on such table (i.e. to generate all the variants based on the Chinese Hanzi table and consider both Preferred Variants and other variants as generated from Character Variants as reserved variants and not delegated). As an example, for a Japanese gTLD application: ".学会" (fully Kanji), a set of variants for the string could be generated based on the Chinese table, { ".学會"; ".學会"; ".學會"; ".孝会"; ".孝會" } and considered as reserved variants.

An example of such implementation can be found in Section 3.1 of the ".ASIA" CJK (Chinese Japanese and Korean) IDN Policies[10] and the ".ASIA" Japanese IDN table at IANA.[11]

6.4 Chinese ccTLD Experience

For SLD under synchronized TLD ".中国/中國" (.xn--fiqs8s /xn--fiqz9s), current registration number stands at around 320,000, among which over 77 percent have variant forms. The registration policy follows principles of RFC 3743; the language character table follows the CDNC language variant table. The registration of Chinese domain names under ".中国/中國" are bundled. The same conceptual Chinese domain name label under ".中国/中國", in Simplified and Traditional form respectively, must be delegated to the same registrant. The system executes the registration and ensures that the relationship is correct. To date, none disputes are filed concerning malicious use of variant domain names.

For SLD under synchronized TLD ".taiwan (.台灣/台湾)" (xn--kpry57d /xn--kprw13d), current registration number stands at around 40,000 among which over 83 percent have variant forms. The registration policy follows principles of RFC 3743; the language character table follow the CDNC language variant table. The registration of Chinese domain names under ".台灣/台湾" are bundled. To date, none disputes are filed concerning malicious use of variant domain names.

As of 1st September 2011, there are 24,450 ".hongkong (香港)" domain names registered, amongst which 85.5% have variant forms, the TC and SC variant labels for the same domain name are bundled. The registration policy follows RFC 3743. So far, the Hong Kong domain registry has received only 3 dispute cases for Chinese ".香港" domain names and they were all filed in 2008. The arbitration panels have rendered all decisions. There are no dispute cases for ".香港" domain names at the moment.

---

10 http://dot.asia/policies/DotAsia-CJK-IDN-Policies-COMPLETE--2011-05-04.pdf
11 http://www.iana.org/domains/idn-tables/tables/asia_ja_1.1.txt

6.5 Conclusion

Based on RFC 3743, the Chinese ccTLD experience and the considerations above, the following are Chinese users' priority expectations:

- The IDL and its Variant Labels SHOULD belong to the same registrant.
- The SC form and TC form of the applied-for IDL SHOULD be resolvable simultaneously or non-resolvable at all.

7 Issues Related to Language Variant Tables

7.1 The need for Chinese Language Variant Tables for the Root Zone

It is necessary for management of the root zone to consider language variant tables. The case study team considered several important scenarios:

a) ICANN allows applicants to submit a *Variant Table* together with their application

The problem with this approach is that ICANN will have to verify the accuracy of the Variant Table from each applicant. If the accuracy of the table is not verified, there could be serious issues with dispute resolution and security later on. The consequence of different table is that the applicant of the same IDN TLD in the same language may end up having different variants, or worst with the wrong or unintended variants.

b) ICANN allows applicants to submit the *variants of the TLD* with their application

The problem with this approach is that with no Variant Table, there is no technical mechanism to for ICANN to verify whether the requested variant is reasonable or accurate.

If ICANN chooses this approach, it should rely on the community, consulting with linguistic experts or during public comment period, to raise objections if the applications and its variants are not reasonable.

c) ICANN adopts a single unified variant table for the root zone

The problem with this approach is the difficulty for ICANN to adopt a single Variant Table. Already, there are multiple Variant Tables for the same language and locale within the IANA database. For ICANN to define and implement a process to adopt a single Variant Table may take years. However, this would result in the most consistent and predictable variants generation.

7.2 Whether IDN variants at TLD level should be based on language or script

Chinese ideographs are used by multiple languages, such as Chinese, Japanese and Korean.

If IDN variants at the TLD level were generated based on language, then there would be different variants (or no variants) of the same IDN TLD string depending on the language. This would result in the most accurate variant handling for Chinese script with least false positives.

If IDN variants at the TLD level are generated based on CJK script wide tables, then there would be same variants for the same IDN TLD string, regardless of the language. However, this would result in wider false positives, whereby there may be variants been reserved even though such variants may not be linguistically accurate and correct in the applied language.

Besides Chinese, the Japanese uses ideographs extensively. It is very likely that some Japanese IDN TLD applications may have ideographs. In the Japanese context, new and old form ideographs are considered as different strings. This is quite different from Chinese where Chinese considered Traditional and Simplified variants as the same. However, in discussions with the Japanese experts, they felt such false positives are acceptable especially at the gTLD or at the root level where a more conservative approach is needed.

For Koreans, the use of ideographs has been deprecated in favor of Hangul. Therefore, it is unlikely there will be Korean TLDs using ideographs. In the unlikely event that there is a Korean TLD application in ideographs, a similar approach can be applied.

Thus, we prefer variants to be generated based on language but, since this is apparently impossible at the root, we can accept a script based system.

As explained in Section 2, Han characters are currently used in both Chinese and Japanese communities but in different contexts. Therefore given the global nature of gTLDs and the root zone, it should be appropriate to consider the Han script for protection purposes (e.g. contention sets), and considering the Chinese and Japanese languages for usability purposes (e.g. delegation of gTLDs).

More specifically, in considering contention sets, regardless of whether a gTLD applicant indicates that it is applying for a Chinese or a Japanese TLD string, if the string applied for consists fully of Han characters (i.e. contains only Hanzi or Kanji), ICANN should utilize the Chinese Language Variant Table to produce the set of variants implicated by the applied for string. However, for delegation of gTLDs, ICANN should refer to the Language Tag declared by the applicant accordingly. That is, if the applicant declared that the application is for a Japanese gTLD, then only the applied for string should be delegated and all variants reserved. If the applicant declared that the application is for a Chinese gTLD, then the applied for string along with the Preferred Variant(s) should be delegated (and other variants reserved).

Furthermore, for Chinese gTLDs, unlike in the case of ccTLDs (e.g. for .CN or .TW) where one form of the Chinese characters is predominantly used, gTLDs in the root zone is inherently global in context and requires the consideration of an environment where some of the users will be using Simplified Chinese (e.g. in Mainland China, Singapore, etc.), while others may be using Traditional Chinese (e.g. in Hong Kong, Macau, Taipei, etc.). Therefore, it is necessary to consider both the Chinese table used by .CN and the Chinese table used by .TW together in the compilation of Preferred Variants for gTLDs. In summary, for Chinese gTLDs, the following 3 sets of TLDs should be delegated together:

1. Applied-for IDL

2. Preferred Variant in Simplified Chinese (based on zh-CN table)

3. Preferred Variant in Traditional Chinese (based on zh-TW table)

In most cases for the context of gTLDs, the above 3 sets should likely result in two TLD labels, and in some cases one TLD label. The reason that the result should likely result in two TLD labels to be delegated is because it is very likely that the Applied for string (1.) would be either fully in Simplified Chinese (2.) or fully in Traditional Chinese (3.). And sometimes, a TLD string would happen to be the same for both Simplified Chinese and Traditional Chinese, and therefore resulting in just 1 TLD string to be delegated (e.g. the experience of resulting in 1 TLD

string to be delegated is observed for the ".香港" (hongkong) and ".新加坡" (Singapore) IDN ccTLD fast track strings).

7.3 Considerations for a process to define the root Variant Tables

a) May have more than one Variant Tables for the root

If ICANN decide to handle variants based on language, it is possible there are multiple Variant Tables, for different language and for each locale/country.

b) Reference to Standards defined by an International recognized Standard Body, such as Unicode Consortium or ISO/IEC JTC1/SC2.

The major collections of CJK Unified Ideographs and its Extension A containing about 26,000 characters are in Basic Multilingual Plane (BMP). The remaining CJK Extension B, C, and D containing 44,000 characters are allocated in the Supplementary Plane.

In total, there are more than 70,000 characters in CJK Unified Ideographs and its Extensions. Supporting the entire set of characters may be difficult for devices (e.g. mobile phones) with limited resources.

In addition, many of these characters were defined for academic study or historical preservation. Not all characters are used in normal modern communication. Allowing all characters to IDN TLD would also increase the complexity in managing variant relationship.

One of the recommendations, International Ideographs Core (IICORE), from Ideographic Rapporteur Group (IRG) recommends a subset consisting of less than 10,000 characters.

c) Relying on long established practices within the ICANN community, such as Chinese Domain Name Consortium (CDNC)

The issues of Chinese variants have been extensively discussed within the ICANN community over numerous years. In particularly, the CDNC has been working on this for many years.

d) Calling upon industry and linguistic experts

ICANN should consult independent industry and linguistic experts to define and verify the Variant Tables before they are adopted by ICANN. Such expert group could come together on a regularly basis to review and to update the Variant tables.

e) IANA to maintain the Variant Tables

ICANN should maintain (i.e. on the IANA website) an authoritative Variant Tables for the root, available for anyone to access, verify and use.

7.4 The standard for Variant Tables

The earliest work on variant tables for the Chinese script is RFC 3743 for Chinese, Japanese and Korean domain name registration guideline. It comprises on various concepts for variant handling, such as bundling, atomic IDL Package, and reserved variants. It also defines a standard table as well as an algorithm to generate preferred variant and reserved variants.

The Chinese community, building on the concepts and principles of RFC 3743, defined RFC

4713, which is a more specific recommendation for Chinese domain name registration and administration.

RFC 4290 also derives its basic principles and concepts from RFC 3743 and attempts to make a generic recommendation for internationalized domain name registration. It retains many of the concepts without going into the specifics of the algorithm for generation of variants.

8 Issues Related to Evaluation, Allocation, Delegation and Operation of Chinese IDN Variant TLDs

In this section we enumerate a set of issues relating to evaluation, contention, allocation, delegation, and operation of IDN variant TLDs.

8.1 Evaluation Issues (when ICANN evaluate the TLD application)

8.1.1 Evaluation Panels

Evaluation involves different aspects of the new gTLD application such as string similarity, technical and financial readiness, and DNS stability. While some of these evaluation steps can be automated or do not require specific knowledge and experience, there are aspects that require human judgment, such as determining the extent of similarity between two strings or handling an applied for IDL that contains a Chinese character not listed in the LVT. It is difficult to judge these aspects properly without the assistance and advice of experts in different areas. While members of the Technical Panel will evaluate the technical and financial readiness, and impact on DNS stability by the applied-for TLD, it is desirable that experts with knowledge and experience of linguistics, especially the Chinese language, be included in the Panel to look at string similarity, identify conflicts with geographic names, and verify the list of discovered variant labels for the applied for IDL.

Apart from the processes and procedures documented in RFC3743, the Chinese domain name community has accumulated over time a great deal of knowledge and experience in handling variant issues related to registration of Chinese domain names at the second or lower level. The handling of Chinese character variant is a well-understood and practiced process among the members of the Chinese Domain Name Consortium (CDNC). CDNC and its members are ready to offer their knowledge and experience in formulating and documenting the process to handle variants for Chinese TLDs.

String similarity review is a core part of the String Evaluation process. This review has to determine whether the applied-for IDL is similar to one or more of the applied-for or delegated IDN TLDs to such an extent that confusion or security issues will be caused. It is important to note that this review may involve considerations beyond variants. Similarities can be visual, phonetic, or semantic. Also, there may be a need to look at related variant tables in addition to the LVT for this language. This review requires research work and human judgment, both of which are difficult to automate and difficult for people unfamiliar with the Chinese script. It is desirable to ensure that some members in the Evaluation Panel will have the knowledge and expertise to conduct string-similarity review for Chinese IDLs.

In the ICANN Draft Applicant Guidebook, the 4th paragraph of section 2.2.1.1 “String Similarity Review” says “… similarity review will be conducted by an independent String Similarity Panel.” Because existence of variant labels is also a factor to be considered, it is advisable that the String

Similarity Panel has members who have relevant expertise on Chinese variant.

The selection process for members of the different evaluation panel(s) should be clearly documented. In particular, the eligibility for panel members, the process to nominate and select panel members, and the scope and responsibilities of panel members has to be included. In addition to the Chinese-speaking communities, the process should address whether experts from other communities using the Han script should be included on the Panel(s).

8.1.2 Discovery of Chinese Variant Labels

Purposes of the String Reviews are to ensure that the applied-for gTLD is not identical or confusingly similar to an existing TLD and to identify if the applied-for gTLD is confusingly similar to another applied-for gTLD. With the possible existence of variant labels for an applied-for gTLD in Chinese, the String Reviews will have to cover all variant labels in the IDL package of the applied-for gTLD, not just the gTLD itself. It is advisable that ICANN will proactively identify the IDL packages of all applied-for gTLDs in Chinese because the applicant may not have provided the variant labels of the applied-for gTLD, or he may provide incomplete or incorrect variant information.

It is advisable to have an automated process for the discovery of Chinese Variant Labels pertaining to an applied for IDL to minimize the likelihood of human error and to ensure consistency of the evaluation process. For the SLD Chinese IDNs and synchronized IDN TLDs, the Chinese IDN community has been using the methodology based on RFC3743 (discussed in Appendix C) and the Language Variant Table (LVT) developed by CDNC (Chinese Domain Name Consortium) . To ensure that there is a consistent treatment for all applications for Chinese IDN TLDs, it may be worth adopting one single unified LVT for IDLs in the Han script.

Section 1.3.2 of the Draft Applicant Guidebook states that each applicant is expected to provide the "top level IDN tables" for the language or script for the applied-for gTLD. This may lead to inconsistency when evaluating applications for gTLDs with the same language or script. This is because the top level IDN tables provided by different applicants may be different and may contain wrong entries. It is worth considering establishing just one top level IDN table for the Han script. A process for creating and administering the "unified top level IDN tables" will have to be formulated.

Although an automated process would lead to consistent results, it is advisable that the list of discovered Variant Labels be verified and confirmed by members of at least one of the Panels for the Evaluation phase.

Point c under the 4th paragraph of section 1.3.3 of the draft applicant guidebook says that "Applicant submits an application for a gTLD string and does not indicate variants to the applied-for gTLD string. ICANN will not identify variant strings unless scenario (b) above occurs." This is not desirable because this may lead to the possibility of multiple gTLDs being approved by ICANN which are in fact variant labels of one another.

If the applicant has specified variant labels that are not in the discovered IDL Package for the applied for IDL, the applicant should be informed of the discrepancy. The applicant should then clarify whether he would still like to apply for the same labels or not. The application can proceed only if the applicant agrees to adopt the IDL Package as discovered during the Evaluation phase.

If any of the discovered variant labels is the same as one of the labels in the IDL Package of an existing delegated Chinese IDN TLD, the IDL will not pass the Evaluation. If any of the discovered variant labels is the same as one of the labels in the IDL Package of the Chinese IDN TLD applied for in another new gTLD application, there is a string contention. Different scenarios should also be considered to ensure that all possibilities are covered.

Finally, section 2.2.1.4 of the draft applicant guidebook is on "Geographic Names Review". Because an applied-for gTLD may have variant labels, consideration for geographic names has to be extended to all labels in the IDL Package of the applied-for gTLD.

8.1.3 Technical and Financial Readiness

The Evaluation phase also has to address the technical and financial readiness of the applicant in terms of handling variant TLDs. The applicant should be aware of and ready to address issues arising from having to administer additional TLDs which are Preferred Variant Labels of the applied for TLD. These include providing sufficient system capacity and facilities to handle the possible multiple zones for both the applied-for IDL and the Preferred Variant Labels, providing system functions to enforce the registration policies that apply to the different labels in the IDL Package, reserving the Character Variant Labels from registration by other registrants, and providing appropriate support for additional EPP extensions and additional information to be displayed by WHOIS.

The applicant should also be ready to handle the administration and management of the LVT for their IDN TLDs. Adding or deleting characters from the LVT will require regeneration of the IDL packages of existing CDN registrations and communication with the registrars and registrants. Consideration is required on publication of changed LVTs and informing the public to ensure that applicants previously interested in registering Chinese domain names including one or more of the added characters are informed and given the opportunity to act with appropriate priority. An appeal process may have to be in place for applicants or registrants who are affected by changes in the LVT.

8.1.4 Extended Evaluation

All issues stated for the Initial Evaluation also apply to the Extended Evaluation.

8.2 Contention/Objection/Dispute Issues (when ICANN rules that there is a contention for the applied-for gTLD, or when objections or disputes are received for the new TLD)

String contention, objection filing, and dispute resolution are also depicted as distinct phases of the new gTLD application lifecycle. Since it is expected that the issues are similar for these three phases, they are grouped together under this issue report. Issues related to variant for these phases should also be clearly documented.

The presence of IDN variants will lead to more scenarios to be considered when deciding whether there is a string contention.

- What if the Application A is in contention with the variant of Application B?
- What if variant of Application A is in contention with variant of Application B?
- What if A and B of the above two cases happen to be in different languages?
- What if the variant is a minor case that none of applicants care about?

As defined in section 1.1.2.10 of the Draft Applicant Guidebook (DAG), string contention applies

only when there is more than one qualified application for the same or similar gTLD strings. With the possible existence of variant labels for the applied-for gTLDs, it is advisable to extend the scope "same or similar gTLD strings" to "same or similar gTLD strings, or gTLDs whose IDL packages have labels which are identical or similar to one another".

Similarly, the definition of "contention set" as in the 5th paragraph of section 1.1.2.10 of the DAG is impacted by the presence of variant. It is advisable to expand the existing definition "Groups of applied-for strings that are either identical or similar are called contention sets" to "Groups of applied-for strings that are either identical or similar, or whose IDL packages have labels which are identical or similar, are called contention sets"

IDN variants should be taken into account in Objection Procedures, particularly in "string confusion objection" and "legal right objection." Where a rights holder files a legal rights objection, should the "variant" forms of its marks, names, or signs in IDN characters on which the objection is based be considered and accepted? For example, the World Intellectual Property Organization (WIPO) "世界知识产权组织" (U+4E16 U+754C U+77E5 U+8BC6 U+4EA7 U+6743 U+7EC4 U+7EC7) may object to an application for "世界知識產權組織" (U+4E16 U+754C U+77E5 U+8B58 U+7522 U+6B0A U+7D44 U+7E54), which differ in substituting Traditional characters for equivalent Simplified ones.

8.3 Allocation Issues (when ICANN decides which string(s) is/are to be allocated for the new TLD)

It is advisable that the whole IDL Package that passes the Evaluation be allocated. The fact that all labels in the IDL Package are allocated does not mean that all labels in the IDL Package have to be put into the zone file (i.e. activated). As defined in RFC 3743, only the Original Label and Preferred Variant Labels should be inserted into the zone file and the Character Variant Labels should be reserved. The allocation of the whole IDL Package will prevent future applications involving IDLs that are Variant Labels of this application from passing the Evaluation.

8.4 Delegation Issues (when ICANN decides whether the applicant is the right authority to be delegated the new TLD string(s) and whether there is a fee involved)

It is important that all considerations on delegation of Chinese variants within the new gTLD process be clearly documented in the respective process document. Delegation involves decisions on whether only some or all labels in the IDL Package should be put into the root zone, whether the whole IDL Package will be delegated to one applicant or the IDL Package will be split and delegated to different applicants, the type of agreement to be struck between the applicant and ICANN, and whether an application fee or quarterly fee will be levied and whether those fees are based on IDL Package rather than individual variant labels. In particular, ICANN should take into account concerns of the Chinese communities that it is not acceptable for a label in traditional Chinese to be delegated while the same label in simplified Chinese is not. [IDNA2008].

8.4.1 Delegation of Variant Labels

There are issues related to the usability and the number of Variant Labels to be considered when determining whether only some or all Variant Labels should be put into the root zone. Although all Variant Labels in the IDL Package are considered equivalent to the applied-for IDL, not all of them will be used in practice. The LVT allows the distinction between Preferred Variant Labels, which are labels that will be used in practice, and Character Variant Labels, which are identified as

equivalent but are no longer used commonly. Furthermore, if the applied-for IDL is long (six characters or more), the number of Variant Labels may be very large. It may be impractical to insert all of these labels into the root zone, not only from a capacity point of view, but also from an operational one. On the other hand, delegating the applied-for Original IDL but not the Preferred Variant Label (or vice versa) will lead to unfair treatment of and confusion for users in different Chinese-speaking communities.

8.4.2 Security and Stability of the DNS

Disputes and security concerns such as phishing may arise if the Variant Labels in an IDL Package are not delegated to the same registry. This is because these labels are linguistically equivalent and in most cases can be easily mixed up by normal users due to their visual similarity. Even if the labels are delegated to the same registry, the registry may have to ensure that the same restriction will be applied to registration of Chinese domain names at the second or lower levels under the TLDs.

Character Variant Labels in an IDL Package are labels that are not recommended to be used. Consideration is required as to how these labels are blocked or reserved so that they are not inserted into the root zone either for this applicant or for another applicant.

8.4.3 Registry Fees to ICANN

Another key aspect of delegation is whether the structure of the fees to the registry has to be different from that for a single TLD and, if so, what the fee structure and amount should be.

Since the applied-for IDL is actually equivalent to the Preferred Variant Labels, and the IDL Package is meant to be treated as an atomic unit, it is advisable that ICANN should charge only the application fee for one TLD for the whole IDL Package of the applied for IDL. By the same token, if there is an annual registry fee for each IDL registered under the applied for IDN TLD, it will be desirable that the annual fee for one domain name under the TLD should be charged for the whole IDL Package under the delegated applied-for and Variant TLDs.

8.4.4 Contractual Provisioning Requirements

ICANN provides different types of agreement frameworks for ccTLD registries to suit their diverse needs. These types can continue to be used for delegation involving Variant Labels.

In particular, it may be worth having only one contract for the whole IDL Package instead of separate contracts for the IDL and the different Preferred Variant Labels, as these are equivalent and should be used together.

Irrespective of the type of agreement, it should be mentioned that the registry should ensure a good end user experience with the introduction of variant TLDs. The contract should also state any aliasing or synchronization requirements that are committed by the registry. It is advisable that a mechanism be devised to monitor whether such requirements have been realized or not. The types and frequencies of reports showing data related to variant labels should also be specified.

8.4.5 Delegating a reserved variant TLD at a later date

By definition, a Character Variant Label is a variant TLD which has been allocated but NOT delegated or activated for use by the applicant. When certain criteria are satisfied, e.g., when there is a good justification, a reserved Character Variant Label can be “released” and delegated to the

applicant. The following are issues to be considered:

- The conditions under which a reserved Character Variant Label can be delegated for use by the applicant
- The types of reasons that will be accepted as justification for delegating a previously reserved variant TLD
- Whether there will be a time limit after which such requests will no longer be entertained
- The process for reviewing such requests, including the selection of reviewers and the timing of the process

8.4.6 Un-delegating a Variant Label which was delegated previously

After a variant Label TLD has been delegated, it may be possible that it has to be taken away from the applicant at a later date. Possible reasons for this decision are changes to the root Variant Table, a dispute through PDDRP which has been decided to the favor of the complainant, or the failure of the applicant to meet the contractual requirements with ICANN. The following are issues to be considered:

- The conditions under which a delegated variant Label TLD can be undelegated
- Undelegating a TLD is a drastic action and will impact the registry, registrars, and registrants who have registered domain names under the TLD and probably been using the domain names for websites and email addresses. This not only leads to loss of business by registry and registrars, but also to angry customers who have been inconvenienced by such action. It is advisable to have an appeal process in place for the affected registry to appeal such a decision. The process for raising and handling the appeal has to be formulated and published.
- Whether undelegation should be restricted to Reserved Character Variant Labels only
- Who can decide and who has the authority to decide whether a delegated variant TLD should be undelegated
- Protection of rights of the existing registrars and registrants

8.5 Operation Issues (operational consideration after the new TLDs have been delegated)

8.5.1 Impact to Registry Operations

- DNS Resolution
    - Delegation of applied-for label and preferred variant labels will consume more computing resources because more zone files will be generated and there may be a requirement to maintain synchronization between the IDLs under the applied-for IDL and the Preferred Variant Label. This will also propagate to secondary name services and invariably increase the cost of running the DNS.

- DNSSEC
    - There is no functional impact on DNSSEC caused by variant TLDs. They appear as different zones that need to be signed and for which keys are to be generated and managed.

- Shared Registration System (SRS)
    - SRS is a critical registry function for enabling multiple registrars to provide domain name registration services in the TLD. SRS must include the EPP (Extensible

Provisioning Protocol) interface between the registry and the registrars. Extensions must be made to the EPP to enable registration of SLDs under the applied-for TLD and the variant TLDs. If different registries implement different extensions, it will be difficult and complicated for registrars to interface with these registries implementing different extensions. It is therefore desirable to standardize these extensions to support IDN TLDs with preferred variant labels.

- WHOIS
  - Registrants of domain names under IDN TLDs will likely provide non-ASCII contact information. There are two issues here:
    - The WHOIS protocol has not been internationalized. It has no mechanism to signal encodings. In absence of protocol specifications, various parties have adopted ad hoc solutions to address this issue. With the increasing adoption of IDNs, these ad hoc solutions would lead to inconsistent user experience as well as interoperability issues.
    - When a user queries the WHOIS for a domain name with Variant Labels, it is desirable to standardize the behavior in terms of whether only the queried domain name will be displayed or whether all or some of its Variant Labels will be displayed. If some or all of the Variant Labels are to be displayed, the format for displaying those labels (in particular whether both the U-labels and A-labels are to be shown) will have to be considered.

- Data Escrow

  A specific data format has been specified in the New gTLD Guidebook for registries to submit the registration data to the data escrow service provider[12]. The escrow data format currently supports variant TLDs, so no issues here.

- Trademark Clearing House (TCH)
  - This is a service that a registry can use to check whether an applied-for SLD conflicts with a registered trademark. A variant label of the IDL package that conflicts with a trademark can be blocked from registration according to the prevalent trademark-related policies of the registry.
  - Trademark Clearing Houses should be aware of IDN variant issues. – To determine whether there is a match between the applied-for SLD and a trademark in the TCH, the registry will have to query the TCH for every variant label in the IDL package of the applied-for SLD.

- The Registry has to formulate policies on how domain names are allocated, delegated, reserved, and synchronized under the new TLD recognizing that the new TLD may have variant TLDs. The Registry has to decide whether the same domain name under the new TLD and the variant TLDs have to be activated for use by the same registrant, or whether some of these have to be reserved rather than activated.

- For the convenience of registrars, registrants and others who might be affected by the variant domain names, the registry is supposed to publish its IDN table for domain name registration at the second or lower levels to public, especially when this IDN table is different from the one

---

12 See http://tools.ietf.org/html/draft-arias-noguchi-registry-data-escrow-01

applicable to TLDs in the IANA repository.

- While the IDL Package is expected to be treated as an atomic unit, the Registry may consider charging the Registrar the fee for more than one domain name for the IDL Package. Irrespective of the fee structure, the best practice is to delegate the full Traditional and the full Simplified forms of the applied-for Original IDL to the same registrant. The Registry may also consider limiting the numbers of active variants and/or other incremental fee structure to the Registrar if more than two labels (beyond the Traditional Chinese only and the Simplified Chinese only) are requested for delegation. The change in fees depends on the complexity and the multiplicity of the labels, but manageable in operation as long as such policy is well documented and published.

8.5.2 Impact to Dispute Resolutions

- The UDRP applies to disputes for domain names registered under existing gTLDs but apparently not for the gTLDs themselves. With the introduction and possible proliferation of new gTLDs (and some of these may be IDNs), new polices are needed to handle disputes between TLD registries (including IDN TLDs) and trademark owners. Post-Delegate Dispute Resolution Policy (PDDRP) is such a policy addressing such need.

- Some issues arising from Dispute Resolution for TLDs with variant labels are as follows:
    - Should new gTLD trademark measures, specifically Post-Delegate Dispute Resolution Policy (PDDRP), take into account the IDN variants of both the disputed domain names and pertinent right holders' trademarks?
    - Should the trademark clearinghouse operator accept the submission of variant forms of a trademark in IDN characters? Should the verification service provider take into account the variants of a trademark in IDN characters when verifying the identicalness of the given trademark and an applied domain name string? Should a new gTLD registry take into account the variants of a trademark in IDN characters in sunrise registration or trademark claim services?

- Some issues related to Dispute Resolution for SLDs under TLDs are as follows:
    - When someone raises a dispute of a domain name under the new TLD, should the same domain name and/or its preferred variant labels under the variant TLDs (if any) be also considered together or separately? What if these TLDs are administered by different registries? What if the variant labels of that domain name are held by different registrants?
    - Should new gTLD trademark measures, specifically Uniform Rapid Suspension Policy (URS), take into account the IDN variants of both the disputed domain names and pertinent right holder's trademarks?

9 Considerations for different roles

In this section, we focus on the issues which different entities would have to deal with when more than one Chinese Variant TLD is delegated at the root level as well as sub-level domain registration service is launched according to RFC 3743 and 4713, including but not limited to IANA, root server operators, registries, registrars, registrants, system administrators and software vendors. This section complements the previous sections and identifies several additional issues.

9.1 IANA

Once the applied-for strings are evaluated and approved, IANA shall take the responsibility to write the approved strings into the DNS root zone. This means IANA manages and maintains the registration data for new gTLD, publish the registration data to relevant parties and distribute the registration data related to zone file to root server operators.

IANA has already successfully delegated several Chinese variants ccTLDs. But facing the Chinese gTLD applicants, IANA is expected to confirm the issues below to meet the requirements in the Section 6. Consequently, IANA's solution will have a direct or indirect impact on the variant management mechanisms of TLD registries.

- Will the Whois system support IDL package specified in RFC 3743 for Chinese Whois query? For example, will the Whois system return all the variant labels or just the delegated ones? What if someone input an allocated but not delegated variant label?
- What kind of variant label resolution mechanism will IANA adopt on the root? For example, DNAME, a new RR type, multiple variant zones or others?
- Will IANA take the responsibility of building and maintaining the root Chinese variant table? If not, who should do that?

9.2 Root server operators

According to SSAC report "SAC020", the root zone can accommodate 2-5 times the number of TLDs without introducing technical instability. As far as we evaluate it, Chinese variant gTLD put no more requirements than other gTLDs on the root server operation.

9.3 Registries

Issues related to Evaluation, Allocation, Delegation and Operation of Chinese IDN Variant TLDs have been mentioned in Section 8.

9.4 Registrars

Registrars are the parties who provide IDN variant domain name registration related services to the public. In most cases, registrars provide domain name service to registrants and end users directly, sometimes resellers act on their behalf. Any requirement and changes on the variant domain names will affect the registrants and end users directly.

Noting that registrars are entitled to provide IDN variant registration services of different registries using different language tables, registrars may find it necessary to provide a clear explanation on registering various variant TLD, including its language tables and registration policies, to avoid any potential confusions. Registrars also need to consider the following questions:

- What are the charges for variant domain names? Package or separate?
- Does the registrar have to make change on its EPP client system?
- Will there be any change on the data escrow?
- How will the registrar instruct its resellers on the sale of variant domain name?

9.5 Registrants

The Chinese community began research and development on IDNs as early as in the year 2000. In 2003, second-level Chinese domain name registrations was officially introduced under .CN to

cater to the need of Chinese Internet users, and the feedback from registrants has been overwhelmingly positive.

Based on the experiences gained from registration on Chinese variant ccTLDs, registrants are also expecting resolution of the following questions on Chinese variant gTLDs:

- Will it cause confusion if the registration policies on Chinese gTLDs differ from Chinese IDN ccTLDs?
- Will it cause confusion for the registrant if different Chinese TLDs provide different variant registration policy and IDN tables?
- Has the registrant had registered any Chinese IDN ccTLDs before? If so, did RFC3743 and RFC4713 suffice to his/her need for Chinese IDN gTLD?

9.6 Domain Name related service providers

There are optional services to registrants provided by registrars or third parties including managing DNS, web hosting (HTTP, HTTPS, etc), mail hosting (SMTP, IMAP, etc), storage hosting (FTP, WebDAV, etc), digital security certificates (normally known as SSL) and others. All of these services use domain names as identifiers. The diversified expectations from customers who hold variant domain names demand flexible resolution and will significantly increase the complexity of the current system. Bearing this in mind, the service providers may need to take following questions into consideration:

- What are the requirements on delegation of variant TLDs to satisfy the requirements of domain name related service providers?
- Will there be any adjustment on their services and infrastructure to meet the demand by Chinese variant domain names?
- Will additional configuration assistant tools be needed by external solution providers or to be developed in house?

9.7 Software/Application vendors

There are increased complexities by Chinese variant TLDs with combinations of different devices, different operating systems, different input methods, different Unicode code point managements (especially conversion, recognition, processing to Chinese characters). For Software / application vendors, the following suggestions may be considered:

- The Software/Application vendors should provide convenient configuration for variant TLD related domain names.
- The Software/Application vendors should provide variant domain names conversion tools, especially the switch over between the SC and TC.

Above all, software vendors (and also service providers) are expected to evaluate their products on supporting variant TLDs, and communicate with customers on impact with their products' application logic and business logic.

10 Conclusion

This report starts focusing on issues of Han script variants in IDN top level domains, and develops several findings & recommendations that considered very important to mention is this report.

It is recognized that Chinese (or Han) script is widely used in the CJK area but there are different requirements in the different language communities, and specifically, less concern (if any) about the variant issue in these communities. Therefore, with the understanding of representatives of the Japanese and Korean communities, this issue report focuses primarily on the Chinese language requirements, with additional input from the Japanese and Korean communities.

Chinese IDL and it's variants labels should be belong to the same registrant/applicant, and the Simplified and Traditional Chinese forms of the applied-for IDL should be resolvable simultaneously or non-resolvable at all. In the past ten years, experts of JET and CDNC worked a lot on issues and solutions, which are addressed in RFC3743 and RFC4713, and deployed for many years in Mainland China, Taiwan, Hong Kong, etc.

Based the communities' knowledge and experience, we can predict that delegation of only one of both Simplified and Traditional forms of applied-for IDL, or their delegation to different applicant will cause major security and marginalization issues.

The Chinese case study team considers that language variant tables are a critical element in selecting eligible top-level domains, and therefore the policy on how those tables are developed is very important.

The variant issues will impact the TLD application, evaluation, allocation, delegation and operation. IANA, root server operators, registries, registrars, registrants, domain name related service providers, software/application providers and other stakeholders need to consider the variant issues when they adopt the variant TLDs.

11 Select Bibliography

11.1 Internet Drafts and RFCs

Klensin, J., "Suggested Practices for Registration of Internationalized Domain Names (IDN)", RFC 4290, December 2005.

Konishi, K., Huang, K., Qian, H., and Y. Ko, "Joint Engineering Team (JET) Guidelines for Internationalized Domain Names (IDN) Registration and Administration for Chinese, Japanese, and Korean", RFC 3743, April 2004.

Lee, X., Mao, W., Chen, E., Hsu, N., and J. Klensin, "Registration and Administration Recommendations for Chinese Domain Names", RFC 4713, October 2006.

Seng, J., Yoneya, Y., Huang, K., and Kyongsok, K., “Han Ideograph (CJK) for Internationalised Domain Names”, Internet Draft, Available at <http://tools.ietf.org/html/draft-ietf-idn-cjk-01>

11.2 Chinese Dictionaries and Information Processing

Halpern, J. and Kerman, J. The Pitfalls and Complexities of Chinese to Chinese Conversion, September 1999. Available at <http://www.kanji.org/cjk/c2c/c2cbasis.htm>

Hanyu da zidian bianji weiyuanhui,eds. Hanyu da zidian (汉语大字典 "Comprehensive Dictionary of Chinese Characters"). 8 vols. Wuhan: Hubei cishu chubanshe. 1986–1989.

Jeng-Wei Lin, Jan-Ming Ho, Li-Ming Tseng, and Feipei Lai. 2008. Variant Chinese Domain Name Resolution. 7, 4, Article 11 (November 2008), 29 pages. DOI=10.1145/1450295.1450296

Lu Shuxiang, ed. Xiandai Hanyu cidian (现代汉语词典 "The Contemporary Chinese Dictionary"). Beijing: Commercial Press. 1973. ISBN 7-100-03477-9

National Language Committee, People’s Republic of China. 1986. A complete set of simplified Chinese characters.

11.3 Japanese Dictionaries and Information Processing

Japanese Standards Association, "Code of the Japanese graphic character set for information interchange", JIS X 0208-1978, -1983 and -1990.

Japanese Agency for Cultural Affairs. Joyo Kanji: "regular-use Chinese characters" (常用漢字表). Available < http://www.bunka.go.jp/kokugo_nihongo/pdf/jouyoukanjihyou_h22.pdf>

11.4 Korean Dictionaries and Information Processing

Framework Act on the Korean Language” (2011) Available at <http://law.go.kr/LSW/lsInfoP.do?lsiSeq=112364#0000>

11.5 Unicode and ISO Standards

The Unicode Consortium, "The Unicode Standard, Version 6.0", (Mountain View, CA: The Unicode Consortium, 2011. ISBN 978-1-936213-01-6). <http://www.unicode.org/versions/Unicode6.0.0/>.

The Unicode Consortium, "Chapter 12: East Asian Scripts. The Unicode Standard, Version 6.0",

(Mountain View, CA: The Unicode Consortium, 2011. ISBN 978-1-936213-01-6). < http://www.unicode.org/versions/Unicode6.0.0/ch12.pdf> (note: this includes discussions about Unicode Source Separation Rules for CJK)

ISO/IEC, "ISO/IEC 10646:2011. International Standard -- Information technology - Universal Multiple-Octet Coded Character Set (UCS)", 2011.

Annex S of ISO/IEC 10646:2001: Procedure for the unification and arrangement of CJK Ideographs.

11.6 ICANN Related Documents

Internet Corporation for Assigned Names and Numbers (ICANN). Guidelines for the Implementation of Internationalised Domain Names (2003). Marina Del Rey, CA: ICANN. Retrieved September 19, 2011, from <http://www.icann.org/en/general/idn-guidelines-20jun03.htm>

Internet Corporation for Assigned Names and Numbers (ICANN). (2011) New gTLD draft Applicant Guidebook. Marina Del Rey, CA: ICANN. Retrieved September 19, 2011, from <http://www.icann.org/en/topics/new-gtlds/rfp-clean-19sep11-en.pdf>

A Chinese Case Study Team

The Chinese Case Study team was constituted, and first met, at the 41st ICANN meeting held in Singapore in June 2011. Subsequent to that meeting, it met bi-weekly by telephone to consider the set of questions, discuss and analyze their relevance and implications regarding the Chinese script. In early August, three sub teams were formed to develop sections of this report, and they met weekly to complete the task. The report also went through two readings and revisions by the whole team.

A.1 Team Membership

The Case Study team was composed of the following members:

Name	Role
Xiaodong Lee	Case Study Coordinator
Chris Dillon	Team Member
Hong Xue	Team Member
James Seng	Team Member
Jian Zhang	Team Member
Jonathan Shea	Team Member
Joseph Yee	Team Member
June Seo	Team Member
Shian-Shyong Tseng	Team Member
Wei Wang	Team Member
Yangwoo Ko	Team Member
Yoshiro Yoneya	Team Member
Zhoucai Zhang	Team Member
Edmon Chung	Observer
Yang Yu	Observer
Steve Sheng	Case Study Liaison
Francisco Arias	Subject Matter Expert (Registry Operations)
Kim Davies	Subject Matter Expert (Security)
Nicholas Ostler	Subject Matter Expert (Linguistics)
Andrew Sullivan	Subject Matter Export (Protocols)

A.2 Declarations of Interest

In order to ensure transparency by sharing relevant information on any interests the team's members have in relation to the areas of study, team members were asked to provide written statements declaring their interests.

These statements are published online at the case study team's website.[13]

A.3 Recognition

[13] https://community.icann.org/display/VIP/Chinese

The team recognizes the support provided by CDNC, the host organization for the Chinese case study team. In particular, we thank TWNIC (a member of CDNC) for their support of the face-to-face meeting, which was held in Taipei on August 21-22.

B Other References

1. The following characters assigned in the specified code ranges look similar to or related to CJK ideographs. However, in ISO/IEC 10646 and Unicode, they have been defined as scripts different from CJK Unified Ideographs. Hence, those code points shall be blocked (forbidden), no longer deemed to be variants of Chinese characters.
   Note: They could be regarded as "cross-script characters visual similarity" .

   CJK MISCELLANEOUS 3192-319F
   CJK STROKES 31C0-31E3
   ENCLOSED CJK LETTERS AND MONTHS 3200-32FF
   SMALL FORM VARIANTS FE50-FE6F
   HALFWIDTH AND FULLWIDTH FORMS FF00-FFEF
   KANGXI RADICALS 2F00-2FDF
   CJK RADICALS SUPPLEMENT 2E80-2EFF
   ENCLOSED IDEOGRAPHIC SUPPLEMENT 1F200-1F2FF
   IDEOGRAPHIC DESCRIPTION CHARACTERS 2FF0-2FFF

2. The following characters assigned in the specified code ranges look identical to CJK ideographs. However, in ISO/IEC 10646 and Unicode, they have been treated as "duplicate encoded characters" for downwards compatibility. Hence, those code points should also be blocked (forbidden).
   Note: Those characters could be regarded as "within-script character visual similarity", but cannot be included in domain names as variants.

   CJK COMPATIBILITY 3300-33FF
   CJK COMPATIBILITY IDEOGRAPHS F900-FAFF
   CJK COMPATIBILITY FORMS FE30-FE4F
   CJK COMPATIBILITY IDEOGRAPHS SUPPLEMENT 2F800-2FA1F
   Selected from ISO/IEC 10646-2011

C Brief Overview of Key RFCs for Chinese Variant handling

In this appendix, we provide a brief overview of key RFCs that current Chinese domain name operators use for handling Chinese variants. We include them as appendix because in sections of the document, these RFCs are referenced and sometimes relied on to raise certain issues. Interested readers may wish to learn more about them but would otherwise be unwilling to read through the whole RFCs, so this quick overview provide some context.

C.1 RFC 3743: Joint Engineering Team (JET) Guidelines for Internationalized Domain Names (IDN) Registration and Administration for Chinese, Japanese, and Korean, 2004

This section is an excerpt from RFC 3743 and the paper "Variant Chinese Domain Name Resolution" by Jeng-Wei Lin, Jan-Ming Ho, Li-Ming Tseng, and Feipei Lai. 2008.

RFC 3743 defines a set of IDN registration and administration guidelines for applying restrictions to CJK scripts and the zones that use these scripts. A domain registry could define its own local rules for permitted characters and the handling of IDLs and their variants. The Language Variant Table (LVT) mechanism is used to enforce language-based character variant preferences.

In a LVT, there are three columns. The first column lists a valid character that is permitted to be used in an IDL, the second column lists its preferred variants, and the third column lists other variants. As shown in figures 1 and Figure 2 below, LVTtw and LVTcn are the LVTs for Traditional Chinese and Simplified Chinese, submitted to IANA (Internet Assigned Numbers Authority) by TWNIC (Taiwan Network Information Center, the domain registry for .tw) and CNNIC (China Internet Network Information Center, the domain registry for .cn) respectively. Both tables come from the same CDNC table, which has four columns.

Valid Codepoint		Preferred Variants	Character Variants
台	(U+53F0)	台	台檯颱臺颱
檯	(U+6AAF)	台	台檯颱臺颱
颱	(U+7C49)	台	台檯颱臺颱
臺	(U+81FA)	台	台檯颱臺颱
颱	(U+98B1)	台	台檯臺颱
：		：	：
湾	(U+6E7E)	湾	灣湾
灣	(U+7063)	湾	灣湾
：		：	：
大	(U+5927)	大	大
：		：	：
学	(U+5B66)	学	學学孪
學	(U+5B78)	学	學学孪
孪	(U+6588)	学	學学孪
：		：	：
发	(U+53D1)	发	發髮髮髮
髮	(U+5F42)	发	發髮髮髮
發	(U+767C)	发	發髮髮髮
髮	(U+9AEA)	发	發髮髮髮
髮	(U+9AEE)	发	發髮髮
：		：	：
発	(U+767A)	—	—
：		：	：
余	(U+4F59)	余	余餘馀
餘	(U+9918)	余馀	余餘馀
馀	(U+9980)	馀	余餘馀

Figure 1: A snapshot of LVTcn, IANA IDN variant table for simplified Chinese, submitted by CNNIC. The first column (valid code point) and the third column (character variants) of this table are exactly the same as the LVTtw, but the preferred variants are different.

Valid Codepoint		Preferred Variants	Character Variants
台	(U+53F0)	台檯颱臺颱	台檯颱臺颱
檯	(U+6AAF)	檯	台檯颱臺颱
颱	(U+7C49)	颱	台檯颱臺颱
臺	(U+81FA)	臺	台檯颱臺颱
颱	(U+98B1)	颱	台檯臺颱
：		：	：
湾	(U+6E7E)	灣	灣湾
灣	(U+7063)	灣	灣湾
：		：	：
大	(U+5927)	大	大
：		：	：
学	(U+5B66)	學	学學孪
學	(U+5B78)	學	学學孪
孪	(U+6588)	學	学學孪
：		：	：
发	(U+53D1)	發髮	發髮髮髮
髮	(U+5F42)	發	發髮髮髮
發	(U+767C)	發	發髮髮髮
髮	(U+9AEA)	髮	發髮髮髮
髮	(U+9AEE)	髮	發髮髮
：		：	：
発	(U+767A)	—	—
：		：	：
余	(U+4F59)	余餘	余餘馀
餘	(U+9918)	餘	余餘馀
馀	(U+9980)	餘	余餘馀

Figure 2: A snapshot of LVTtw, IANA IDN variant table for traditional Chinese, submitted by TWNIC. The first column (valid code point) and the third column (character variants) of this table are exactly the same as the LVTcn, but the preferred variants are different.

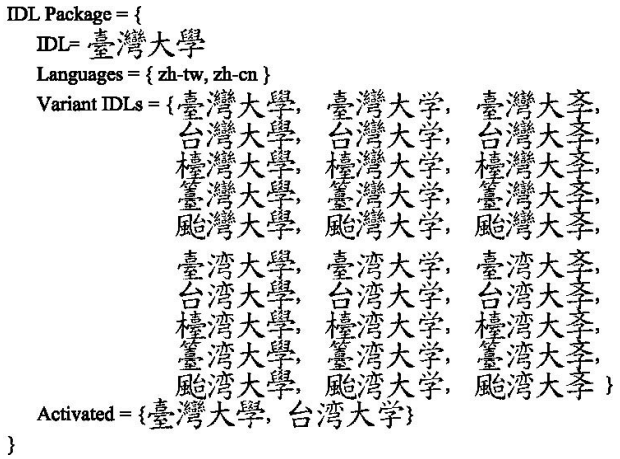

Figure 3: An example IDL package.

When an IDL is registered, a collection of its variant IDLs, known as an IDL package, is created according to the zone-specified LVTs. All variant IDLs in the IDL package are unavailable to other name holders. In addition to the registered IDL, the name holder can request the domain registry to activate some of the variant IDLs in the IDL package, for free or for an extra fee. Activated IDLs are stored in the zone files and thus become resolvable, and other IDLs are reserved. If a domain registry allows an IDL to associate with more than one language or script, multiple LVTs will be used to create the IDL package. For example, TWNIC and CNNIC employ a registration policy that allows an IDL to associate with both Traditional Chinese and Simplified Chinese [RFC 4713]. When an IDL 臺 灣 大 學 (National Taiwan University, U+81FA U+7063 U+5927 U+5B78) is registered, an IDL package is created according to LVT of ".tw" and LVT of ".cn". By substituting one or more characters in the IDL with their variants, the package includes 30 IDLs that are mutually variant IDLs of each other, as shown in Figure 3. While two of them, 臺灣大學 (U+81FA U+7063 U+5927 U+5B78) and 台湾大学 (U+53F0 U+6E7E U+5927 U+5B66), are activated, others are reserved.

An IDL package is an atomic unit for IDL registration. Once an IDL package is created, no IDLs can be inserted into or removed from the IDL package during its lifetime. However, the name holders can dynamically request the domain registry to activate and deactivate some variant IDLs. When the IDL package is destroyed, due to either being unregistered or revoked, all IDLs in the package are available again to all name holders at the same time.

C.2 RFC 4713: Registration and Administration Recommendations for Chinese Domain Names, RFC 4713, 2006.

RFC 4713, which is contributed by CDNC members, describes how ".CN" and ".TW" apply the principles of RFC 3743 to manage Simplified Chinese and Traditional Chinese domain name equivalence, RFC4290 is also adopted and referred by RFC4713.

The key point of RFC4713 is that both Zh-CN and Zh-TW are select as the language tags by default (mentioned in section 3.2), it make sure the Simplified Chinese domain labels and Traditional Chinese domain labels are put into the zone variant and added into the zone file.

This RFC need to considered with the ".CN" and ".TW" language variant tables registered in IANA together, the code point list of ".CN" and ".TW" language variant tables is same.